



UNIVERSITAT  
POLITÈCNICA  
DE VALÈNCIA

---

# Study of molecular bases behind sex differences in non-alcoholic fatty liver disease and multiple sclerosis through transcriptomic meta-analysis

April 2025

---

Author: José Francisco Català Senent

Supervisors: Dr. Francisco García García  
Dr. Marta R. Hidalgo García

Tutor: Dr. Máximo Ibo Galindo Orozco





*A mis padres*



# Agradecimientos

Quiero empezar esta tesis agradeciendo a todas las personas que, de una forma u otra, la han hecho posible. A todos aquellos me habéis acompañado a lo largo de mi travesía investigadora desde hace ya bastantes años. Espero no dejarme a nadie, pero si lo hago, perdonadme, sois muchos.

Me gustaría comenzar dando las gracias a mis padres y a mi hermana, porque sin todo su apoyo y su ayuda esta tesis no existiría.

Agradecer también a mis directores Paco y Marta por la oportunidad que me dieron hace más 5 años de comenzar este proyecto. Por todo lo que me han enseñado, por todos los esfuerzos para sacar esta tesis adelante y por su optimismo. No sé si soy realmente consciente de la suerte que he tenido teniéndooos como directores.

A todo el equipazo de la UBB, aquellos que entramos juntos allá por 2018 (Rubén Grillo, Adolfo, Sergio, Irene Pérez, Raúl, Pau, Sandra, Franc) y hemos hecho gran parte de este camino juntos. A Irene Soler y Carla, que llegaron más tarde, pero con las que he compartido tantas horas en la UBB. A Rubén Sánchez, que fue quien me enseñó todo lo que sé sobre el *clúster*. A Zoraida, por toda su ayuda y apoyo. A Máximo Ibo, tutor de esta tesis. Finalmente, a *Kim Jong-un* (Quim, Jhon y José), a Xavi y a toda la gente del CIPF que ahora no soy capaz de recordar.

A los Residentes en Valencia (Koke, Alberto, Luz, Miguel y David), con los que empecé en la bioinformática, y a Sara, que llegó después, por todos los buenos momentos, por los *dual cores*, por *Seq-Bios*, por los pardales, etc. No podría

haber encontrado a mejores personas en mi vuelta a las clases allá por 2016. Además, me gustaría aprovechar estas líneas para sumarme al llamamiento de reencuentro que propuso en los agradecimientos de su tesis el ya Dr. Martín-Arana.

A la *Secta de Biólogos* (Pepe, Nico, Ana Elena, Miguel, Anabel, Marc, Silvia, Susana, Lauri, Juanan, Xinxe, Diego, Javi, Patri, Marian, Rebeka, Tere y Silbia) por tantos momentos, como diría Ana Elena: sois los mejores.

A mi grupo de amigos del colegio que han apoyado sin acabar de entender del todo lo que hacía, por las cenas de los viernes en *la China* que ahora han convertido en almuerzos, especialmente a Salva, a Adri y a Broco. A Ger y Dori, por las tardes aprendiendo arquitectura en la tetería, por los viajes por la isla, por acogerme en Londres y por todo.

A la gente que conocí en INCLIVA. A la gente de los lunes y de la plataforma de bioinformática con las compartí tantas horas en *Narnia*. A Laura Castillo, por apoyarme y aguantar mis quejas durante el final de esta tesis. A la gente de medicina de precisión y de informática, he aprendido muchísimo gracias a vosotros y vosotras.

A Elena G. Biosca, por acogerme en 2006 en aquel laboratorio 3.58 de Burjassot donde empezó a enseñarme a investigar: gracias por todo. Agradecer también a la toda la gente con la que he compartido laboratorio en Bacplant, en el IVIA, en el CITA y en Glasgow.

Finalmente, quiero agradecer a toda la comunidad que, de forma altruista, ofrece sus conocimientos en foros como *Stack Overflow* o *Biostarts*, ayudando a otros usuarios y al avance de la ciencia.

# Abstract

Sex differences in health are one of the primary biases in medicine today. Traditionally, males have been used as models in research and medicine, ignoring the fact that the development, symptoms, and response to a pathology are not necessarily the same in both sexes. Fortunately, in recent decades, researchers and institutions have made significant advances to reduce this bias, with the aim of providing more efficient treatment of diseases.

In this context, this doctoral thesis has contributed in reducing sex bias in medicine by studying the molecular basis of sex-related alterations in two diseases with a significant impact on the population: non-alcoholic fatty liver disease (NAFLD) and multiple sclerosis (MS). NAFLD comprises a range of liver disorders that begin with an accumulation of fat in the hepatocytes (NAFL) and progress to the development of inflammation (NASH), which can lead to loss of liver function and cancer. The latest studies estimate the overall prevalence of the disease at around 30% (with an upward trend), being higher in men (40%) than in women (26%). Among women, a lower prevalence is observed in premenopausal women, while after menopause, it is equal to that of men. This high prevalence of the pathology makes NAFLD a worldwide problem and the most common liver disease in developed countries.

On the other hand, MS is a chronic autoimmune, inflammatory, and degenerative disease of the central nervous system. This pathology presents different forms and stages and can, in the most severe cases, seriously affect the quality of life of those who suffer from it. As is usual in neurodegenerative diseases,

MS is more prevalent in women than in men and is mainly diagnosed in young adults.

In the NAFLD work, 7 studies, including 323 patients, were selected from the 114 baseline work. Meta-analyses identified molecular mechanisms significantly overexpressed in premenopausal women versus men, including processes associated with DNA regulation, vinculin binding, responses to interleukin-2 (IL-2), negative regulation of neuronal death, and ion and cation transport. In males, we discovered an overrepresentation of genes associated with negative regulation of interleukin-6 (IL-6) and the establishment of planar polarity involved in neural tube closure. The differences in immune response observed between premenopausal men and women in the transition from NAFL to NASH suggest the role of IL-2 and IL-6 in the differential progression of NAFLD as a function of sex.

On the other hand, in the chapter on MS, 122 publications were reviewed, of which 9 met the defined criteria. These 9 studies (5 of blood samples and 4 of brain tissue samples) included 474 individuals. The blood and brain tissue meta-analyses identified, respectively, 1 and 13 MS-associated genes altered differently between men and women. In the case of functional analyses in the brain, sex-associated immune patterns were detected, among others. Women were more affected in the proinflammatory environment and innate immune responses related to the myeloid lineage, whereas adaptive responses were more affected in men.

Overall, the results of this dissertation reaffirm the usefulness of the systematic review and meta-analysis strategy as a tool for the reuse of openly published data. In this case, the application of this methodology has allowed us to detect a set of molecular alterations affected differently between women and men in both non-alcoholic fatty liver disease and multiple sclerosis. Particularly noteworthy is the presence of changes associated with interleukins 2 and 6 in both diseases, which highlights the role that the immune system plays in sex-associated alterations. The results of this work open the door to further research to translate these results into clinical practice.

# Resumen

Las diferencias de sexo en salud son uno de los mayores sesgos existentes en la medicina actual. Tradicionalmente se han usado machos como modelo en investigación y en medicina, obviando que el desarrollo, los síntomas y la respuesta frente a una patología no es necesariamente igual en ambos sexos. Afortunadamente, aunque aún queda trabajo por hacer, en las últimas décadas se han realizado grandes avances por parte de investigadores e instituciones para reducir este sesgo, con el objetivo de proporcionar un tratamiento más eficiente de las enfermedades.

En este contexto, esta tesis doctoral ha contribuido en la reducción del sesgo de sexo en medicina, estudiando las bases moleculares alteradas en función del sexo en dos enfermedades con un importante impacto en la población. Las patologías seleccionadas fueron la enfermedad del hígado graso no alcohólico (NAFLD, por sus siglas en inglés) y la esclerosis múltiple (MS, por sus siglas en inglés). El NAFLD comprende un abanico de alteraciones hepáticas que comienza con una acumulación de grasa en los hepatocitos (NAFL, por sus siglas en inglés) y avanza con la aparición de inflamación (NASH, por sus siglas en inglés), pudiendo desembocar en pérdida de la función hepática y cáncer. Los últimos estudios estiman una prevalencia global de la enfermedad alrededor del 30% (con una tendencia al alza), siendo mayor en hombres (40%) que en mujeres (26%). Dentro de las mujeres, se observa una prevalencia inferior en las mujeres premenopáusicas, mientras que tras la menopausia esta se iguala con la de los hombres. Esta alta prevalencia de la patología hace del NAFLD un problema mundial y la enfermedad hepática más común en los países desarrollados. Por otro lado, la MS es una enfermedad crónica autoin-

mune, inflamatoria y degenerativa del sistema nervioso central. Esta patología presenta distintas formas y estadios, pudiendo, en los casos más graves, afectar notablemente la calidad de vida de quienes la padecen. Como es habitual en las enfermedades neurodegenerativas, la MS presenta una mayor prevalencia en mujeres que en hombres, siendo principalmente diagnosticada en adultos jóvenes.

En el trabajo de NAFLD se seleccionaron 7 estudios, incluyendo 323 pacientes, de los 114 trabajos de partida. Los metaanálisis identificaron mecanismos moleculares significativamente sobreexpresados en las mujeres premenopáusicas frente a los hombres, incluyendo procesos asociados con la regulación del ADN, la unión de vinculina, las respuestas a la interleucina-2 (IL-2), la regulación negativa de la muerte neuronal y, el transporte de iones y cationes. En los hombres, descubrimos la sobrerrepresentación de genes asociados con la regulación negativa de la interleucina-6 (IL-6) y el establecimiento de la polaridad planar implicada en el cierre del tubo neural. Las diferencias en la respuesta inmunitaria observadas entre hombres y mujeres premenopáusicas en el paso de NAFL a NASH sugieren el papel de las IL-2 e IL-6 en la progresión diferencial del NAFLD en función del sexo.

Por otro lado, en el capítulo sobre la MS se revisaron 122 publicaciones, de las cuales 9 cumplieron con los criterios definidos. Estos 9 estudios (5 de muestras de sangre y 4 de muestras de tejido cerebral) incluían un total de 474 individuos. Los metaanálisis de sangre y tejido cerebral identificaron, respectivamente, 1 y 13 genes asociados a la MS alterados de forma distinta entre hombres y mujeres. En el caso de los análisis funcionales en el cerebro se detectaron, entre otros, patrones inmunitarios asociados al sexo. En mujeres se observó una mayor afectación del entorno proinflamatorio y las respuestas inmunitarias innatas relacionadas con el linaje mieloide, mientras que las respuestas adaptativas se vieron más afectadas en los hombres.

En conjunto, los resultados de esta disertación reafirman la utilidad de la estrategia de revisión sistemática y metaanálisis como una herramienta para la reutilización de datos publicados en abierto. En este caso, la aplicación de esta metodología nos ha permitido detectar un conjunto de alteraciones moleculares afectadas de distinta forma entre mujeres y hombres, tanto en la enfermedad del hígado graso no alcohólico como en la esclerosis múltiple. Particularmente, es destacable la presencia de cambios asociados a las interleucinas 2 y 6 en ambas enfermedades, lo cual resalta el papel que el sistema inmune juega en las alteraciones asociadas el sexo. Los resultados de este trabajo abren la puerta a investigaciones más profundas que permitan trasladar estos resultados a la práctica clínica.



# Resum

Les diferències de sexe en salut són un dels majors biaixos existents en la medicina actual. Tradicionalment s'han usat mascles com a model en investigació i en medicina, obviant que el desenvolupament, els símptomes i la resposta enfront d'una patologia no és necessàriament igual en tots dos sexes. Afortunadament, encara que encara queda treball per fer, en les últimes dècades s'han realitzat grans avanços per part d'investigadors i institucions per a reduir aquest biaix, amb l'objectiu de proporcionar un tractament més eficient de les malalties.

En aquest context, aquesta tesi doctoral ha contribuït en la reducció del biaix de sexe en medicina estudiant les bases moleculars alterades en funció del sexe en dues malalties amb un important impacte en la població. Les patologies seleccionades van ser la malaltia del fetge gras no alcohòlic (NAFLD, per les seues sigles en anglés) i l'esclerosi múltiple (MS, per les seues sigles en anglés). El NAFLD comprén un ventall d'alteracions hepàtiques que comença amb una acumulació de greix als hepatòcits (NAFL, per les seues sigles en anglés) i avança amb l'aparició d'inflamació (NASH, per les seues sigles en anglés), podent desembocar en pèrdua de la funció hepàtica i càncer. Els últims estudis estimen una prevalença global de la malaltia al voltant del 30% (amb una tendència a l'alça), sent major en homes (40%) que en dones (26%). Dins de les dones, s'observa una prevalença inferior en les dones premenopàusiques, mentre que després de la menopausa aquesta s'igualava amb la dels homes. Aquesta alta prevalença de la patologia fa del NAFLD un problema mundial i la malaltia hepàtica més comuna als països desenvolupats.

D'altra banda, la MS és una malaltia crònica autoimmunità, inflamatòria i degenerativa del sistema nerviós central. Aquesta patologia presenta diferents formes i estadis podent, en els casos més greus, afectar greument la qualitat de vida dels qui la pateixen. Com és habitual en les malalties neurodegeneratives, la MS presenta una major prevalença en dones que en homes, sent principalment diagnosticada en adults joves.

En el treball de NAFLD es van seleccionar 7 estudis, incloent 323 pacients, dels 114 treballs de partida. Els metaanàlisis van identificar mecanismes moleculars significativament sobreexpressats en les dones premenopàusiques enfront dels homes, incloent processos associats amb la regulació de l'ADN, la unió de vinculina, les respostes a la interleucina-2 (IL-2), la regulació negativa de la mort neuronal i el transport d'ions i cations. En els homes, descobrim la sobrerepresentació de gens associats amb la regulació negativa de la interleucina-6 (IL-6) i l'establiment de la polaritat planar implicada en el tancament del tub neural. Les diferències en la resposta immunitària observades entre homes i dones premenopàusiques en el pas de NAFL a NASH suggereixen el paper de les IL-2 i IL-6 en la progressió diferencial del NAFLD en funció del sexe.

D'altra banda, en el capítol sobre la MS es van revisar 122 publicacions, de les quals 9 van complir amb els criteris definits. Aquests 9 estudis (5 de mostres de sang i 4 de mostres de teixit cerebral) inclouen un total de 474 individus. Els metaanàlisis de sang i teixit cerebral van identificar, respectivament, 1 i 13 gens associats a la MS alterats de manera diferent entre homes i dones. En el cas dels anàlisis funcionals en el cervell es van detectar, entre altres, patrons immunitaris associats al sexe. En dones es va observar una major afectació de l'entorn proinflamatori i les respostes immunitàries innates relacionades amb el llinatge mieloide, mentre que les respostes adaptatives es van veure més afectades en els homes.

En conjunt, els resultats d'aquesta dissertació reafirmen la utilitat de l'estratègia de revisió sistemàtica i metaanàlisis com una eina per a la reutilització de dades publicades en obert. En aquest cas, l'aplicació d'aquesta metodologia ens ha permès detectar un conjunt d'alteracions moleculars afectades de diferent forma entre dones i homes tant en la malaltia del fetge gras no alcohòlic com en l'esclerosi múltiple. Particularment, és destacable la presència de canvis associats a les interleucines 2 i 6 en totes dues malalties, la qual cosa ressalta el paper que el sistema immune juga en les alteracions associades al sexe. Els resultats d'aquest treball obrin la porta a investigacions més profundes que permeten traslladar aquests resultats a la pràctica clínica.

# Contents

|                                                                                                                               |             |
|-------------------------------------------------------------------------------------------------------------------------------|-------------|
| <b>Abstract</b>                                                                                                               | <b>vii</b>  |
| <b>Contents</b>                                                                                                               | <b>xiii</b> |
| <b>Abbreviations</b>                                                                                                          | <b>xxi</b>  |
| <b>1 Introduction</b>                                                                                                         | <b>1</b>    |
| 1.1 Sex perspective in research, medicine and health . . . . .                                                                | 1           |
| 1.1.1 Sex and gender: definitions . . . . .                                                                                   | 2           |
| 1.1.2 Sex differences in medicine and health . . . . .                                                                        | 2           |
| 1.1.3 Sex differences in research . . . . .                                                                                   | 3           |
| 1.1.4 Sex differences in diseases . . . . .                                                                                   | 5           |
| 1.2 Molecular biology and omic sciences . . . . .                                                                             | 12          |
| 1.2.1 Omic sciences . . . . .                                                                                                 | 13          |
| 1.2.2 Transcriptomics . . . . .                                                                                               | 14          |
| 1.3 Systematic review and meta-analysis . . . . .                                                                             | 17          |
| 1.3.1 Systematic review . . . . .                                                                                             | 18          |
| 1.3.2 Meta-analysis . . . . .                                                                                                 | 19          |
| <b>2 Objectives</b>                                                                                                           | <b>21</b>   |
| <b>3 Hepatic steatosis and steatohepatitis: a functional meta-analysis of sex-based differences in transcriptomic studies</b> | <b>23</b>   |
| 3.1 Abstract . . . . .                                                                                                        | 25          |
| 3.2 Background . . . . .                                                                                                      | 27          |

|       |                                                                                         |           |
|-------|-----------------------------------------------------------------------------------------|-----------|
| 3.3   | Materials and methods . . . . .                                                         | 28        |
| 3.3.1 | Literature review . . . . .                                                             | 28        |
| 3.3.2 | Study exclusion criteria . . . . .                                                      | 28        |
| 3.3.3 | Bioinformatics analysis . . . . .                                                       | 28        |
| 3.3.4 | Data acquisition and preprocessing . . . . .                                            | 29        |
| 3.3.5 | Differential gene expression and functional enrichment analysis . . . . .               | 30        |
| 3.3.6 | Differential pathway activation and functional analysis . . . . .                       | 31        |
| 3.3.7 | Meta-analysis . . . . .                                                                 | 32        |
| 3.3.8 | Availability of data and materials . . . . .                                            | 32        |
| 3.4   | Results . . . . .                                                                       | 34        |
| 3.4.1 | Systematic review and exploratory analysis . . . . .                                    | 34        |
| 3.4.2 | Individual analysis . . . . .                                                           | 37        |
| 3.4.3 | Meta-analysis . . . . .                                                                 | 37        |
| 3.4.4 | Gene functional meta-analysis . . . . .                                                 | 38        |
| 3.4.5 | Pathway functional meta-analysis . . . . .                                              | 38        |
| 3.4.6 | Metafun-NAFLD platform . . . . .                                                        | 39        |
| 3.5   | Discussion . . . . .                                                                    | 43        |
| 3.5.1 | Perspectives and significance . . . . .                                                 | 46        |
| 3.6   | Conclusions . . . . .                                                                   | 46        |
| 3.7   | Availability of data and materials . . . . .                                            | 46        |
| 3.8   | References . . . . .                                                                    | 47        |
| 3.9   | Acknowledgements . . . . .                                                              | 53        |
| 3.10  | Funding . . . . .                                                                       | 53        |
| 3.11  | Authors' contributions . . . . .                                                        | 53        |
| 3.12  | Author details . . . . .                                                                | 54        |
| 3.13  | Supplementary material . . . . .                                                        | 54        |
| 4     | <b>A deep transcriptome meta-analysis reveals sex differences in multiple sclerosis</b> | <b>61</b> |
|       | Abstract . . . . .                                                                      | 63        |
| 4.1   | Abstract . . . . .                                                                      | 63        |
| 4.2   | Background . . . . .                                                                    | 65        |
| 4.3   | Materials and methods . . . . .                                                         | 66        |
| 4.3.1 | Literature Review and Study Selection . . . . .                                         | 66        |
| 4.3.2 | Bioinformatic Analyses . . . . .                                                        | 67        |
| 4.3.3 | Data Acquisition and Preprocessing . . . . .                                            | 67        |
| 4.3.4 | Comparisons . . . . .                                                                   | 67        |
| 4.3.5 | Differential Gene Expression and Meta-analyses . . . . .                                | 69        |
| 4.3.6 | Functional Enrichment Analysis . . . . .                                                | 69        |
| 4.3.7 | Metafun-MS Platform . . . . .                                                           | 70        |
| 4.4   | Results . . . . .                                                                       | 71        |

|           |                                                                                    |            |
|-----------|------------------------------------------------------------------------------------|------------|
| 4.4.1     | Systematic Review and Exploratory Analysis . . . . .                               | 71         |
| 4.4.2     | Differential Gene Expression Analysis . . . . .                                    | 71         |
| 4.4.3     | Meta-analysis of DGE Profiles . . . . .                                            | 75         |
| 4.4.4     | Meta-analysis of DGE in Blood . . . . .                                            | 76         |
| 4.4.5     | Meta-analysis of DGE in Brain Tissue . . . . .                                     | 76         |
| 4.4.6     | Functional Enrichment of the Brain-Tissue DGE Meta-analysis . . . . .              | 78         |
| 4.4.7     | Metafun-MS platform . . . . .                                                      | 82         |
| 4.5       | Discussion . . . . .                                                               | 84         |
| 4.5.1     | Meta-analysis in blood . . . . .                                                   | 84         |
| 4.5.2     | Meta-analysis in brain tissue . . . . .                                            | 86         |
| 4.5.3     | Strengths and limitations . . . . .                                                | 89         |
| 4.6       | Conclusions . . . . .                                                              | 90         |
| 4.7       | Additional files . . . . .                                                         | 91         |
| 4.8       | Availability of data and materials . . . . .                                       | 91         |
| 4.9       | Computer Code and Software . . . . .                                               | 91         |
| 4.10      | References . . . . .                                                               | 92         |
| 4.11      | Acknowledgements . . . . .                                                         | 104        |
| 4.12      | Funding . . . . .                                                                  | 105        |
| 4.13      | CRediT authorship contribution statement . . . . .                                 | 105        |
| 4.14      | Author details . . . . .                                                           | 105        |
| 4.15      | Supplementary material . . . . .                                                   | 107        |
| <b>5</b>  | <b>General discussion</b>                                                          | <b>111</b> |
| 5.1       | Sex differences in health and research . . . . .                                   | 111        |
| 5.2       | Research synthesis and Open Science . . . . .                                      | 116        |
| 5.2.1     | Research synthesis . . . . .                                                       | 116        |
| 5.2.2     | Open Science . . . . .                                                             | 117        |
| <b>6</b>  | <b>Conclusions</b>                                                                 | <b>121</b> |
|           | <b>References</b>                                                                  | <b>123</b> |
|           | <b>Artificial intelligence statement</b>                                           | <b>149</b> |
| <b>I</b>  | <b>Interpretation of meta-analysis results</b>                                     | <b>151</b> |
| I.1       | Evaluation of the combined effect of the primary studies . . . . .                 | 152        |
| I.2       | Evaluation of the heterogeneity and the influence of the primary studies . . . . . | 153        |
| <b>II</b> | <b>Scientific contributions during the period of this doctoral dissertation</b>    | <b>157</b> |



# List of Figures

|      |                                                                                                                                                    |    |
|------|----------------------------------------------------------------------------------------------------------------------------------------------------|----|
| 1.1  | Decision tree for the incorporation of the sex factor in research studies . . .                                                                    | 5  |
| 1.2  | Prevalence of a set of autoimmune or inflammatory diseases in females and males . . . . .                                                          | 9  |
| 1.3  | Prevalence of a set of neurodegenerative diseases in females and males . . .                                                                       | 11 |
| 1.4  | Microarray and sequencing flow cell . . . . .                                                                                                      | 16 |
| 1.5  | Flow of information through the different phases of a systematic review following the PRISMA 2009 statement . . . . .                              | 19 |
| 3.1  | Data analysis workflow . . . . .                                                                                                                   | 29 |
| 3.2  | Flow diagram of our systematic review of the literature and selection of studies for meta-analysis, according to the PRISMA statement guidelines . | 35 |
| 3.3  | Representation of results of the bioinformatics analysis . . . . .                                                                                 | 40 |
| 3.4  | Differential functional profiling by sex . . . . .                                                                                                 | 42 |
| 3.S1 | UpSet plots of our gene functional analysis for each individual study . . .                                                                        | 55 |
| 3.S2 | Forest plots for significant GO terms from our functional gene meta-analysis.                                                                      | 56 |
| 3.S3 | Forest plots of significant GO terms from our functional pathways meta-analysis. . . . .                                                           | 57 |
| 3.S4 | Funnel plots of the significant GO terms from our functional gene meta-analysis. . . . .                                                           | 58 |
| 3.S5 | Funnel plots of significant GO terms from our functional pathway meta-analysis. . . . .                                                            | 59 |
| 4.1  | Flowchart describing the work performed . . . . .                                                                                                  | 68 |
| 4.2  | PRISMA diagram of the systematic review . . . . .                                                                                                  | 72 |
| 4.3  | Boxplot of LFC values for the DGE in the SDID comparison . . . . .                                                                                 | 75 |

|      |                                                                                                               |     |
|------|---------------------------------------------------------------------------------------------------------------|-----|
| 4.4  | Upset plot of the meta-analysis results in blood studies . . . . .                                            | 77  |
| 4.5  | Upset plot of the meta-analysis results in brain tissue samples . . . . .                                     | 80  |
| 4.6  | Upset plots of the functional enrichment results in brain tissue for BP GO terms and KEGG pathways . . . . .  | 81  |
| 4.7  | Bar chart of the number of significant BP GO terms in the IDF and IDM comparisons . . . . .                   | 82  |
| 4.8  | Dot plot of a representative selection of the specific significant BP GO terms in IDF and IDM . . . . .       | 83  |
| 4.S1 | Principal component analysis plot of study GSE62584 . . . . .                                                 | 107 |
| 4.S2 | Bar chart representing the possible causes of a positive or negative LFC/LOR in the SDID comparison . . . . . | 108 |
| I.1  | Example of forest plot . . . . .                                                                              | 153 |
| I.2  | Example of funnel plots . . . . .                                                                             | 155 |
| I.3  | Example of influence plots . . . . .                                                                          | 156 |



# List of Tables

|     |                                                                                                                                         |     |
|-----|-----------------------------------------------------------------------------------------------------------------------------------------|-----|
| 1.1 | Known sex differences in a selection of diseases . . . . .                                                                              | 7   |
| 3.1 | Characteristics of the studies selected for inclusion in our analysis in terms<br>of disease stage, sex, BMI, and patient age . . . . . | 36  |
| 3.2 | Number of significant GO terms and KEGG pathways in individual studies<br>after gene functional analysis . . . . .                      | 37  |
| 3.3 | Significant functions in the functional meta-analyses . . . . .                                                                         | 41  |
| 4.1 | Description of selected studies . . . . .                                                                                               | 73  |
| 4.2 | Distribution of samples by sex and experimental group . . . . .                                                                         | 73  |
| 4.3 | The number of significantly differentially-expressed genes by study (and tissue)                                                        | 74  |
| 4.4 | Results of meta-analyses by comparison and tissue . . . . .                                                                             | 76  |
| 4.5 | Meta-analysis results of blood samples . . . . .                                                                                        | 78  |
| 4.6 | Meta-analysis results from brain tissue samples . . . . .                                                                               | 79  |
| 5.1 | Significant GO terms associated with IL-2 or IL-6 in both studies . . . . .                                                             | 113 |



# Abbreviations

**AD** Alzheimer's disease.

**ADN** Ácido desoxirribonucleico, àcid desoxiribonucleic.

**AI** Artificial intelligence.

**ALS** Amyotrophic lateral sclerosis.

**BBB** Blood brain barrier.

**BH** Benjamini & Hochberg p-value adjust method.

**BMI** Body mass index.

**BP** Biological processes of Gene Ontology.

**BY** Benjamini & Yekutieli p-value adjust method.

**cDNA** Complementary DNA (DNA reverse-transcribed from RNA).

**CI** Confidence interval.

**CIS** Clinical isolated syndrome.

**CNS** Central nervous system.

**CRISPR** Clustered regularly interspaced short palindromic repeats.

**DGE** Differential gene expression.

**DNA** Deoxyribonucleic acid.

**E2** Estradiol.

**EST** Expressed sequence tag.

**FAIR** Findable, accessible, interoperable, and reusable (data principles).

**FDR** False discovery rate (also BH).

**GEO** Gene Omnibus Expression database.

**GO** Gene Ontology.

**GSA** Gene set analysis.

**HGNC** HUGO gene nomenclature committee.

**IDF** Impact of the disease in females (comparison).

**IDM** Impact of the disease in males (comparison).

**IL** Interleukin.

**IL2RA** Interleukin-2 receptor alpha.

**KEGG** Kyoto Encyclopedia of Genes and Genomes.

**LFC** Logarithm of the fold change.

**lncRNA** Long non-coding ribonucleic acid.

**logFC** Logarithm of the fold change.

**LOR** Logarithm of the odds ratio.

**MA** Meta-analysis.

**MD-GSA** Multidimensional gene-set analysis.

**MHC** Major histocompatibility complex.

**mRNA** Messenger ribonucleic acid.

**MS** Multiple sclerosis.

**NAFL** Non-alcoholic steatosis.

**NAFLD** Non-alcoholic fatty liver disease.

**NASH** Non-alcoholic steatohepatitis.

**NCBI** National Center for Biotechnology Information, US.

**NDD** Neurodegenerative disorder.

**NK** Natural killer (cells).

**NOD** Nucleotide-binding oligomerization domain.

**PA** Parkinson's disease.

**PAA** Pathway activity analysis.

**PCA** Principal component analysis.

**PCR** Polymerase chain reaction.

**PPMS** Progressive primary multiple sclerosis.

**PRISMA** Preferred reporting items for systematic reviews and meta-analyses.

**PRMS** Progressive-relapse multiple sclerosis.

**RMA** Robust multi-array average (normalization method).

**RNA** Ribonucleic acid.

**RNA-Seq** Ribonucleic acid sequencing.

**RRMS** Relapsing-remitting multiple sclerosis.

**SAGE** Serial analysis of gene expression.

**SAT** Subcutaneous adipose tissue.

**scRNA-Seq** Single-cell ribonucleic acid sequencing.

**SD** Standard deviation.

**SDID** Sex-differential impact of the disease (comparison).

**SPMS** Secondary-progressive multiple sclerosis.

**TMM** Trimmed mean of M values (normalization method).

**UNESCO** United Nations educational, scientific, and cultural organization.

**US** United States (of America).

**VAT** Visceral adipose tissue.

## Chapter 1

# Introduction

### 1.1 Sex perspective in research, medicine and health

The sex of individuals is a determining factor in the species in which this element is present. In humans, since prehistoric times, sex has been a pivotal factor in the construction of society. However, this aspect of individuals has long been overlooked in key fields such as research, medicine, and health. This doctoral thesis seeks to help fill this gap in the knowledge of science, deciphering the molecular mechanisms that differentially affect males and females in the diseases in which these differences have been described, such as non-alcoholic fatty liver disease and multiple sclerosis.

To this end, this section will introduce the role that the sex perspective has played in research, medicine, and health and how this factor has evolved up to the present day.

### ***1.1.1 Sex and gender: definitions***

The concepts of sex and gender are often used interchangeably in the scientific literature, although their meaning is not the same. Traditionally, the term sex refers to the biological characteristics that define humans as female or male, i.e., chromosomes, reproductive organs, and physiological functions [1, 2]. On the other hand, gender is a social construct that may or may not correspond to sex. It refers to the set of socially constructed roles and relationships, personality traits, attitudes, behaviors, values, relative power, and influence that society differentially attributes to each gender [1].

Both sex and gender have a modifying effect on health, disease and medicine [3]. However, in this work we will evaluate the effect of sex, as a purely biological concept, on different diseases.

In this thesis, unless otherwise specified, when naming women and men we will refer to the sexual aspect of the terms.

### ***1.1.2 Sex differences in medicine and health***

Sex is a crucial factor in medical practice. Genetic, biochemical, anatomical, and physiological aspects are determining elements in diagnosing and treating diseases, and are influenced by the sex of the individual. However, throughout history, men have been considered the standard in medical studies without taking into account all the clinical implications for women [4, 5]. In the beginning of anatomical research, most of the bodies used for dissection and drawing were male, while references to representations of female bodies tended to the ideas of Aristotle's observation "the female is, as it were, a mutilated male" [4, 6] or Galen's idea of women as "imperfect" representing a "turned in" version of the men [4, 7]. This bias has survived until recent times. A review of 31 anatomy textbooks published between 1890 and 1989 in the United States found that the number of female figures was only 30% of those with a defined sex [8]. More recently, a study analyzing medicine and general surgery textbooks from 2013 determined the percentage of females with clearly discernable sex was 39.58% and 46.74%, respectively [9].

One of the first publications that focused on the underrepresentation of the female sex in medicine was published by Ayanian and Epstein [10] in 1991 in the journal *The New England Medicine*. In this work, it was observed that, in hospitals in Massachusetts and Maryland (United States), women hospitalized for coronary artery disease underwent fewer diagnostic tests (coronary angiography) and fewer major therapeutic procedures (revascularization) than men.



As a conclusion, the authors suggested that women may not have equivalent access to these procedures at a time when the incidence of coronary heart disease among women was increasing.

These biases in textbooks and medicine have negative consequences for those groups that escape the usual tendency. Historically, there has been little research into the genetics, biochemistry and anatomy of women and girls in the health sciences [4]. By escaping the prototypical signs, particular diseases are underdiagnosed (or diagnosed late) in women. In addition, the presumption of equality for both sexes in therapeutic strategies may cause them to be inappropriate for women, leading to inadequate treatment and prognosis. An example of how the lack of consideration of sex differences in pharmacokinetics and pharmacodynamics [11] can have worrying consequences is that 8 of the 10 drugs withdrawn from the market in the United States between 1997 and 2001 presented a higher health risk for women than men [12].

For all these reasons, despite the advancements made, it is essential to continue working to address the existing sex gap in medicine, striving toward more precise medical practices.

### *1.1.3 Sex differences in research*

Concerning differences in health and biomedical research, a paper published in 2019 evaluated the occurrence of the sex factor in about 11.5 million contemporary scientific publications (1980-2016) [13]. For this aim, the authors grouped the journals into three disciplines: public health, clinical medicine, and biomedical research. The result of this work was that only 54% of public health studies and 43% of clinical medicine papers contained individuals of both sexes. For biomedical research, the results were more worrisome, as almost 70% of the studies did not indicate the sex of the individuals. This study also concludes that over the last 40 years, there has been an increase in the number of studies that include sex information, especially in the fields of clinical medicine and public health, but in the case of biomedical investigations, sex is still widely underestimated.

As previously discussed, females have historically been excluded from experiments based on alleged data variability caused by their reproductive cycle [14, 15]. However, recent studies indicate that this extra variability in females does not exist. By a meta-analysis of microarray expression data, Becker, Prendergast, and Liang [16] showed that the variability at the gene expression level detected in females was no more significant than in males, neither in mice nor

in human females. The work of Prendergast, Onishi, and Zucker [17] concluded that monitoring the estrogen cycle in rats is unnecessary for neuroscience research. Other comprehensive studies, such as those conducted by McCarthy [18] and by Shansky and Woolley [19], also debunk the myth that females are variable and difficult to study [15].

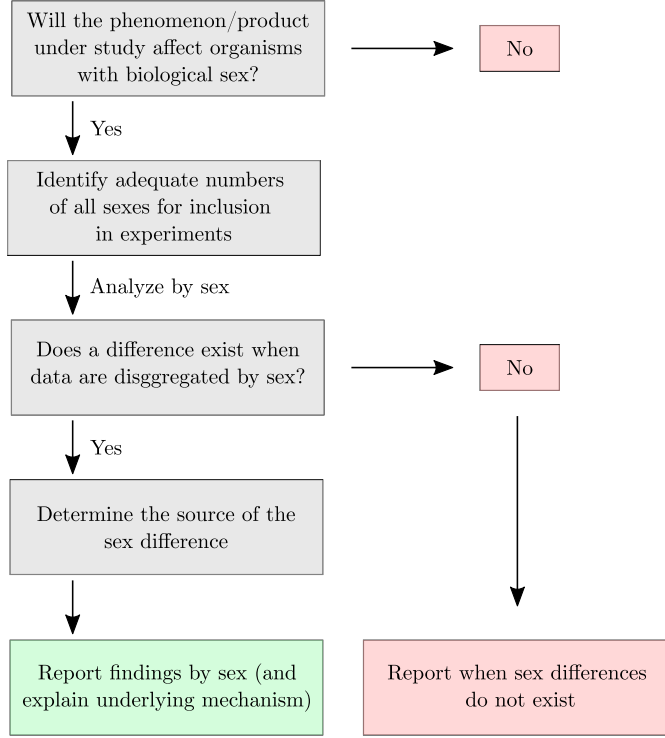
Another argument commonly used to avoid the inclusion of the sex variable in studies is the thought that this requires a doubling of the sample size to be studied. However, an efficient experimental design allows the incorporation of the sex and/or gender variables while maintaining control over the variance. Research indicates that using multifactorial experimental designs can incorporate the sex variable in the analysis with only a 14-33% increase in sample size. [20, 21].

As a general rule, the sex variable should be considered whenever it can influence the result. The paper published by Tannenbaum et al. [14] shows us a decision tree to determine the need to include the sex factor in our analysis and our results (Fig. 1.1).

Incorporating sex differences in research not only solves a clinical problem, but also improves research by increasing reproducibility and experimental efficiency [14]. Pooling data without regard to sex implies a lack of precision and may lead to misinterpretation, masking alterations associated with this factor.

Measures can be taken at several levels to reduce sex-related bias in biomedical research. The first one is funding; institutions can require the introduction of the sex variable in funding proposals; in fact, different agencies such as the National Institutes of Health (US Department of Health and Human Services) and the European Union have recognized this factor as a crucial aspect in personalized medicine and a biological variable necessary for its inclusion in research [22, 23, 24, 25]. In the case of the European Union, the Horizon Europe funding program takes into account that its projects integrate the gender dimension in their research [26]. Scientific publishers can also address the problem, making journals insist on including the sex dimension in published papers as a requirement for ethical and reproducible science [13].

One of the essential objectives of this doctoral thesis, detailed in Chapter 2, is to contribute to reducing these sex-associated biases in research.



**Fig. 1.1:** Decision tree for the incorporation of the sex factor in research studies. Adapted from Tannenbaum et al. [14].

#### 1.1.4 Sex differences in diseases

The existence of diseases that exhibit different behaviors based on sex is now widely recognized. These differences, which include aspects such as prevalence, age of onset, progression, and treatment response, ultimately have their origins at the genetic level. These differences appear at the moment of the formation of a zygote carrying XX or XY chromosomes [3, 27], and this chromosomal difference affects different levels. Within the sex chromosomes, we find genes that are unique to one chromosome, such as the *SRY* gene on the Y chromosome, while others have homologs with slight differences between the X and Y chromosomes (e.g., *ZFX/ZFY* or *UTX/UTY*) [3].

The primary source of differences between females and males is arguably the *SRY* gene, which causes the development of the testes in males [28]. An ini-

tial surge in testosterone levels, which in humans occurs at the end of the first trimester of pregnancy, causes the definitive masculinization of the reproductive organs. These organs are responsible for establishing lifelong sex differences in plasma levels of gonadal steroid hormones (such as testosterone, estradiol, and progesterone). Throughout life, sex hormones act at multiple levels in the differentiation of tissues according to the sex of the individual [3, 27].

On the epigenetic aspect, in human males, genetic imprinting on the X chromosome comes exclusively through the maternal route, while females receive one X chromosome from each of their parents. Moreover, to prevent sex-associated dose differences in this chromosome, females undergo random inactivation of one of the X chromosomes. However, the escape of some genes from this silencing causes these genes to show higher expression levels in females [27]. Therefore, sex-related molecular differences originating from sex chromosomes and the genetic inheritance received from parents are found in all human cells and persist throughout life, independent of sex hormones [3].

The amalgamation of genetic and hormonal factors contributes to the development of distinct biological systems in men and women, resulting in variations in disease susceptibility, presentation, and treatment response. As a result, sex plays a crucial role in influencing physiology and disease through genetic, epigenetic, and hormonal mechanisms.

As mentioned in previous sections, research into sex differences in medicine has significantly advanced in recent decades. At present, we know of multiple diseases in which this phenomenon occurs, among which we could highlight coronary heart disease, chronic lung disease, stroke, neurodegenerative diseases, chronic liver disease, autoimmune diseases, and some types of cancer. Table 1.1 summarizes some of these known sex differences.

This doctoral dissertation explores the molecular basis underlying sex differences in non-alcoholic fatty liver disease (NAFLD) and multiple sclerosis (MS). For this reason, the following paragraphs will provide deeper description of the existing knowledge on sex differences in metabolic disorders, autoimmune and neurodegenerative diseases.

**Table 1.1.** Known sex differences in a selection of diseases. Adapted from Mauvais-Jarvis et al. [3] with additional information from Lonardo et al. [29], Català-Senent et al. [30] and López-Cerdán et al. [31].

| Disease                           | Male sex                                                                                      | Female sex                                                                                         |
|-----------------------------------|-----------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------|
| Alzheimer's disease               | Lower prevalence                                                                              | Higher prevalence; clinical course is faster                                                       |
| Autoimmune disease                | Lower prevalence (in general)                                                                 | Higher prevalence (in general)                                                                     |
| Cancer                            | Higher prevalence and mortality; genetic cell autonomous predisposition                       | Lower prevalence and mortality for some cancers; higher expression of X-encoded tumour suppressors |
| Chronic kidney disease            | More rapid rate of progression; testosterone might be deleterious                             | Higher prevalence; estrogens might be protective                                                   |
| Heart disease                     | Younger age; more obstructive coronary artery disease                                         | Older age; more coronary microvascular dysfunction                                                 |
| Ischaemic stroke                  | Younger age of onset                                                                          | Older age of onset; sex-specific risk factors                                                      |
| Multiple sclerosis                | Worse and faster progression; higher probability of developing progressive multiple sclerosis | More susceptible; diagnosed at an earlier age (on average)                                         |
| Non-alcoholic fatty liver disease | Higher prevalence and severity than in women during the reproductive age                      | Same or higher prevalence than in men, in women after menopause                                    |
| Parkinson's disease               | Greater risk of Parkinson's disease                                                           | Higher mortality rates and require earlier documented placements in nursing care                   |

### *Metabolic disorders*

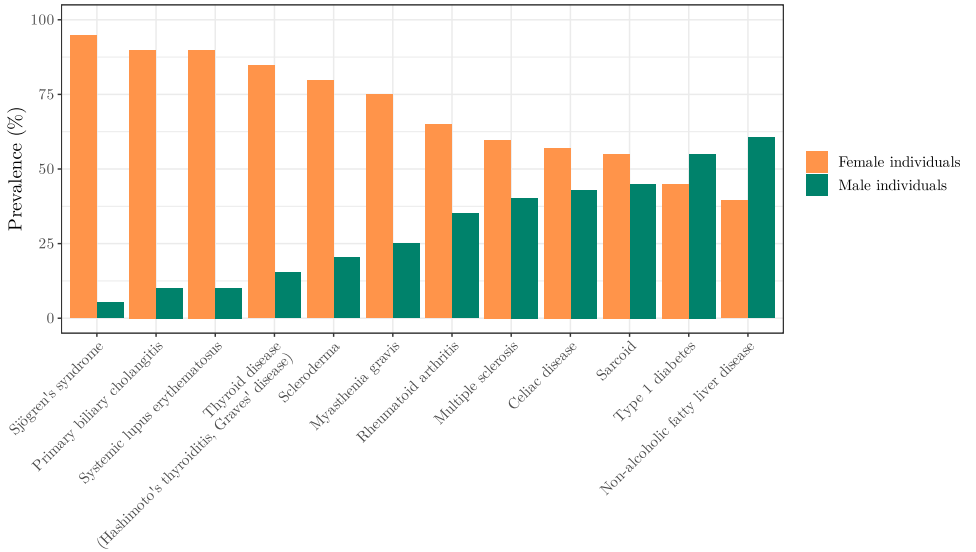
Sex differences in metabolic disorders, such as obesity or hepatic steatosis, have been widely described in mammals and other animals. In general, such disorders are more prevalent in males than in females [32, 33, 34]. Traditionally, these differences have been associated with the role of sex hormones [35, 36]. However, more recent studies suggest that the complement of the X chromosome also plays a role [32, 37].

The liver is a major metabolic organ and is one of the most sexually dimorphic organs [38], resulting in significant sex-related differences in its physiological and pathological processes [39]. An study in mice found more than 6000 differentially expressed genes have been identified between males and females [39].

Adipose tissue, a crucial element in sex differences in metabolic diseases, is more than just a fat storage unit. It functions as an endocrine organ, playing a vital role in regulating metabolism and ensuring energy homeostasis through the secretion of adipokines and cytokines, such as leptin, adiponectin, tumor necrosis factor  $\alpha$  and interleukin 6 [32, 40]. In adult mammals, most adipose tissue is white adipose tissue, either in subcutaneous adipose tissue (SAT) or in the abdominal region as visceral adipose tissue (VAT). The distribution of body fat follows a different pattern according to sex. Premenopausal women store fat in the SAT depots around the hips and thighs, whereas men store fat in the VAT depots in the deep abdominal region [41]. Variations in fat distribution between sexes are partially linked to specific differences in lipoprotein lipase activity within adipose depots, which is crucial in fat accumulation [42]. Additionally, adipose-mediated sex hormone metabolism also contributes in dimorphic adipose distribution. [32]. There is increasing evidence that a more substantial accumulation of VAT, such as that in men, may be associated with an increased risk of metabolic disease [43]. On the other hand, in premenopausal women, their fat distribution may improve glucose metabolism, partially protecting against the development of type 2 diabetes [44]. After menopause, a shift from SAT to VAT in women causes the loss of this protection against metabolic diseases [32, 42].

### Autoimmune diseases

Autoimmune diseases encompass more than 80 chronic disorders that affect about 3-5% of the population [45, 46]. Within this group of alterations, we find diseases such as type 1 diabetes, thyroid disease, systemic lupus erythematosus, rheumatoid arthritis, celiac disease, or multiple sclerosis. For women, the risk of autoimmune diseases is up to four times higher than for men, being one of the main causes of morbidity and mortality in young and middle-aged women (Fig. 1.2) [47, 48]. Several explanations have been proposed for these differences: sex hormones, the X chromosome, environmental factors, microchimerism, and the microbiome. However, the precise mechanisms behind these sex differences have not yet been fully clarified [47].



**Fig. 1.2:** Prevalence of a set of autoimmune or inflammatory diseases in females and males. Image adapted from Mauvais-Jarvis et al. [3], Di Florio et al. [49] and Teng et al. [50].

One hypothesis put forward by Kronzer, Bridges, and Davis [47] proposes that one factor that may explain the sex-related autoimmune bias would be antibodies. Specifically, it posits that women have an evolutionarily conserved tendency towards greater B-cell activation and the production of higher levels of antibodies, resulting in a higher incidence of autoimmune diseases caused by these proteins. This hypothesis would be based on women having more antibodies than men at baseline [51, 52, 53, 54] and in response to infection or vaccination [51, 54]. In addition, higher levels of autoantibodies have also

been observed in women [53]. Kronzer *et al.* propose that this excessive production of antibodies would be an evolutionary response to infections, which have been and continue to be the leading cause of infant mortality [55], since these proteins protect both mothers and their children, via breast milk.

The modulation of the immune system greatly relies on interleukins (IL), which are small proteins essential for the proliferation, maturation, migration, and adhesion of these cells [56, 57]. Some interleukins exhibit pleiotropy, meaning they can have different effects based on their concentrations and the specific context in which they act. Additionally, there are notable sex differences in the levels and mechanisms of action of several of these proteins, including IL-2, IL-6, IL-8, and IL-17 [58, 59, 60, 61]. As key regulators of the immune system, interleukins are vital in the context of autoimmune diseases. For instance, insufficient levels of IL-2 can contribute to the development of autoimmune disorders [62, 63]. Understanding sex differences and the roles of interleukins in modulating immune responses is crucial for learning their impact on numerous diseases.

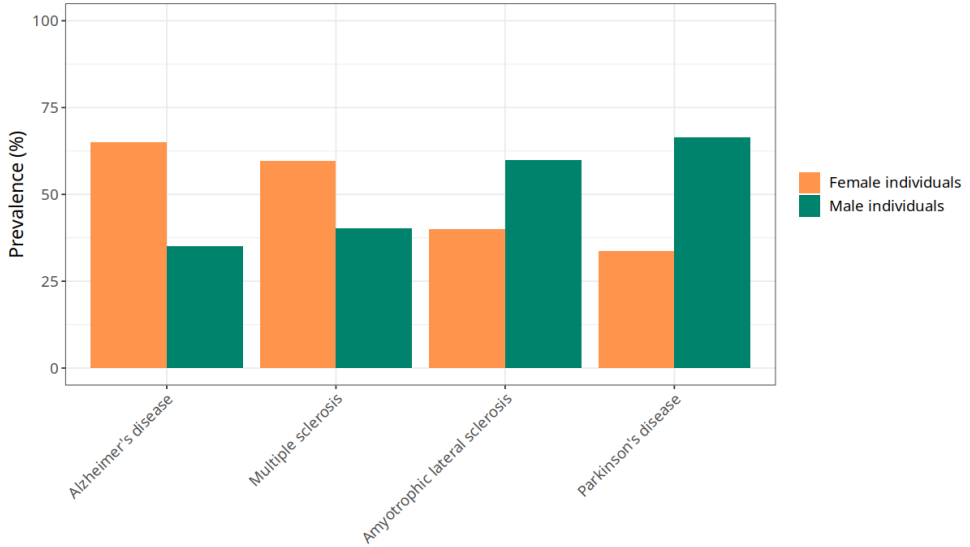
Sex differences in the number and function of lymphocyte populations, which play a pivotal role in the development and progression of autoimmune diseases, are also known [64]. Women have a stronger pro-inflammatory response than men, displaying increased lymphocyte proliferation, activation, and cytokine production. On the contrary, men exhibit an expansion of regulatory cell subsets [64]. Moreover, in addition to the numerous genes that exhibit differential expression in lymphocytes, there are differences associated with the X chromosome. The X chromosome, which is larger than the Y chromosome, contains numerous genes related to immune function [65]. As previously mentioned, some genes (about 15% of them) escape the process of inactivation of one of the X chromosomes that occurs during embryogenesis in females [27, 64, 66]. This escape from complete inactivation results in an overexpression of immune-related genes from the X chromosome in female lymphocytes, which has been associated with autoimmune diseases [64, 66].

### *Neurodegenerative diseases*

Neurodegenerative disorders (NDDs) encompass a group of neuropathological conditions, including Parkinson's disease (PD), Alzheimer's disease (AD), amyotrophic lateral sclerosis (ALS), and multiple sclerosis (MS). These conditions are distinguished by a gradual decline in neuronal functions and eventual cell death [67, 68]. NDDs can have dramatic effects on individuals, leading to memory problems and changes in motor functions that make it difficult to



perform everyday tasks for those affected people. As shown in Fig. 1.3 and Table 1.1, these diseases exhibit sex differences in their onset, prevalence, and development.



**Fig. 1.3:** Prevalence of a set of neurodegenerative diseases in females and males. Image adapted from Mauvais-Jarvis et al. [3], McCombe and Henderson [69], and Alzheimer's association report [70].

The role of the immune system and immune cell interactions in NDDs has been previously described, playing a significant role in the onset and progression of these disorders [67, 71, 72]. Several studies have shown that immune responses, especially innate immune responses, vary between males and females [45, 67, 73].

Besides that, immune and inflammatory reactions in inflammatory neurological diseases, like MS, have been related to the development of these diseases [74]. Neuroinflammation in both inflammatory neurological and neurodegenerative diseases is characterized by the activation of microglia and astrocytes, leading to the secretion of pro-inflammatory cytokines and chemokines that further stimulate immune cell trafficking to the central nervous system [75, 76]. Microglia are non-neuronal cells of the central nervous system (CNS) that play a role in neuronal development, modulation of synaptic transmission, and adult neurogenesis [77, 78]. As the resident immune cells of the brain parenchyma, they are the first line of sentinels against tissue damage or invading pathogens,

acting as a key element between the nervous and immune systems [78, 79, 80]. Sex-associated molecular alterations in microglia have been previously described [72, 81] and may play a role in the sex differences in these diseases [82].

## 1.2 Molecular biology and omic sciences

Molecular biology is a branch of biology that explores the structure, function, and interactions of biological molecules within living organisms at the molecular level. It delves into the study of DNA, RNA, proteins, and other essential macromolecules, unraveling the intricate mechanisms that govern life processes.

Deciphering the DNA double helix structure in 1953 [83, 84] laid the foundation for molecular biology, providing a fundamental framework for understanding genetic information transfer. This field has witnessed remarkable advancements in recent decades, driven by rapid technological progress and groundbreaking discoveries.

In the ensuing years, the development of techniques such as polymerase chain reaction (PCR) [85], DNA sequencing [86, 87], and recombinant DNA technology [88] revolutionized molecular biology, enabling scientists to manipulate and analyze genetic material with unprecedented precision. The advent of these tools facilitated the exploration of gene expression, regulation, and the identification of specific genes associated with various diseases.

Recent decades have also seen the emergence of innovative technologies, such as CRISPR-Cas9 gene editing [89, 90], which allows for precise modification of genes with unprecedented accuracy. This breakthrough has opened up new avenues for therapeutic applications, as well as a better understanding of the genetic basis of diseases.

In summary, the field of molecular biology has undergone a transformative journey in recent decades, driven by technological innovation and a growing body of knowledge. This progress has not only deepened our understanding of the molecular mechanisms underpinning life but has also opened up exciting possibilities for medical applications and the development of novel therapeutic interventions.

### 1.2.1 *Omic sciences*

Omic sciences, also known as "omics", collectively refer to a set of interdisciplinary fields that investigate and analyze large-scale biological data at various molecular levels. These fields include genomics, transcriptomics, proteomics, metabolomics, and other "omics" branches, each focusing on a specific aspect of biological molecules. Overall, the goal of omics sciences is to identify, characterize, and quantify all biological molecules that are involved in the structure, function, and dynamics of a cell, tissue, or organism [91].

The development of omic sciences has been a transformative force in modern biology, revolutionizing our understanding of the complexity of living organisms. Genomics, which involves the study of an organism's entire genome, marked a significant breakthrough with the completion of the Human Genome Project in 2003 [92, 93]. This monumental effort provided a complete map of the human genetic code and laid the groundwork for subsequent omics disciplines. Transcriptomics, also known as expression profiling, is the study of the expression levels of all species of transcript in a given cell population [94, 95], shedding light on gene expression patterns and regulatory mechanisms. Proteomics delves into the study of all proteins within a biological system, offering insights into their functions, interactions, and modifications [96]. Metabolomics focuses on the comprehensive analysis of small molecules involved in cellular processes, providing a snapshot of the metabolic state of an organism [97]. In addition to those already mentioned, the list of omic sciences is vast and includes other disciplines such as metagenomics or lipidomics [98, 99, 100]. In recent years, thanks to the lower costs of the techniques, the number of studies that approach their research from a multi-omic perspective is increasing, obtaining an integrative vision of the phenomena that occur in biological systems and driving innovations that can positively impact human health [101].

To achieve the objectives of this thesis in uncovering the molecular basis of non-alcoholic fatty liver disease and multiple sclerosis, we have utilized transcriptomics data.

### 1.2.2 Transcriptomics

Transcriptomics is the branch of molecular biology that focuses on the study of the transcriptome, which is the complete set of RNA molecules, including mRNA, non-coding RNA, and other RNA species, produced by the cells of an organism [95]. The main goal of transcriptomics is to understand the expression patterns of genes and how they contribute to the functioning of cells and organisms. In medicine, transcriptomics studies contribute to identifying disease biomarkers, understanding disease mechanisms, and developing targeted therapies.

#### *Development of transcriptomics*

Like other omics, transcriptomics has undergone a rapid development related to improvements in techniques and lower costs.

The first approaches to transcript profiling date back to the 1980s, using Sanger sequencing to sequence individual transcripts previously copied to cDNA and termed expressed sequence tags (ESTs) [102, 103, 104]. The term "transcriptome" was first used in the mid-1990s [104, 105] with the emergence of two new methods for transcript characterization:

- **Serial analysis of gene expression (SAGE)** [106] was developed from the EST methodology to enhance the efficiency of tag generation and enable some measurement of transcript abundance. cDNA is generated from the RNA but is digested into 11 bp "tag" fragments using restriction enzymes that cut at a specific sequence and 11 base pairs along from that sequence. These cDNA tags are subsequently concatenated head-to-tail into long strands (>500 bp) and sequenced. After deconvoluting the sequences into their original tags, if a reference genome is accessible, these tags can sometimes be aligned to determine their corresponding genes [107]. In cases where a reference genome is not available, the tags can be directly utilized as diagnostic markers if they are found to exhibit differential expression [107]. Over time, SAGE variations have been developed to enhance the accuracy of tag identification by extending the length of tags [108].
- **DNA microarrays** [109], also called DNA chips, are based on quantifying the hybridization of labeled cDNA strands to probes fixed on a solid substrate. The fluorescence intensity at each probe location on the array indicates the transcript abundance for that probe sequence [104, 108].

Due to its low cost and high throughput, this technique became widely used in the 2000s [108]. This technique has a significant limitation since, unlike SAGE, it requires prior knowledge of the transcripts to design the probes.

High-throughput sequencing, emerging in the early 2000s, enabled the development of **RNA-Seq**. This technique, involving the sequencing of cDNA transcripts, overcomes limitations of previous technologies by enabling both transcript discovery and quantification. RNA-Seq allows assembly against reference genomes or *de novo* and is highly effective in detecting alternative splicing events [108]. Additionally, it permits the quantification of RNAs over a wider dynamic range and with a sensitivity of two to three orders of magnitude higher than microarrays [110]. These advantages turned RNA-Seq into the most popular transcript profiling approach over the last decade.

In recent years, transcriptomic techniques have significantly improved. Long-read sequencing technologies have enabled the study of individual transcripts, circumventing previous limitations in transcript reconstruction from short reads and working with genes with a large number of isoforms [108, 111]. The development of **single-cell RNA sequencing** (scRNA-Seq) in 2009 made it possible to understand the transcriptome at the cellular level [112]. More recently, **spatial transcriptomics** has emerged [113], overcoming restrictions of scRNA-Seq, such as the loss of cell positional information, allowing us to acquire transcriptomic data from intact tissue sections and, in turn, obtain information on spatial distribution and elucidate cellular interaction patterns [114].

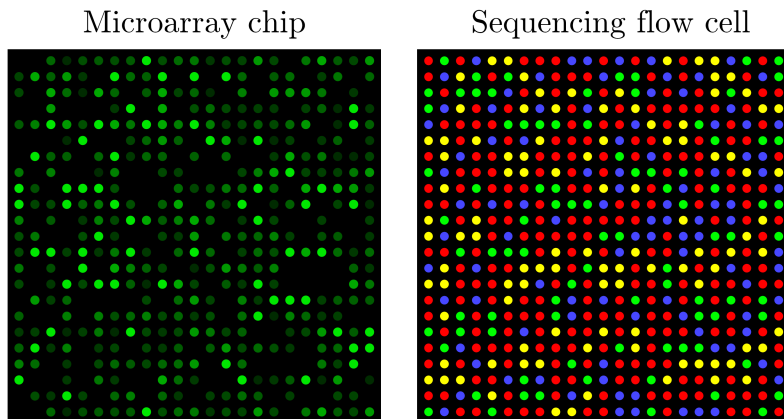
### *Analysis of transcriptomics data*

Before conducting transcriptomic analysis, it is essential to preprocess the collected data, which will differ based on the technology employed. This step aims to get a matrix containing the quantification of the biological features of interest (genes, transcripts, exons, etc.) for each sample. The preprocessing process for the most common transcriptomic technologies, microarrays and RNA-Seq, is briefly explained below.

As a result of a microarray analysis, we obtain an image quantifying the fluorescence of each spot on the chip (Fig. 1.4). The fluorescence intensity indicates the abundance of the probe associated with each microarray position. These probes are connected to the biological features, which allows us to generate a matrix that quantifies the abundance of these genes, transcripts, etc. Addition-

ally, possible artifacts must be identified and eliminated from further analysis [104, 115].

RNA-Seq data are also obtained from images, but the process of converting these to sequences is usually managed by the sequencer itself. For example, Illumina’s sequencing-by-synthesis method results in an array of clusters distributed over the surface of a flow cell. The flow cell is recorded each sequencing cycle, with tens or hundreds of cycles in total (Fig. 1.4). The color in each cycle indicates the nucleotide present at that moment in each of the clusters. A demultiplexing process converts this series of images into sequences of nucleotides, which, after quality control, are usually aligned against a reference genome (or *de novo* if no such genome is available) [104]. Finally, quantifying the aligned sequences at the level of biological features generates the count matrix needed for subsequent analyses.



**Fig. 1.4:** Microarrays and RNA-Seq are based on image analysis in different ways. In a microarray chip, each spot corresponds to a defined oligonucleotide probe, and the abundance of a specific sequence is directly detected through fluorescence intensity (Affymetrix, Santa Clara, CA). Conversely, in a high-throughput sequencing flow cell, spots are sequenced one nucleotide at a time, with the color at each round indicating the next nucleotide in the sequence (Illumina Hiseq, San Diego, CA). Variations of these techniques may involve the use of more or fewer color channels. Figure from Lowe et al. [104], published under Creative Commons Attribution License.

Among the range of analyses that can be performed with transcriptomic data, differential expression analysis is one of the most common, if not the most common. Differential expression analysis is the methodology used in this thesis to elucidate sex-related differences in NAFLD and MS, since this type of analy-

sis allows the comparison of abundances at the transcriptomic level between individuals from different groups.

The differential expression analysis method [111, 116, 117] is conditioned by technology, type of data, and number of samples, among other factors. Most of these methods require normalization of the data matrix to perform comparisons correctly. After conducting the differential expression analyses, we obtain for each biological element analyzed (genes in the case of this thesis) a measure of the magnitude of the change (logFC, in our case), a measure of variability, a p-value that will be adjusted later, and method-specific statistics. These results are the necessary starting point for subsequent functional analyses or differential expression meta-analyses.

Functional profiling gives us a more integrative view of what is happening at the molecular level. This is achieved by analyzing sets of elements (genes, in our case) associated with each functional term. There are multiple databases that associate genes with functional terms, some of the most popular being Gene Ontology (GO) [118], the Kyoto Encyclopedia of Genes and Genomes (KEGG) [119], and Reactome [120]. Multiple functional profiling methods are available to combine these genomic results from experiments and biological information from databases. The choice of method will depend on the source data, the experimental design, and the biological question to be answered. Two different functional profiling methods are used throughout this thesis: the multidimensional gene-set analysis (MD-GSA) method proposed by Montaner and Dopazo [121] and the Hipathia pathway activity analysis method developed by Hidalgo et al. [122].

### 1.3 Systematic review and meta-analysis

The amount of available omics datasets is currently increasing at a vertiginous rate. This can be seen in databases such as Gene Expression Omnibus (GEO), an international public repository that archives gene expression and epigenomics data sets, doubling the number of studies stored in it every five years [123]. Many reasons can explain this increase in the amount of information in repositories; among them are the lower costs of technology and the growing commitment to a more open science.

Open science is a movement that promotes collaboration, transparency, and access to scientific research [124, 125, 126]. One of the key aspects of open science is the FAIR data principles, which aim to make research data Findable, Accessible, Interoperable, and Reusable [127]. Usually, the omics generated for

a study are only used to answer the biological questions posed in the investigation for which they were obtained. However, this information can often be squeezed to answer other biological questions. Therefore, to get the most out of the investment in science, scientific funding by public institutions increasingly demands that the results obtained be published under the principles of open science [124, 128, 129, 130]. Systematic review and meta-analysis processes provide the opportunity to reanalyze these open data, helping to optimize resources and enabling the advancement of science by extracting new knowledge from them [131].

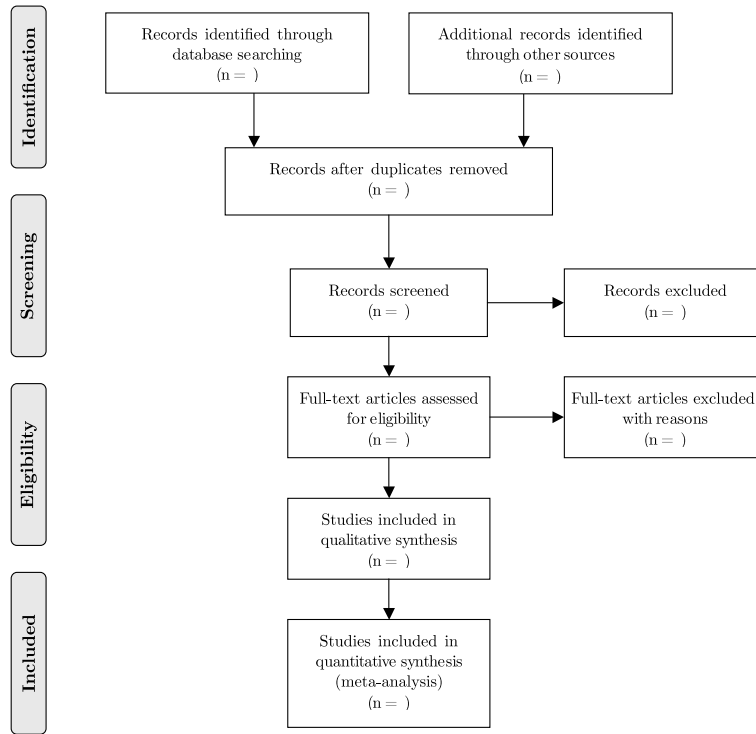
### *1.3.1 Systematic review*

The systematic review methodology allows the synthesis of previous knowledge to answer a research question. This process should be carried out transparently and reproducibly, trying to include previous evidence on the subject and assessing the evidence's quality to reduce possible biases [132, 133, 134].

The history of systematic reviews can be traced back to the 18th century, with the first recognized modern systematic review being a 1753 publication examining all previous scurvy publications. [135, 136]. Several papers promoting or employing systematic approaches can be located in the initial decades of the 20th century; it was primarily during the 1970s that the field of research synthesis started to emerge prominently due to the imperative to employ explicit, transparent, and rigorous methodologies to support the credibility of reviews [137].

In order to ensure reliable quality standards for systematic reviews, the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) statement was published in 2009 [132, 138]. This text, recently updated in 2021 as PRISMA 2020 [134, 139], contains the requirements and steps for a rigorous systematic review. Fig. 1.5 shows the flow of information through the different phases of a systematic review following PRISMA 2009 since the systematic reviews carried out in this thesis were performed prior to the publication of PRISMA 2020.





**Fig. 1.5:** Flow of information through the different phases of a systematic review following the PRISMA 2009 statement. Adapted from Moher et al. [138].

### 1.3.2 Meta-analysis

Clinical research aims to obtain precise and quality results that can be used as a basis for clinical practice. However, clinical studies only sometimes achieve this goal and occasionally present results that differ from each other [140]. One way to address this problem is through meta-analysis, a statistical methodology that allows a combined result to be obtained from the results of individual independent studies previously selected in a rigorous systematic review process [141, 142, 143].

The use of meta-analysis began in social sciences and psychology as a solution to the irregularities that arose in individual studies due to the different contexts, types of subjects, and other factors with which these studies were

conducted [144]. Over time, the use of this methodology spread to different areas of science, including medicine and genomics [142, 143].

Among the advantages of meta-analyses, a key benefit is the greater statistical power in obtaining the combined effect of the individual studies and quantifying their variability [141, 142]. Another significant advantage is the ability to answer questions not asked in the original studies, thereby maximizing scientific resources [141, 143]. However, it is important to note that this methodology also has limitations. The most crucial of these is the need for appropriate statistical methods and, equally importantly, the need for careful interpretation of results, especially when there are discrepancies between the results of individual studies [141]. This caution is essential to ensure the reliability and validity of the meta-analysis.

### *Meta-analysis models*

There are two main meta-analysis models: the fixed-effects model and the random-effects model [141, 143, 145, 146].

- The fixed-effects model assumes that a common effect underlies all studies and, as a result, all variance in observed effect sizes is variability due to random sampling error (within-study variability).
- The random-effects model assumes that study effect sizes vary from one study to the next. Under this model, variance in observed effect sizes is attributable to both sampling error (within-study variance) and statistical heterogeneity (between-study variance).

Choosing one or the other model is a crucial decision when performing a meta-analysis [145, 146]. The random-effects model tends to be considered more realistic since it does not require zero variability between studies. Furthermore, its results can be generalized to a broader and more diverse population [141]. Due to the heterogeneity among the studies analyzed, the random-effects model was used in this doctoral thesis.

Appendix I provides a detailed explanation of the interpretation of the results obtained from the meta-analyses.

## Chapter 2

# Objectives

As previously exposed, historically, there has been a bias in medicine, research, and health in favor of males. The consequences of this bias include misdiagnosis and inadequate treatment dosages, among others, in women. In recent decades, an effort has emerged in the scientific community to address this bias. Within this context, this doctoral thesis seeks to use widely used methods such as systematic review and meta-analysis to contribute to this bias reduction.

Therefore, the **main objective** of this doctoral thesis is to study the molecular bases that affect men and women differently in non-alcoholic fatty liver disease and multiple sclerosis through meta-analysis of transcriptomic studies.

In order to achieve this goal, the **specific objectives** of this study are as follows:

- **Objective 1:** To study the molecular basis differentially affected between premenopausal women and men in the development of non-acholic fatty liver disease, specifically in the transition from NAFL to NASH stage (**Chapter 3**):
  - Perform a systematic review of the studies available in public databases on NAFLD, including the factors: sex, age, and stage of the disease.

- Conduct differential expression analysis of genes from selected studies, together with their subsequent functional enrichment and meta-analysis of the results of the latter.
- Carry out a signaling pathways analysis from selected studies and their subsequent functional enrichment, as well as a meta-analysis of the latter results.
- **Objective 2:** To identify molecular mechanisms underlying observed MS sex differences that may guide novel therapeutic approaches tailored for males or females (**Chapter 4**):
  - Accomplish a systematic review of studies on multiple sclerosis, including information on patient sex, in publicly available databases.
  - Conduct differential gene expression analyses and subsequent meta-analyses to identify gene alterations between healthy individuals and MS patients, distinguishing by sex and considering potential tissue origins of the samples.
  - Perform functional enrichment at the tissue level to characterize sex differences found in the genes.

## Chapter 3

# Hepatic steatosis and steatohepatitis: a functional meta-analysis of sex-based differences in transcriptomic studies

José F. Català-Senent<sup>1,2</sup>, Marta R. Hidalgo<sup>1</sup>, Marina Berenguer<sup>3,4,5,6</sup>,  
Gopanandan Parthasarathy<sup>7</sup>, Harmeet Malhi<sup>7</sup>, Pablo Malmierca-Merlo<sup>1,8</sup>,  
María de la Iglesia-Vayá<sup>9</sup> and Francisco García-García<sup>1,2</sup>

Published in **Biology of Sex Differences**

[DOI: 10.1186/s13293-021-00368-1]



### 3.1 Abstract

**Background:** Previous studies have described sex-based differences in the epidemiological and clinical patterns of non-alcoholic fatty liver disease (NAFLD); however, we understand relatively little regarding the underlying molecular mechanisms. Herein, we present the first systematic review and meta-analysis of NAFLD transcriptomic studies to identify sex-based differences in the molecular mechanisms involved during the steatosis (NAFL) and steatohepatitis (NASH) stages of the disease.

**Methods:** Transcriptomic studies in the Gene Expression Omnibus database were systematically reviewed following the PRISMA statement guidelines. For each study, NAFL and NASH in premenopausal women and men were compared using a dual strategy: gene-set analysis and pathway activity analysis. Finally, the functional results of all studies were integrated into a meta-analysis.

**Results:** We reviewed a total of 114 abstracts and analyzed seven studies that included 323 eligible patients. The meta-analyses identified significantly altered molecular mechanisms between premenopausal women and men, including the overrepresentation of genes associated with DNA regulation, vinculin binding, interleukin-2 responses, negative regulation of neuronal death, and the transport of ions and cations in premenopausal women. In men, we discovered the overrepresentation of genes associated with the negative regulation of interleukin-6 and the establishment of planar polarity involved in neural tube closure.

**Conclusions:** Our meta-analysis of transcriptomic data provides a powerful approach to identify sex-based differences in NAFLD. We detected differences in relevant biological functions and molecular terms between premenopausal women and men. Differences in immune responsiveness between men and premenopausal women with NAFLD suggest that women possess a more immune tolerant milieu, while men display an impaired liver regenerative response.

**Keywords:** Non-alcoholic fatty liver disease, Sex characteristics, Precision medicine, Computational biology, Transcriptome profiling





## 3.2 Background

Non-alcoholic fatty liver disease (NAFLD) encompasses a spectrum of liver disorders—ranging from fat accumulation in hepatocytes (NAFL) to non-alcoholic steatohepatitis (NASH)—which can lead to the development of cirrhosis or liver cancer. While steatosis characterizes NAFL, NASH exhibits the additional histological features of inflammation, ballooning, and fibrosis. Awareness of NAFLD has recently increased due to its worldwide impact on health; indeed, NAFLD now represents the most common liver disease in developed countries [1]. While the estimated worldwide prevalence of NAFLD is around 25%, it is more common in South America and the Middle East and less prevalent in Africa [2].

Of note, there exists significant interindividual heterogeneity regarding NAFLD progression, which has a considerable impact on the clinical consequences to the individual. Patients in the initial stage of the disease (NAFL) display a low risk of adverse outcomes [3]; however, progression to NASH increases the possibility of both hepatic and extrahepatic complications.

Interestingly, NAFLD has a higher prevalence in men and postmenopausal women than premenopausal women [4]. Indeed, the liver shows the second-largest amount of sexual dimorphism in humans [5], with both physiological and pathological hepatic processes, such as the detoxifying metabolism of cholesterol and the prevalence of hepatic diseases, respectively, differing between men and women [6]. Therefore, patient sex will likely affect NAFLD progression and treatment outcomes [7]; however, most NAFLD studies do not consider differences between sexes in their analysis [8]. We sought to explore the molecular mechanisms underpinning the likely existence of sex-related differences in NAFLD patients for the above-mentioned reasons.

In this context, we carried out the first (to the best of our knowledge) systematic review and meta-analysis of transcriptomic studies in NAFLD. We selected studies that had deposited data in the Gene Expression Omnibus (GEO) datasets database [9] to guarantee the eligibility of selected studies. We used these analyses to clarify the molecular mechanisms underpinning differences in NAFL and NASH in premenopausal women and men (postmenopausal women were not included in the analysis due to the overall low number of cases). Overall, sex-based differences may have important clinical implications for patients, and this knowledge may help to develop personalized approaches to NAFLD management.

### 3.3 Materials and methods

#### 3.3.1 Literature review

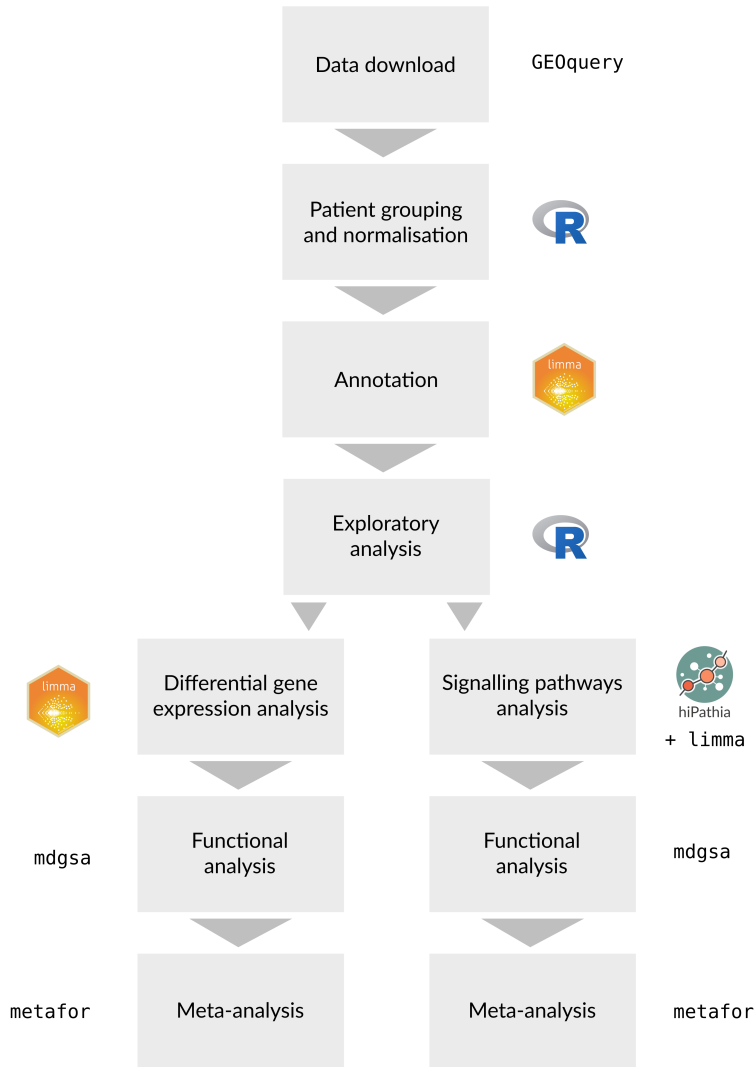
A review was conducted in June 2020 following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement guidelines [10]. The GEO datasets database was searched [9] using the keywords “NAFLD”, “NAFL”, “steatosis”, “NASH”, and “steatohepatitis” for transcriptomic studies published in English.

#### 3.3.2 Study exclusion criteria

The following exclusion criteria were applied: (i) studies conducted in organisms other than humans, (ii) studies without information regarding the sex of participants or that did not include both sexes, (iii) studies without individuals from the NAFL and NASH stages of the disease, and (iv) studies in which the disease had not been diagnosed with a biopsy. In the latter case, the requirement for NAFL and NASH diagnoses by biopsy was added as less invasive methods (such as conventional imaging techniques) cannot accurately detect mild steatosis or differentiate NAFL from NASH [11]. This step aimed to minimize the input of false-positive data into the statistical analyses caused by an incorrect disease stage classification.

#### 3.3.3 Bioinformatics analysis

The same three-step transcriptomic analysis strategy was applied to each separate study: (i) data acquisition and preprocessing, (ii) differential gene expression and functional enrichment analysis, and (iii) differential pathway activation and functional analysis. The functional results of all studies were integrated using meta-analytical techniques (Fig. 3.1). Bioinformatics analysis was carried out using the programming language R 3.6.0 [12]; information regarding the packages used and their version numbers are provided in the supplementary material.



**Fig. 3.1.** Data Analysis Workflow

### 3.3.4 Data acquisition and preprocessing

During data preprocessing, the nomenclature of the distinct disease stages (by grouping patients into Control, NAFL, and NASH groups) and the association of the probe identifiers with their corresponding genes were standardized. In

those studies containing treated and untreated patients, treated patients were eliminated from further analysis. For studies that used microarrays, probe codes were transformed to their respective Entrez identifiers from the NCBI database. For repeated probes, the median of their expression values was calculated. When available, data normalized by the original authors of the studies were used. Otherwise (studies GSE61260 and GSE66676), raw data were normalized using the Robust Multichip Average (RMA) algorithm [13]. The gene nomenclature in RNA-Seq studies was also standardized to the Entrez identifiers, and the raw data matrix was processed using the TMM standardization method [14] followed by a log transformation of the data.

Women were separated into premenopausal and post-menopausal groups to meet the objective of this work. As selected studies failed to include this specific information, women aged under 50 years were assumed as premenopausal based on studies indicating an average age of menopause at around 48–52 years [15, 16].

After data normalization, the detection of possible anomalous effects within the studies was carried out by completing an exploratory clustering and a principal component analysis (PCA).

### ***3.3.5 Differential gene expression and functional enrichment analysis***

Differential expression analyses of selected studies were carried out using the *limma* and *edgeR* packages [17, 18] to detect differentially expressed genes when comparing NAFL and NASH in premenopausal women versus men. A linear model was adjusted for each gene, which included possible batch effects, contrasting (NASH.W – NAFL.W) and (NASH.M – NAFL.M), where NASH.W, NAFL.W, NASH.M, and NAFL.M corresponded to NASH-affected premenopausal women, NAFL-affected premenopausal women, NASH-affected men, and NAFL-affected men, respectively. Statistics regarding the differential expression were calculated and *p* values adjusted using the Benjamini & Hochberg (BH) method [19].

Based on the differential gene expression results, functional enrichment analysis was then performed using gene set analysis (GSA) [20]. First, genes were ordered according to their *p* value and the sign of the contrast statistic. Next, GSA was performed using the logistic regression model implemented in the *mdgsa* R package [21] along with their corresponding functional annotations

obtained from the Gene Ontology (GO) [22] and Kyoto Encyclopedia of Genes and Genomes (KEGG) PATHWAY [23] databases.

Due to their hierarchical structure, the gene annotations with GO terms applied in the *mdgsa* package were propagated to inherit the annotations of the ancestor terms. Excessively specific or generic annotations (blocks smaller than ten or larger than 500 words) were subsequently filtered out. Finally, functions with a BH-adjusted  $p$  value under 0.05 were considered significant.

For the two functional elements (GO terms and KEGG paths), the number of overrepresented elements shared by the studies were analyzed. These results were graphically represented as UpSet plots [24] to demonstrate the common elements between the different sets. First, the overrepresented elements were compared as separate graphs to detect the common functions in each group: GO terms in premenopausal women, GO terms in men, KEGG pathways in premenopausal women, and KEGG pathways in men. Next, any significant GO terms in each study were visualized in the same plot to highlight significant terms with different signs among the studies.

### 3.3.6 Differential pathway activation and functional analysis

The *hipathia* algorithm [25] was used to perform Pathway Activity Analysis (or PAA). This method transforms gene expression values for stimulus-response signaling sub-pathways into activation levels that ultimately trigger cellular responses. In this current study, the *hipathia* R package was used to analyze 1654 GO terms and 142 UniProt functions [26] associated with 146 KEGG routes [23]. The activation signal of each sub-pathway was computed from gene expression values using this algorithm. These values were used to detect significant differential activations in the (NASH.W – NAFL.W) – (NASH.M – NAFL.M) contrast pair (as defined above). The functional annotations included in *hipathia* and the differential activation results were integrated using the GSA method to identify differences at the functional level (GO terms and UniProt functions) in relation to NAFLD stages.

### 3.3.7 Meta-analysis

Once the gene functional analysis was applied to each study, a functional meta-analysis was carried out to summarize the results. Similarly, once the pathway functional analysis was applied to each study, the results were summarized via a homologous functional meta-analysis.

Both meta-analyses were performed following the methodology described in detail in García-García [27]. Briefly, the *metafor* R package [28] was used to evaluate the combined effect within a random-effects model, which more precisely detects over-represented elements when compared to performing studies individually at higher statistical power. Likewise, variability in the individual studies was considered in the global estimation of the measured effect so that less-variable results had a higher weight in the overall calculation of the logarithm of the odds ratio (LOR) [29]. Incorporating the variability between experiments in the model of random effects provides more statistically robust results and better integration of selected studies. Finally, the suitability of this selection was confirmed by evaluating the heterogeneity of each study in the global model and the application of cross-validation techniques.

For each meta-analysis and function,  $p$  values (adjusted using the BH method), LORs, and 95% confidence intervals (CIs) were calculated. A term was considered significant when its adjusted  $p$  value was lower than 0.05; a positive LOR in significant functions indicated a greater overrepresentation in premenopausal women than in men. In contrast, a negative LOR indicated higher representation in men than in premenopausal women. Finally, funnel plots and forest plots were used to assess the variability and measure each study's contribution to the meta-analysis.

A total of 8223 elements were analyzed in the gene functional meta-analysis (7994 GO terms and 229 KEGG pathways); 1654 GO terms and 142 UniProt functions [26] were assessed in the functional pathway meta-analysis in association with 146 KEGG pathways [23].

### 3.3.8 Availability of data and materials

The large volume of data and results generated in this study is freely available through the metafun-NAFLD platform: <https://bioinfo.cipf.es/metafun-NAFLD>. Access allows any user to analyze the results shown in this manuscript and review other results that may be of interest. The front-end was developed using the Bootstrap library. All graphics used in this tool were implemented

with Plot.ly except for the exploratory analysis cluster plot, which was generated with the *ggplot2* package.

In this study, transcriptomic data from the Gene Expression Omnibus database with accession numbers GSE48452, GSE61260, GSE66676, GSE83452, GSE89632, GSE126848, and GSE130970 were analyzed.

## 3.4 Results

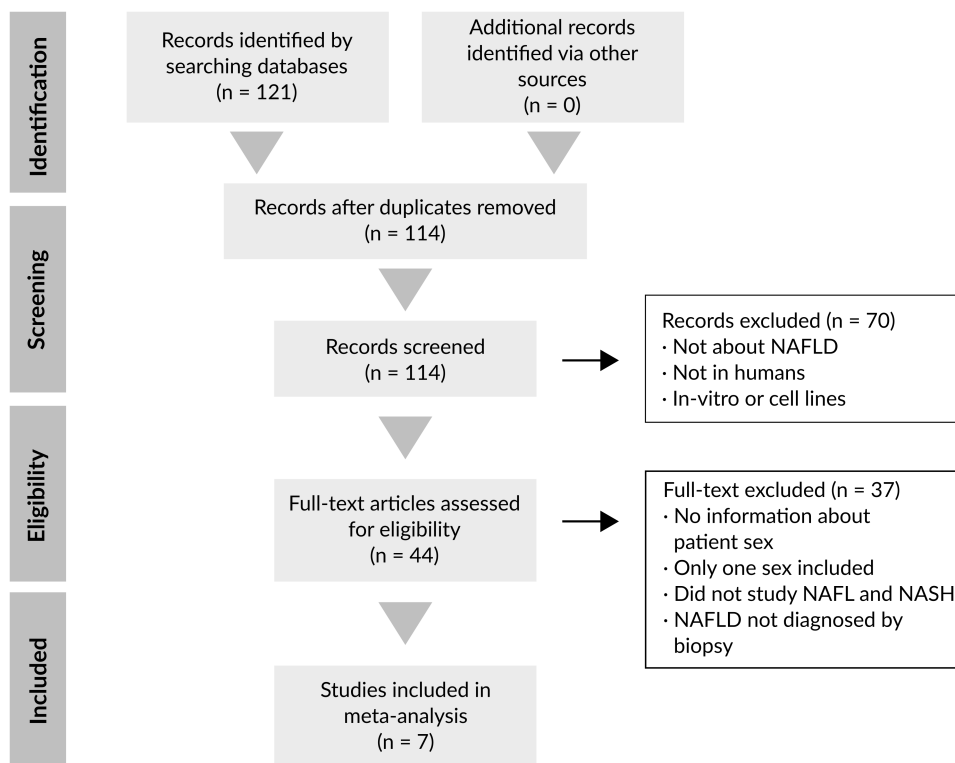
### 3.4.1 Systematic review and exploratory analysis

As shown in the PRISMA flow diagram (Fig. 3.2), our systematic review yielded 114 non-duplicated studies. After applying the previously described exclusion criteria, we employed a total of seven studies in our overall analysis. We excluded studies for the following reasons: studies not in humans or unrelated to NAFLD ( $n = 55$ ), in vitro studies ( $n = 15$ ), lack of information regarding sex or inclusion of only one sex ( $n = 24$ ), no patients in NAFL and NASH disease stages ( $n = 12$ ), and diagnosis not based on histology ( $n = 1$ ). The strict nature of these inclusion and exclusion criteria allowed the selection of comparable studies to ensure the reliability of the subsequent analytical strategies.

The seven selected studies [30, 31, 32, 33, 34, 35, 36] included a total of 323 eligible patients (Table 3.1); 164 individuals were men and 159 women under the age of 50 (49.2% men and 50.8% women, respectively). By disease stage, 148 belonged to the NAFL group and 175 to the NASH group (45.8 and 54.2%, respectively).

As previously indicated, we only used data from non-treated individuals in study GSE83452. (Table 3.1) demonstrates the number of individuals from each study eligible for inclusion based on our aforementioned criteria. In an exploratory analysis, PCA of the GSE83452 study demonstrated that samples clustered into two groups. Further analysis of this study identified technical factors as the source of this grouping rather than any other clinical characteristic, as samples grouped according to their order (the first 64 samples grouped into one cluster and the last 88 into another). We considered this a batch effect, which we accounted for in our subsequent analyses.





**Fig. 3.2.** Flow diagram of our systematic review of the literature and selection of studies for meta-analysis, according to the PRISMA statement guidelines.

**Table 3.1.** Characteristics of the studies selected for inclusion in our analysis in terms of disease stage, sex, BMI, and patient age

| GEO accession,<br>Authors, Year [Ref.]        | Patients by disease stage and sex |      |      |      | Age   |                   | BMI   |                |
|-----------------------------------------------|-----------------------------------|------|------|------|-------|-------------------|-------|----------------|
|                                               | Premenopausal women               |      | Men  |      | Mean  | SD                | Mean  | SD             |
|                                               | NAFL                              | NASH | NAFL | NASH |       |                   |       |                |
| GSE48452 /<br>Ahrens M. et al. 2013 [30]      | 10                                | 9    | 4    | 4    | 41.62 | ± 9.35            | 47.18 | ± 11.34        |
| GSE61260 /<br>Horvath S. et al. 2014 [31]     | 10                                | 7    | 12   | 12   | 41.51 | ± 9.09            | 51.55 | ± 10.40        |
| GSE66676 /<br>Xanthakos S.A. et al. 2015 [32] | 20                                | 4    | 6    | 3    | 17.14 | ± 1.66            | 53.55 | ± 8.86         |
| GSE83452 /<br>Lefebvre P. et al. 2017 [33]    | 32                                | 35   | 6    | 46   | 39.61 | ± 11.51           | N/D   | (>25 or obese) |
| GSE89632 /<br>Arendt B.M. et al. 2015 [34]    | 2                                 | 4    | 14   | 9    | 39.55 | ± 8.19            | 30.21 | ± 5.51         |
| GSE126848 /<br>Suppli M.P. et al. 2019 [35]   | 6                                 | 4    | 9    | 12   | 40    | ± 13 <sup>a</sup> | N/D   | (>30)          |
| GSE130970 /<br>Hoang S.A. et al. 2019 [36]    | 4                                 | 12   | 13   | 14   | 44.72 | ± 9.80            | N/D   | N/D            |

<sup>a</sup> The information corresponding to the patients in this study was kindly provided by the author of the study, although individual patient ages were not available.

### 3.4.2 Individual analysis

The functional analysis of individual genes highlighted significant results over-represented in both the groups of pre-menopausal women and men (summarized in Table 3.2). Analysis (using UpSet plots) of the common terms or pathways in the different studies demonstrated no common element in more than four studies (Supplementary Fig. 1); in fact, we discovered opposite results for some elements in different studies (Fig. 3.3a). The functional analysis of individual pathways only exhibited significant results for the GSE83452 study (20 sub-paths and 96 GO terms).

**Table 3.2.** Number of significant GO terms and KEGG pathways in individual studies after gene functional analysis. Positive and negative logarithms of the odds ratio (LORs) demonstrate overrepresentation in premenopausal women and men, respectively

|                  | GO terms                     |                       | KEGG pathways                |                       |
|------------------|------------------------------|-----------------------|------------------------------|-----------------------|
|                  | Positive LOR<br>(Pre. women) | Negative LOR<br>(Men) | Positive LOR<br>(Pre. women) | Negative LOR<br>(Men) |
| <b>GSE48452</b>  | 166                          | 642                   | 5                            | 22                    |
| <b>GSE61260</b>  | 249                          | 35                    | 5                            | 6                     |
| <b>GSE66676</b>  | 721                          | 168                   | 30                           | 2                     |
| <b>GSE83452</b>  | 90                           | 931                   | 2                            | 25                    |
| <b>GSE89632</b>  | 52                           | 67                    | 2                            | 1                     |
| <b>GSE126848</b> | 1168                         | 150                   | 55                           | 5                     |
| <b>GSE130970</b> | 1244                         | 407                   | 31                           | 20                    |

### 3.4.3 Meta-analysis

We performed separate meta-analyses using the results of the two previous approaches to obtain a measure of the overall effect of each of the 8223 elements evaluated (GO terms, KEGG pathways, or UniProt functions) throughout all individual studies, i.e., the LOR and its confidence interval. This effect indicated the magnitude of the overexpression of each term (the LOR value) and its direction (the LOR sign). This sign also helped to identify the group in which the element is over-represented. In our analysis, positive signs indicate over-representation in pre-menopausal women and negative signs over-representation in men.

Along with the LOR, the meta-analysis also provides an adjusted  $p$  value (FDR) that indicates the significance of the results. As a result of our meta-analysis, we identified thirteen significant GO terms (adjusted  $p$  value  $< 0.05$ ), which refer to the molecular mechanisms altered in our comparison (detailed in the following sections and (Table 3.3). Fig. 3.3b shows, for one function, the

LOR of each study and the joint LOR of the meta-analysis. Supplementary Figs. 2 and 3 detail the results of all significant elements. We confirmed the absence of bias and heterogeneity in all cases, as in Fig. 3.3c (Supplementary Figs. 4 and 5).

### 3.4.4 Gene functional meta-analysis

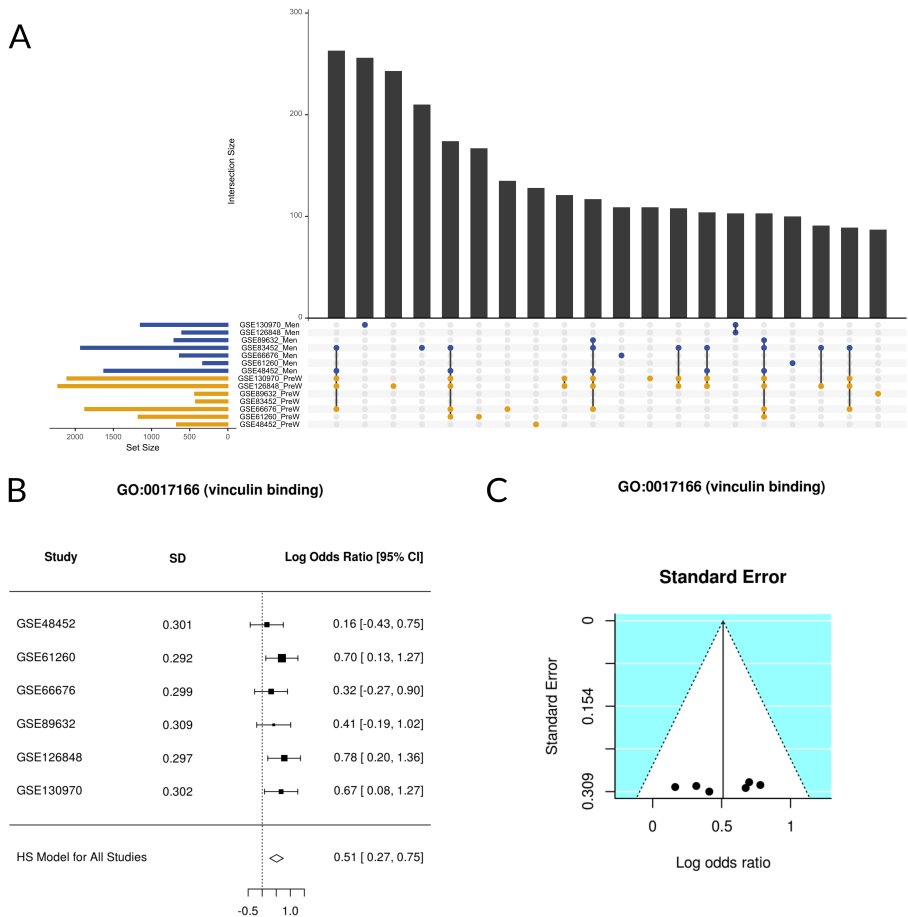
The functional meta-analyses of the individual gene differential expression pipeline results provided a total of nine significant GO terms (Table 3.3 and Fig. 3.4) but no significant KEGG pathways. Among the significant GO terms, we identified six overrepresented in premenopausal women; these elements encompassed a broad spectrum of biological processes and molecular functions, including “*RNA polymerase II regulatory region DNA binding*” (GO:0001012), “*RNA polymerase II regulatory region sequence-specific DNA binding*” (GO:0000977), “*Vinculin binding*” (GO:0017166), “*Cellular response to interleukin-2*” (GO:0071352), “*Negative regulation of neuron death*” (GO:1901215), and “*Response to interleukin-2*” (GO:0070669). We identified three terms overrepresented in men: “*Negative regulation of interleukin-6 secretion*” (GO:1900165), “*Establishment of planar polarity involved in neural tube closure*” (GO:0090177), and “*Regulation of establishment of planar polarity involved in neural tube closure*” (GO:0090178).

### 3.4.5 Pathway functional meta-analysis

The functional meta-analysis of the individual results in the pathway differential activation pipeline returned four significant GO terms. These terms were overrepresented in premenopausal women compared with men, with a LOR value lower than 0.25 (Table 3.3 and Fig. 3.4). These terms relate to the transport of ions and cations: “*Cation transmembrane transport*” (GO:0098655), “*Metal ion transport*” (GO:0030001), “*Regulation of cation transmembrane transport*” (GO:1904062), and “*Inorganic cation transmembrane transport*” (GO:0098662). The meta-analysis of the UniProt functions failed to provide significant results.

### ***3.4.6 Metafun-NAFLD platform***

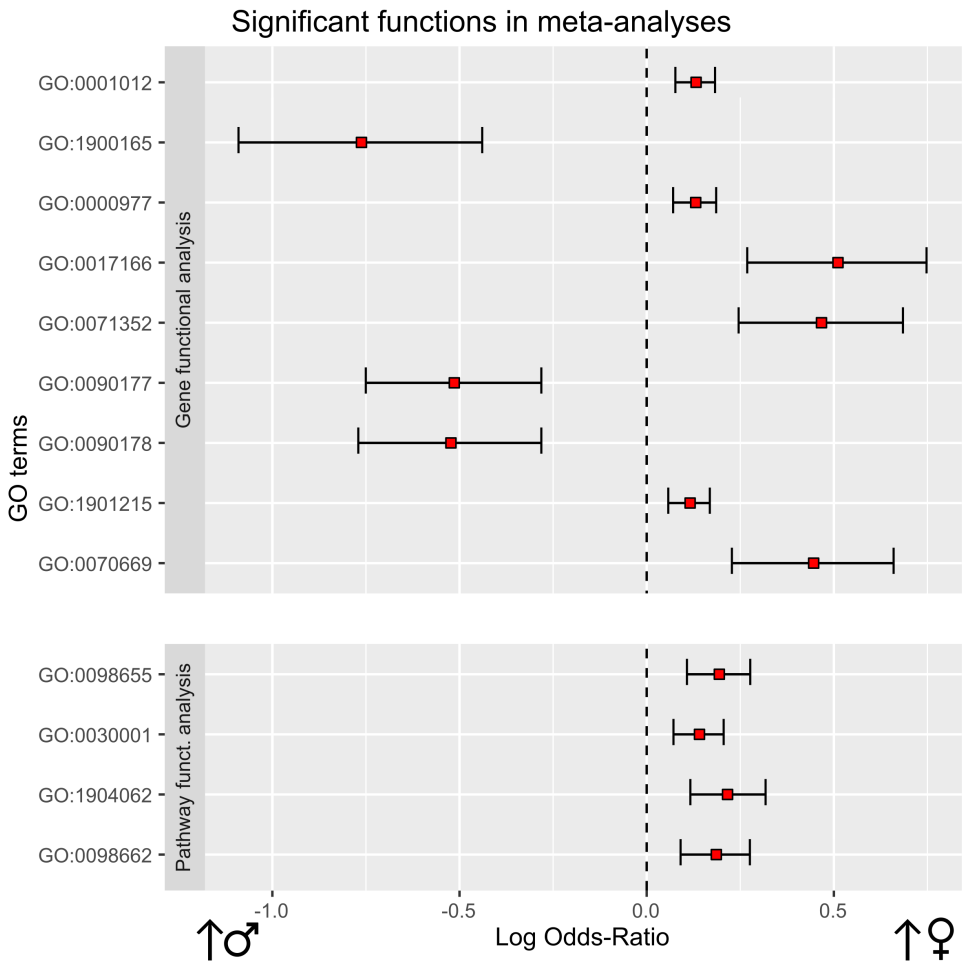
To contribute to data sharing and support the development of subsequent related studies, we generated the Metafun-NAFLD platform (<https://bioinfo.cipf.es/metafun-NAFLD>). This web tool contains the detailed results of all the steps of the bioinformatics pipeline and allows simple interactive exploration by any interested party.



**Fig. 3.3.** Representation of results of the bioinformatics analysis. **a** UpSet plot showing the number of common results throughout individual studies in our functional gene analysis. Only the twenty most abundant interactions are shown. Horizontal bars indicate the number of significant elements in each study (premenopausal [orange] women and men [blue]). The vertical bars indicate the common elements in the sets, indicated with dots under each bar. The single points represent the number of unique elements in each group. Those vertical bars with dots of both colors beneath indicate significant amounts of molecular bases in men and women, in a contradictory way, in different studies. **b** Example of a forest plot from the meta-analysis. This representation allows us to observe the LOR and confidence interval of the GO terms in each study (horizontal bars) and the meta-analysis (diamond). **c** Example of a funnel plot from the meta-analysis. Dots in the white area indicate the absence of bias and heterogeneity

**Table 3.3.** Significant functions in the functional meta-analyses. The logarithm of the odds ratio (LOR), its confidence interval (LOR-CI), the adjusted  $p$  value, and the group in which the function was overexpressed are shown ( $W$ , women;  $M$ , men)

|                                  | GO ID      | Name                                                                           | LOR    | LOR-CI           | Adj. $p$ value | Group |
|----------------------------------|------------|--------------------------------------------------------------------------------|--------|------------------|----------------|-------|
| Gene functional meta-analysis    | GO:0001012 | RNA polymerase II regulatory region DNA binding                                | 0.132  | [0.079, 0.185]   | 0.008          | W     |
|                                  | GO:1900165 | Negative regulation of interleukin-6 secretion                                 | -0.762 | [-1.088, -0.437] | 0.018          | M     |
|                                  | GO:0000977 | RNA polymerase II regulatory region sequence-specific DNA binding              | 0.131  | [0.073, 0.188]   | 0.02           | W     |
|                                  | GO:0017166 | Vinculin binding                                                               | 0.511  | [0.271, 0.75]    | 0.035          | W     |
|                                  | GO:0071352 | Cellular response to interleukin-2                                             | 0.467  | [0.248, 0.687]   | 0.035          | W     |
|                                  | GO:0090177 | Establishment of planar polarity involved in neural tube closure               | -0.514 | [-0.748, -0.279] | 0.035          | M     |
|                                  | GO:0090178 | Regulation of establishment of planar polarity involved in neural tube closure | -0.523 | [-0.768, -0.279] | 0.035          | M     |
| Pathway functional meta-analysis | GO:1901215 | Negative regulation of neuron death                                            | 0.116  | [0.06, 0.171]    | 0.043          | W     |
|                                  | GO:0070669 | Response to interleukin-2                                                      | 0.446  | [0.23, 0.662]    | 0.046          | W     |
|                                  | GO:0098655 | Cation transmembrane transport                                                 | 0.194  | [0.11, 0.279]    | 0.012          | W     |
|                                  | GO:0030001 | Metal ion transport                                                            | 0.141  | [0.074, 0.208]   | 0.027          | W     |
|                                  | GO:1904062 | Regulation of cation transmembrane transport                                   | 0.216  | [0.119, 0.32]    | 0.027          | W     |
|                                  | GO:0098662 | Inorganic cation transmembrane transport                                       | 0.186  | [0.093, 0.278]   | 0.038          | W     |
|                                  |            |                                                                                |        |                  |                |       |



**Fig. 3.4.** Differential functional profiling by sex. This plot shows significant functional terms of each meta-analysis. On the right: biological functions over-represented in women. On the left: biological functions more overrepresented in men. For each function, the LOR (red square) and confidence interval have been represented



### 3.5 Discussion

Although previous studies have analyzed pathobiological mechanisms in NAFLD and (separately) sexual differences in metabolic regulation, the impact of sex on NAFLD pathobiology remains incompletely defined. Epidemiological studies have reported a higher prevalence of NAFLD in men and postmenopausal women than premenopausal women [4]. To better understand the molecular basis of sex-based differences, we present a systematic review and functional meta-analysis of publicly available genomic studies that analyze the differences in NAFLD progression between premenopausal women and men.

Despite the limitations of our approach (presence of studies with different sample sizes and types of platforms), meta-analysis techniques can integrate the selected studies. The individual results of the seven selected studies failed to identify any consistently different GO or KEGG terms between groups across all studies, a finding perhaps related to small sample sizes, underlying genetic and environmental variance in the source population, and methodological heterogeneity. Surprisingly, as shown in Fig. 3.3a, we found certain functions overrepresented in both groups in different studies. A meta-analysis, such as that carried out in this study, can help to improve the identification of consistent, relevant differences across studies, subject to the caveat that patients in each study may have other differences in clinical characteristics. A meta-analysis also increases the overall sample size and hence the strength of findings, which is of particular relevance to diseases such as NAFLD that displays high levels of heterogeneity. Indeed, systematic reviews and meta-analyses represent widely employed strategies in studying sex-based differences in other diseases [37-39].

Our functional meta-analysis identified a total of thirteen GO terms (Table 3.3) that displayed differences between men and premenopausal women using two different approaches—differentially expressed genes (nine terms) and differentially activated pathways (four terms). These terms broadly related to processes such as inflammation, cell binding, and the establishment of polarity during neural tube closure.

A functional meta-analysis of differentially expressed genes identified differences in nine significant GO terms. Among these, our results established the overrepresentation of the “*Vinculin binding*” (GO:0017166) in premenopausal women compared with men. The vinculin cytoplasmic actin-binding protein is enriched in focal adhesions and adherens junctions to contribute to junction stability [40] and plays a vital role as a regulator of apoptosis [41]. Of note, although “cellular adhesion” represents a broad set of processes rather than a specific biologic pathway, this term was previously reported as a GO term

consistently enriched in male NASH patients [42]. Furthermore, Arendt et al. identified differentially expressed genes related to cell–cell adhesion (*ANXA2* and *MAG*) between NAFL and NASH patients [34].

Our analysis also identified sex-based differences related to the inflammatory response, specifically the expression of cytokine-related terms. We found “*Negative regulation of interleukin-6 secretion*” (GO:1900165) overrepresented in men and “*Cellular response to interleukin-2*” (GO:0071352) and “*Response to interleukin-2*” (GO:0070669) overrepresented in premenopausal women. These terms relate to interleukin 2 (IL-2) and interleukin 6 (IL-6), which displayed LOR values close to or above 0.5, thereby indicating a substantial difference between the groups. IL-2 is a known regulator of T cell responses, with low IL-2 levels regulating T cell central tolerance via the formation of regulatory T cells [43]. Higher IL-2 levels induce T cell proliferation and differentiation into memory T cells. IL-2 receptors are also expressed by B cells and cells of the innate immune system [44]. Interestingly, IL-2 signaling is deficient in many chronic liver diseases [45], and levels of soluble IL-2 receptor alpha were found to be higher in children with NASH and advanced fibrosis [46]. Our results suggest that low-level IL-2-induced immune tolerance may mitigate the risk of developing NASH in premenopausal women and raise the intriguing possibility of restoring low-level IL-2 signaling in men as a potential treatment for NAFLD.

IL-6 has been previously implicated in NASH, with increased hepatic expression correlating with disease severity and insulin resistance [47]. IL-6 also regulates hepatic regeneration [48] and represents a significant driver of hepatic carcinogenesis [49]. In NAFLD, IL-6 may enhance hepatic repair and regeneration; however, IL-6 may also promote insulin resistance and hepatocyte apoptosis, thus contributing to NASH development [48, 50]. Indeed IL-6 gene polymorphisms have been associated with increased susceptibility to chronic liver disease [51].

These results suggest that differences in immune mechanisms between men and women may represent one possible factor in the sex-differences observed in NAFLD. For example, the macrophage phenotype in male NAFLD mice presents as pro-inflammatory and pro-fibrotic, but pro-resolution and anti-fibrotic in female mice [52]. We hypothesize that women may suffer from a decreased immune response to lipotoxicity, whereas men may suffer from an impaired liver-healing response to chronic injury perhaps related to deficient IL-6 signaling. These ideas require further examination and validation in additional studies.

A recent study by Herrera-Marcos et al. detected sex-based differences in the control of the *Cidec/Fsp27* gene in the liver, finding that gene expression correlated with the presence and density of liver lipid droplets in mice fed a high-fat diet [53]. Matsusue suggested a role for *Cidec/Fsp27* as a neuronal activator in the steatotic liver [54]; therefore, the significant association with the function “*Negative regulation of neuron death*” (GO:1901215) could relate to the differences detected in this previous study. Upon restricting the gene ontology analysis to *Homo sapiens*, we found the components of the “*Negative regulation of neuron death*” pathway (GO: 1901215) as broad regulators of apoptosis. This finding may suggest an increase in liver cell apoptosis in men who develop advanced NASH at a younger age than women, as premenopausal women possess a level of protection. The observation that hepatocyte apoptosis represents a prominent feature of progressive NASH agrees with this hypothesis [55].

Interestingly, we discovered other significant GO terms that had not previously been associated with NAFLD. For instance, we found the overrepresentation of “*Establishment of planar polarity involved in neural tube closure*” (GO:0090177) and “*Regulation of establishment of planar polarity involved in neural tube closure*” (GO:0090178) in men with a LOR exceeding 0.5 (-0.514 and -0.523). Restricting analyses to *Homo sapiens* demonstrated that these GO terms relate to the Wnt and frizzled family of proteins. Wnt ligands and their cognate receptor frizzled represent critical determinants of liver development, zonation, regeneration, and responses to liver injury [56]. Their overrepresentation in male NASH patients could indicate the activation of disease repair pathways as a compensatory response to ongoing liver injury [57], as observed in preclinical models of NASH.

Our functional pathway meta-analysis identified four significant GO terms involved in cation transport: “*Cation transmembrane transport*” (GO:0098655), “*Metal ion transport*” (GO:0030001), “*Regulation of cation transmembrane transport*” (GO:1904062), and “*Inorganic cation transmembrane transport*” (GO:0098662). Previous studies have associated the alteration of pathways and functions related to cation transport with NAFLD [58]; however, information regarding relevance and significance remains scarce.

### **3.5.1 Perspectives and significance**

Our results support the existence of differences between men and premenopausal women in specific biological processes and molecular functions relevant to NAFLD. Given the well-known heterogeneity of NAFLD progression between men and women, this transcriptomic meta-analysis suggests the use of sex-specific pathophysiologically informed disease biomarkers. These findings may also impact disease diagnosis and prognosis and the design of clinical trials. For example, biological sex-driven differences may allow the enhanced classification of individuals with NAFLD and alter their inclusion in specific clinical trials. Our meta-analysis underscores the importance of recognizing and accounting for sex-related differences in the study of NAFLD pathobiology.

## **3.6 Conclusions**

While differences in clinical characteristics and pathogenic mechanisms in male and female NAFLD patients had been explored previously, details remained scarce. To this end, we selected seven studies suitable for functional meta-analysis and identified thirteen significant GO terms from these data. Of note, specific GO terms were previously identified to have relevance to NAFLD, which supports the notion that sex represents an important variable in understanding NAFLD pathobiology and heterogeneity. We identified novel functional terms that had not been linked to NAFLD previously; therefore, our study may open new avenues for further research. In conclusion, the meta-analysis of transcriptomic data represents a useful tool for the identification of sex-related differences in NAFL and NASH.

## **3.7 Availability of data and materials**

The large volume of data and results generated in this study are freely available through the metafun-NAFLD web tool: <https://bioinfo.cipf.es/metafun-NAFLD>. This study analyzed transcriptomic data available in the Gene Expression Omnibus database with accession numbers GSE48452, GSE61260, GSE66676, GSE83452, GSE89632, GSE126848, and GSE130970.

### 3.8 References

1. Friedman SL, Neuschwander-Tetri BA, Rinella M, Sanyal AJ. Mechanisms of NAFLD development and therapeutic strategies. *Nat Med.* 2018;24:908-922. <https://doi.org/10.1038/s41591-018-0104-9>.
2. Younossi ZM, Koenig AB, Abdelatif D, Fazel Y, Henry L, Wymer M. Global epidemiology of nonalcoholic fatty liver disease-meta-analytic assessment of prevalence, incidence, and outcomes. *Hepatology.* 64:73-84. <https://doi.org/10.1002/hep.28431>.
3. Singh S, Allen AM, Wang Z, Prokop LJ, Murad MH, Loomba R. Fibrosis progression in nonalcoholic fatty liver vs nonalcoholic steatohepatitis: a systematic review and meta-analysis of paired-biopsy studies. *Clin Gastroenterol Hepatol.* 2015;13:643-654.e9. <https://doi.org/10.1016/j.cgh.2014.04.014>.
4. Ballestri S, Nascimbeni F, Baldelli E, Marrazzo A, Romagnoli D, Lonardo A. 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. *Adv Ther.* 2017;34:1291-1326. <https://doi.org/10.1007/s12325-017-0556-1>.
5. Skubic C, Drakulić Ž, Rozman D. Personalized therapy when tackling nonalcoholic fatty liver disease: a focus on sex, genes, and drugs. *Expert Opin Drug Metab Toxicol.* 2018;14:831-841. <https://doi.org/10.1080/17425255.2018.1492552>.
6. Zheng D, Wang X, Antonson P, Gustafsson J-Å, Li Z. Genomics of sex hormone receptor signaling in hepatic sexual dimorphism. *Mol Cell Endocrinol.* Elsevier. 2018;471:33-41. <https://doi.org/10.1016/j.mce.2017.05.025>.
7. Ling G, Sugathan A, Mazor T, Fraenkel E, Waxman DJ. Unbiased, genome-wide in vivo mapping of transcriptional regulatory elements reveals sex differences in chromatin structure associated with sex-specific liver gene expression. *Mol Cell Biol.* 2010;30:5531-5544. <https://doi.org/10.1128/MCB.00601-10>.
8. Lonardo A, Suzuki A. Nonalcoholic fatty liver disease: Does sex matter? *HepatoBiliary Surg Nutr.* 2019;8:164-166. <https://doi.org/10.21037/hbsn.2018.12.04>

9. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, et al. NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res.* 2012;41:D991-D995. <https://doi.org/10.1093/nar/gks1193>.
10. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: the PRISMA statement. *PLoS Med.* 2009;6:e1000097. <https://doi.org/10.1371/journal.pmed.1000097>.
11. Aller R, Fernández-Rodríguez C, lo Iacono O, Bañares R, Abad J, Carrión JA, et al. Consensus document. Management of non-alcoholic fatty liver disease (NAFLD). Clinical practice guideline. *Gastroenterol y Hepatol.* 2018;41:328-349. <https://doi.org/10.1016/j.gastre.2018.05.013>.
12. R Core Team. R: A Language and environment for statistical computing. Vienna, Austria; 2019. <http://www.r-project.org/>
13. Bolstad BM, Irizarry RA, Astrand M, Speed TP. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics.* 2003;19:185-193. <https://doi.org/10.1093/bioinformatics/19.2.185>.
14. Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* 2010;11:R25. <https://doi.org/10.1186/gb-2010-11-3-r25>.
15. Davis SR, Lambrinoudaki I, Lumsden M, Mishra GD, Pal L, Rees M, et al. Menopause. *Nat Rev Dis Prim.* 2015;1:15004. <https://doi.org/10.1038/nrdp.2015.4>.
16. Freeman EW, Sammel MD, Sanders RJ. Risk of long-term hot flashes after natural menopause. *Menopause.* 2014;21:924-932. <https://doi.org/10.1097/GME.000000000000196>.
17. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* Oxford University Press. 2015;43:e47. <https://doi.org/10.1093/nar/gkv007>.
18. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a bioconductor package for differential expression analysis of digi-

- tal gene expression data. *Bioinformatics*. 2010;26:139-140. <https://doi.org/10.1093/bioinformatics/btp616>.
19. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B*. 1995;57:289-300.
  20. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci*. 2005;102:15545-15550. <https://doi.org/10.1073/pnas.0506580102>.
  21. Montaner D, Dopazo J. Multidimensional gene set analysis of genomic data. *PLoS One*. 2010;5:e10348. <https://doi.org/10.1371/journal.pone.0010348>.
  22. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene Ontology: tool for the unification of biology. *Nat Genet*. 2000;25:25-29. <https://doi.org/10.1038/75556>.
  23. Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res*. 2000;28:27-30. <https://doi.org/10.1093/nar/27.1.29>.
  24. Lex A, Gehlenborg N. Sets and intersections. *Nat Methods*. 2014;11:779. <https://doi.org/10.1038/nmeth.3033>.
  25. Hidalgo MR, Cubuk C, Amadoz A, Salavert F, Carbonell-Caballero J, Dopazo J. High throughput estimation of functional cell activities reveals disease mechanisms and predicts relevant clinical outcomes. *Oncotarget*. 2017;8:5160-5178. <https://doi.org/10.18632/oncotarget.14107>.
  26. The UniProt Consortium. UniProt: the universal protein knowledgebase. *Nucleic Acids Res*. 2016;45:D158-D169. <https://doi.org/10.1093/nar/gkw1099>.
  27. García-García F. Métodos de análisis de enriquecimiento funcional en estudios genómicos: Universidad de Valencia; 2016. <https://www.educacion.gob.es/teseo/mostrarRef.do?ref=1307283>.
  28. Viechtbauer W. Conducting meta-analyses in R with the metafor package. *J Stat Softw*. 2010;36. <https://doi.org/10.18637/jss.v036.i03>.

29. Schmidt FL, Hunter JE. Methods of meta-analysis: correcting error and bias in research findings. 3rd ed. SAGE Publications; 2015. <https://doi.org/10.4135/9781483398105>.
30. Ahrens M, Ammerpohl O, von Schönfels W, Kolarova J, Bens S, Itzel T, et al. DNA methylation analysis in nonalcoholic fatty liver disease suggests distinct disease-specific and remodeling signatures after bariatric surgery. *Cell Metab.* 2013;18:296-302. <https://doi.org/10.1016/j.cmet.2013.07.004>.
31. Horvath S, Erhart W, Brosch M, Ammerpohl O, von Schönfels W, Ahrens M, et al. Obesity accelerates epigenetic aging of human liver. *Proc Natl Acad Sci.* 2014;111:15538-15543. <https://doi.org/10.1073/pnas.1412759111>.
32. Xanthakos SA, Jenkins TM, Kleiner DE, Boyce TW, Mourya R, Karns R, et al. High Prevalence of nonalcoholic fatty liver disease in adolescents undergoing bariatric surgery. *Gastroenterology.* 2015;149:623-634.e8. <https://doi.org/10.1053/j.gastro.2015.05.039>.
33. Lefebvre P, Lalloyer F, Baugé E, Pawlak M, Gheeraert C, Dehondt H, et al. Interspecies NASH disease activity whole-genome profiling identifies a fibrogenic role of PPAR $\alpha$ -regulated dermatopontin. *JCI Insight.* *Am Soc Clin Invest.* 2017;2(13):e92264. <https://doi.org/10.1172/jci.insight.92264>.
34. Arendt BM, Comelli EM, Ma DWL, Lou W, Teterina A, Kim T, 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.* 2015;61:1565-1578. <https://doi.org/10.1002/hep.27695>.
35. Suppli MP, Rigbolt KTG, Veidal SS, Heebøll S, Eriksen PL, Demant M, et al. Hepatic transcriptome signatures in patients with varying degrees of nonalcoholic fatty liver disease compared with healthy normal-weight individuals. *Am J Physiol Gastrointest Liver Physiol.* 2019;316:G462-G472. <https://doi.org/10.1152/ajpgi.00358.2018>.
36. Hoang SA, Oseini A, Feaver RE, Cole BK, Asgharpour A, Vincent R, et al. Gene expression predicts histological severity and reveals distinct molecular profiles of nonalcoholic fatty liver disease. *Sci Rep.* 2019;9:12541. <https://doi.org/10.1038/s41598-019-48746-5>.



- 
37. Ryaboshapkina M, Hammar M. Human hepatic gene expression signature of non-alcoholic fatty liver disease progression, a meta-analysis. *Sci Rep.* 2017;7:12361. <https://doi.org/10.1038/s41598-017-10930-w>.
  38. Ruigrok ANV, Salimi-Khorshidi G, Lai M-C, Baron-Cohen S, Lombardo MV, Tait RJ, et al. A meta-analysis of sex differences in human brain structure. *Neurosci Biobehav Rev.* 2014;39:34-50. <https://doi.org/10.1016/j.neubiorev.2013.12.004>.
  39. Casanova Ferrer F, Pascual M, Hidalgo MR, Malmierca-Merlo P, Guerri C, García-García F. Unveiling sex-based differences in the effects of alcohol abuse: a comprehensive functional meta-analysis of transcriptomic studies. *Genes (Basel).* 2020;11:1106. <https://doi.org/10.3390/genes11091106>.
  40. Konishi S, Yano T, Tanaka H, Mizuno T, Kanoh H, Tsukita K, et al. Vinculin is critical for the robustness of the epithelial cell sheet paracellular barrier for ions. *Life Sci Alliance.* 2019;2:e201900414. <https://doi.org/10.26508/lsa.201900414>.
  41. Peng X, Nelson ES, Maiers JL, DeMali KA. Chapter five - new insights into vinculin function and regulation. In: Jeon KW, editor. Academic Press; 2011. p. 191-231. <https://doi.org/10.1016/B978-0-12-386043-9.00005-0>.
  42. Vandel J, Dubois-Chevalier J, Gheeraert C, Derudas B, Raverdy V, Thuillier D, et al. Hepatic molecular signatures highlight the sexual dimorphism of Non-Alcoholic SteatoHepatitis (NASH). *Hepatology.* 2020;hep.31312. <https://doi.org/10.1002/hep.31312>.
  43. Ward NC, Yu A, Moro A, Ban Y, Chen X, Hsiung S, et al. IL-2/CD25: a long-acting fusion protein that promotes immune tolerance by selectively targeting the IL-2 receptor on regulatory T cells. *J Immunol.* 2018;201:2579-2592. <https://doi.org/10.4049/jimmunol.1800907>.
  44. Waters RS, Perry JSA, Han S, Bielekova B, Gedeon T. The effects of interleukin-2 on immune response regulation. *Math Med Biol.* 2018;35:79-119. <https://doi.org/10.1093/imammb/dqw021>.
  45. Saxena S, Nouri-Aria KT, Anderson MG, Eddleston AL, Williams R. Interleukin 2 activity in chronic liver disease and the effect of in vitro alpha-interferon. *Clin Exp Immunol.* 1986;63:541-548.

46. Perito ER, Ajmera V, Bass NM, Rosenthal P, Lavine JE, Schwimmer JB, et al. Association between cytokines and liver histology in children with nonalcoholic fatty liver disease. *Hepatol Commun.* 2017;1:609-622. <https://doi.org/10.1002/hep4.1068>.
47. Wieckowska A, Papouchado BG, Li Z, Lopez R, Zein NN, Feldstein AE. Increased hepatic and circulating interleukin-6 levels in human nonalcoholic steatohepatitis. *Am J Gastroenterol.* 2008;103:1372-1379.
48. Braunersreuther V, Viviani GL, Mach F, Montecucco F. Role of cytokines and chemokines in non-alcoholic fatty liver disease. *World J Gastroenterol.* 2012;18:727. <https://doi.org/10.3748/wjg.v18.i8.727>.
49. Schmidt-Arras D, Rose-John S. IL-6 pathway in the liver: from physiopathology to therapy. *J Hepatol.* 2016;64:1403-1415. <https://doi.org/10.1016/j.jhep.2016.02.004>.
50. Yamaguchi K, Itoh Y, Yokomizo C, Nishimura T, Niimi T, Fujii H, et al. Blockade of interleukin-6 signaling enhances hepatic steatosis but improves liver injury in methionine choline-deficient diet-fed mice. *Lab Invest.* 2010;90:1169-1178. <https://doi.org/10.1038/labinvest.2010.75>.
51. Wang X, Yan Z, Ye Q. Interleukin-6 gene polymorphisms and susceptibility to liver diseases: a meta-analysis. *Medicine (Baltimore).* 2019;98:e18408. <https://doi.org/10.1097/MD.00000000000018408>.
52. Lonardo A, Nascimbeni F, Ballestri S, Fairweather D, Win S, Than TA, et al. Sex differences in nonalcoholic fatty liver disease: state of the art and identification of research gaps. *Hepatology.* 2019;70:1457-1469. <https://doi.org/10.1002/hep.30626>.
53. Herrera-Marcos LV, Sancho-Knapik S, Gabás-Rivera C, Barranquero C, Gascón S, Romanos E, et al. Pgc1a is responsible for the sex differences in hepatic Cidec/Fsp27 $\beta$  mRNA expression in hepatic steatosis of mice fed a Western diet. *Am J Physiol Metab.* 2020;318:E249-E261. <https://doi.org/10.1152/ajpendo.00199.2019>.
54. Matsusue K. A physiological role for fat specific protein 27/cell death-inducing DFF45-like effector C in adipose and liver. *Biol Pharm Bull.* 2010;33:346-350. <https://doi.org/10.1248/bpb.33.346>.
55. Feldstein AE, Canbay A, Angulo P, Taniai M, Burgart LJ, Lindor KD, et al. Hepatocyte apoptosis and fas expression are prominent features of

- human nonalcoholic steatohepatitis. *Gastroenterology*. 2003;125:437-443. [https://doi.org/10.1016/s0016-5085\(03\)00907-7](https://doi.org/10.1016/s0016-5085(03)00907-7).
56. Russell JO, Monga SP. Wnt/ $\beta$ -catenin signaling in liver development, homeostasis, and pathobiology. *Annu Rev Pathol*. 2018;13:351-378. <https://doi.org/10.1146/annurev-pathol-020117-044010>.
57. Zhu L, Baker SS, Shahein A, Choudhury S, Liu W, Bhatia T, et al. Up-regulation of non-canonical Wnt ligands and oxidative glucose metabolism in NASH induced by methionine-choline deficient diet. *Trends cell Mol Biol*. 2018;13:47-56. <https://doi.org/10.31300/tcmb.13.2018.47-56>.
58. Pirola CJ, Sookoian S. Multiomics biomarkers for the prediction of nonalcoholic fatty liver disease severity. *World J Gastroenterol*. 2018;24:1601-1615. <https://doi.org/10.3748/wjg.v24.i15.1601>.

### **3.9 Acknowledgements**

The authors would like to thank the Principe Felipe Research Centre (CIPF) for providing access to its computational cluster to carry out the work associated with this study. We would also like to thank Dr. Kristoffer T.G. Rigbolt for providing information regarding the ages of the samples used in study GSE126848.

### **3.10 Funding**

This research was supported by grants from GV/2020/186 and ISCIII PT17/0009/0015 FEDER. Marina Berenguer was supported by a grant from the ISCIII (PI19/01360).

### **3.11 Authors' contributions**

JFC analyzed the data; MRH and FGG designed and supervised the analysis; PMM designed and implemented the web tool; JFC, MRH, and FGG wrote the manuscript; MRH designed the graphical abstract; MB, GP, MIV, and HM helped in the interpretation of the results; MRH and FGG conceived the work. The authors read and approved the final manuscript.

### 3.12 Author details

<sup>1</sup>Bioinformatics and Biostatistics Unit, Principe Felipe Research Center, Valencia, Spain.

<sup>2</sup>Spanish National Bioinformatics Institute, ELIXIR-Spain (INB, ELIXIR-ES).

<sup>3</sup>Liver Transplantation and Hepatology Unit, Hospital Universitario y Politécnico La Fe, Valencia, Spain.

<sup>4</sup>Grupo de Hepatología, Cirugía HBP y Trasplantes, Instituto de Investigación Sanitaria La Fe, Valencia, Spain.

<sup>5</sup>CIBERehd, Centro de Investigación Biomédica en Red en Enfermedades Hepáticas y Digestivas, Instituto de Salud Carlos III, Madrid, Spain.

<sup>6</sup>Department of Medicine, Universitat de València, Valencia, Spain.

<sup>7</sup>Division of Gastroenterology and Hepatology, Mayo Clinic, Rochester, MN, USA.

<sup>8</sup>Atos Research Innovation (ARI), Spain.

<sup>9</sup>Biomedical Imaging Unit FISABIO-CIPF. Fundación para el Fomento de la Investigación Sanitario y Biomédica de la Comunidad Valenciana. Valencia, Spain.

\*Corresponding authors: mhidalgo@cipf.es (Marta R. Hidalgo), fgarcia@cipf.es (Francisco García-García)

### 3.13 Supplementary material

Supplementary Figures 2 and 4 have been adjusted from the original article to fit the format of this manuscript. All supplementary figures and tables of this article are available both on the article's website (<https://doi.org/10.1186/s13293-021-00368-1>) and in the public repository Zenodo (<https://zenodo.org/doi/10.5281/zenodo.14062682>) for better visualization.



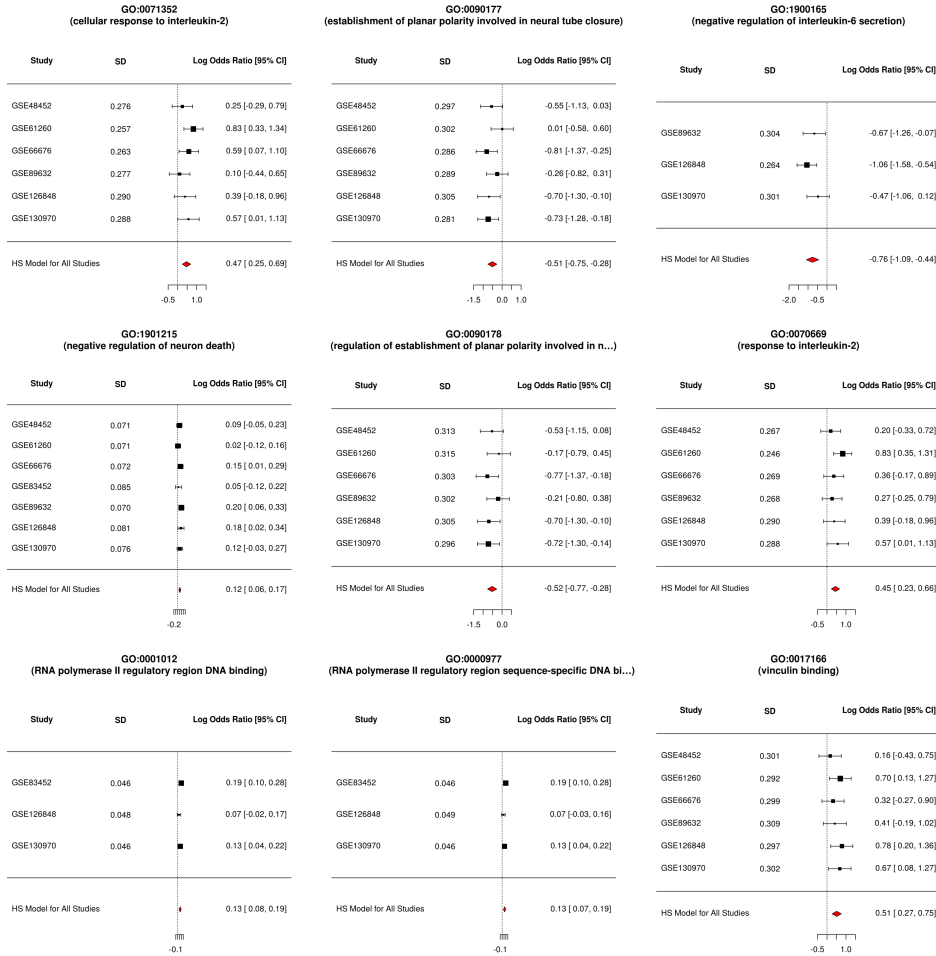
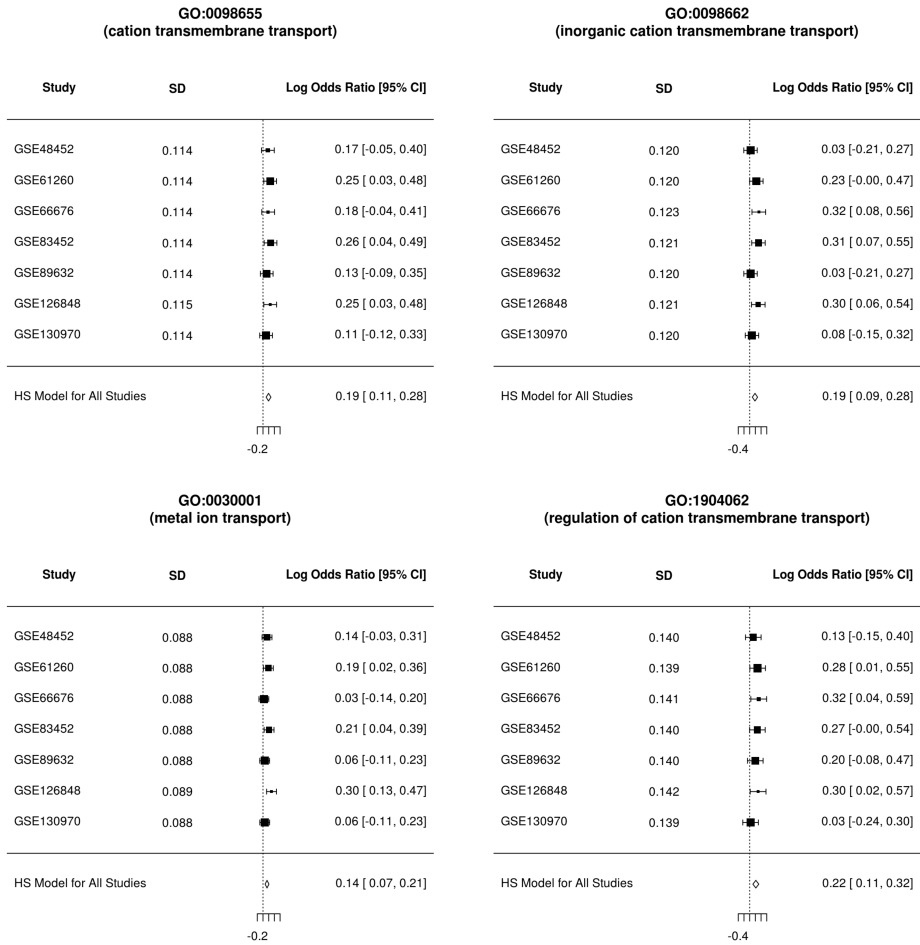
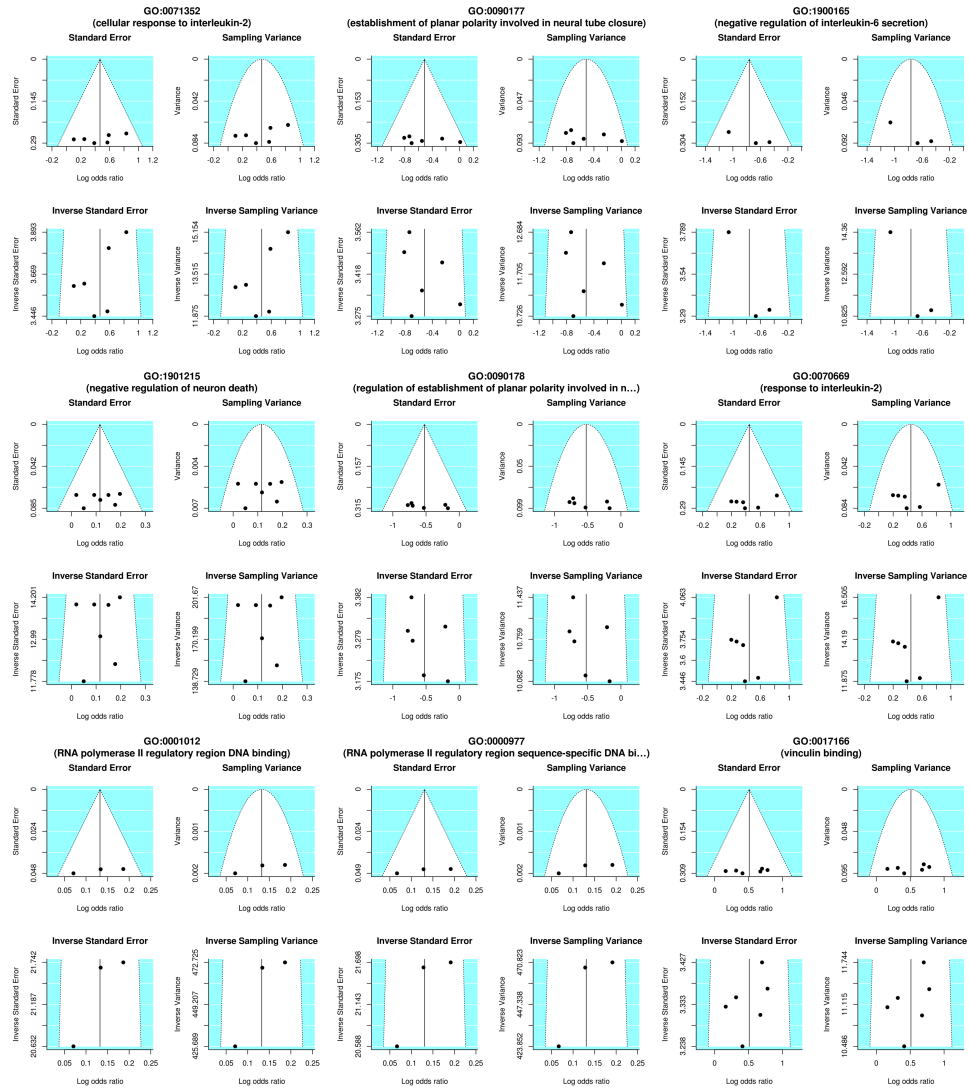


Fig. 3.S2. Forest plots for significant GO terms from our functional gene meta-analysis.

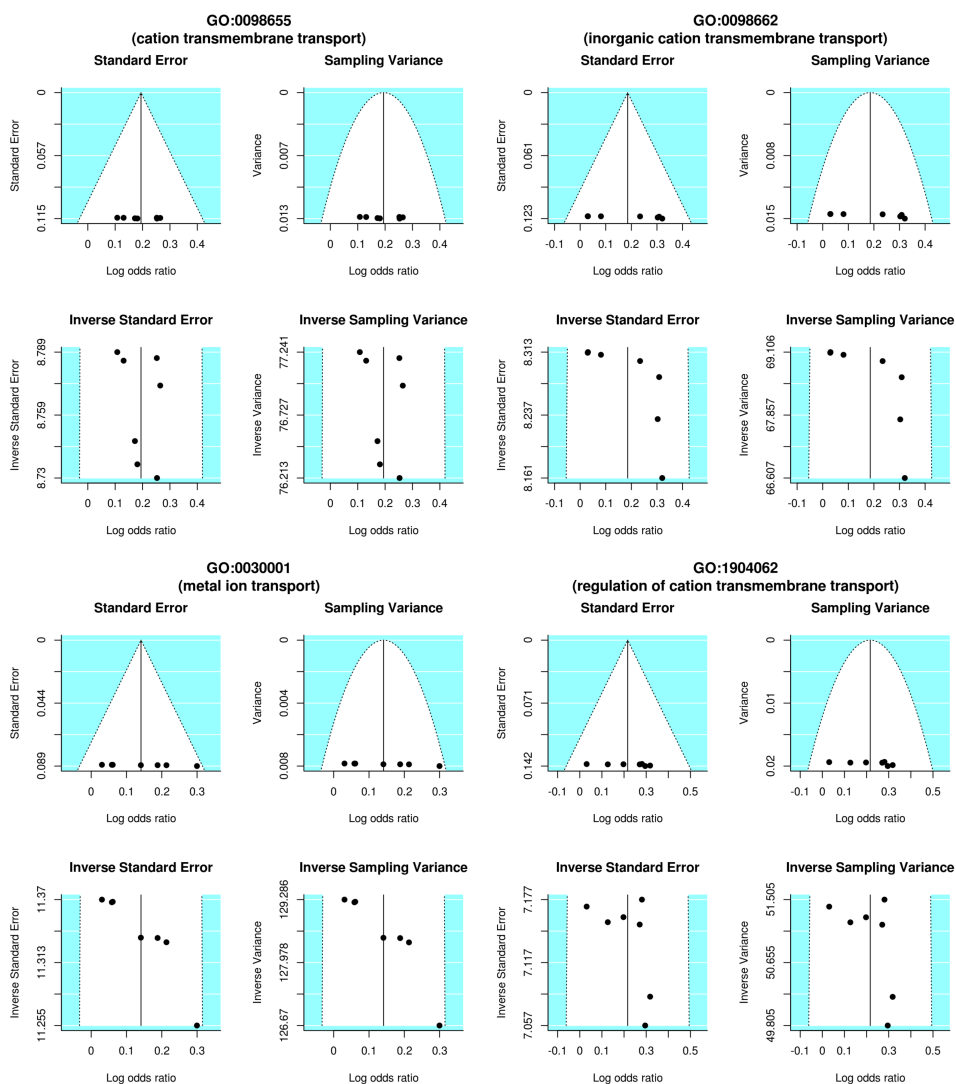


**Fig. 3.S3. Forest plots of significant GO terms from our functional pathways meta-analysis.**



**Fig. 3.S4. Funnel plots of the significant GO terms from our functional gene meta-analysis.** In the absence of bias and heterogeneity, we would expect the points to be distributed in the form of a funnel, within the region of confidence.





**Fig. 3.S5. Funnel plots of significant GO terms from our functional pathway meta-analysis.** In the absence of bias and heterogeneity, we would expect the points to be distributed in the form of a funnel, within the region of confidence.



## Chapter 4

# A deep transcriptome meta-analysis reveals sex differences in multiple sclerosis

José F. Català-Senent<sup>1†</sup>, Zoraida Andreu<sup>2†</sup>, Marta R. Hidalgo<sup>1</sup>, Irene Soler-Sáez<sup>1</sup>, Francisco José Roig<sup>1,3</sup>, Natalia Yanguas-Casás<sup>4</sup>, Almudena Neva-Alejo<sup>1</sup>, Adolfo López-Cerdán<sup>5</sup>, María de la Iglesia-Vayá<sup>5</sup>, Barbara E. Stranger<sup>6</sup> and Francisco García-García<sup>1</sup>

Published in **Neurobiology of Disease**

[DOI: 10.1016/j.nbd.2023.106113]



## 4.1 Abstract

**Background:** Multiple sclerosis (MS), a chronic auto-immune, inflammatory, and degenerative disease of the central nervous system, affects both males and females; however, females suffer from a higher risk of developing MS (2-3:1 ratio relative to males). The precise sex-based factors influencing risk of MS are currently unknown. Here, we explore the role of sex in MS to identify molecular mechanisms underlying observed MS sex differences that may guide novel therapeutic approaches tailored for males or females.

**Methods:** We performed a rigorous and systematic review of genome-wide transcriptome studies of MS that included patient sex data in the Gene Expression Omnibus and ArrayExpress databases following PRISMA statement guidelines. For each selected study, we analyzed differential gene expression to explore the impact of the disease in females (IDF), in males (IDM) and our main goal: the sex differential impact of the disease (SDID). Then, for each scenario (IDF, IDM and SDID) we performed 2 meta-analyses in the main tissues involved in the disease (brain and blood). Finally, we performed a gene set analysis in brain tissue, in which a higher number of genes were dysregulated, to characterize sex differences in biological pathways.

**Results:** After screening 122 publications, the systematic review provided a selection of 9 studies (5 in blood and 4 in brain tissue) with a total of 474 samples (189 females with MS and 109 control females; 82 males with MS and 94 control males). Blood and brain tissue meta-analyses identified, respectively, 1 (*KIR2DL3*) and 13 (*ARL17B*, *CECR7*, *CEP78*, *IFFO2*, *LOC401127*, *NUDT18*, *RNF10*, *SLC17A5*, *STMP1*, *TRAF3IP2-AS1*, *UBXN2B*, *ZNF117*, *ZNF488*) MS-associated genes that differed between males and females (SDID comparison). Functional analyses in the brain revealed different altered immune patterns in females and males (IDF and IDM comparisons). The pro-inflammatory environment and innate immune responses related to myeloid lineage appear to be more affected in females, while adaptive responses associated with the lymphocyte lineage in males. Additionally, females with MS displayed alterations in mitochondrial respiratory chain complexes, purine, and glutamate metabolism, while MS males displayed alterations in stress response to metal ion, amine, and amino acid transport.

**Conclusions:** We found transcriptomic and functional differences between MS males and MS females (especially in the immune system), which may support the development of new sex-based research of this disease. Our study highlights the importance of understanding the role of biological sex in MS to guide a more personalized medicine.

**Keywords:** Multiple sclerosis, biomarkers, functional profiling, meta-analysis, sex-based differences, neurodegeneration

**Highlights:**

- Brain tissue meta-analysis identified 13 differentially affected genes by sex.
- Blood meta-analysis identified *KIR2DL3*: sex-differential pattern in self-tolerance.
- Functional analyses revealed different altered immune patterns in females and males.
- Understanding the role of biological sex in MS to guide a more personalized medicine

## 4.2 Background

Multiple sclerosis (MS), is a chronic autoimmune and neurodegenerative disease with exacerbated inflammatory reactions. In MS, the blood brain barrier (BBB) is disrupted and immune cells reach the central nervous system (CNS) causing demyelination and neurodegeneration. MS is mainly diagnosed in young adults (around 20-40 years) and represents the leading cause of disability in this age group that does not derive from trauma [1, 2]. The disease displays different phenotypes and progression stages characterized by relapses and remissions, being the relapsing-remitting MS (RRMS) the most common form of the disease in female and male [3]. Cognitive, motoric, sensory, visual and emotional fields are affected depending on whether the brain or the spinal cord are damaged.

While the etiology of MS remains unknown, genetic and environmental factors impact disease-onset, additionally sex significantly influences MS [4]. As with other autoimmune and neurodegenerative diseases, MS exhibits a strong sex bias in incidence, progression and disability [3, 5]. While females are more susceptible to MS than males, males are (on average) older than females at MS diagnosis and suffer from worse and more rapid progression. Males are also more likely to develop progressive MS than females [6, 7]. Sex differences in autoimmune and neurological diseases have been previously described [3, 5, 6] and associated with the presence of sex chromosomes (XX vs. XY, gene dosage, imprinting, and transcriptomic profile) and endogenous sex hormones (estrogens, progesterone, androgens, and prolactin) that impact the immune system and promote changes in disease severity [6, 8]. Broadly, estrogens (estradiol (E2)) and prolactin, work as enhancers at least of humoral immunity, and testosterone and progesterone as natural immunosuppressants [5, 9]. Additionally, the transcriptomic profiles and epigenetic modifications on sex chromosomes influence hormonal actions in immune responses in both sexes. Responses to environmental factors such as microbial exposure, diet, and sociological differences, also represent sex-based factors in MS [10]. Therefore, differences in the prevalence and severity of MS between males and females involve complex and poorly understood interactions between genetic, hormonal, and environmental factors. This knowledge gap, in turn, impairs the development of personalized therapeutics based on sex-specific information.

Multi-omics-based studies in large and clinically well-characterized patient cohorts with associated patient sex data may identify critical pathways and networks related to sexual dimorphism, and thereby support the discovery of new biomarkers and targeted sex-specific therapeutic interventions. In silico ap-

proaches using the large volumes of biological information accessible in public repositories of biomedical data, such as GEO (Gene Expression Omnibus) [11], represent a powerful research tool. In particular, meta-analyses of transcriptomic studies can uncover consensus patterns of expression across multiple studies, due to their capacity for integration and greater statistical power than individual studies.

To identify genes involved in sex-related differences in MS, we performed a systematic review and selection of transcriptomic studies of MS from the GEO and ArrayExpress databases [11, 12] that included information on patient sex. We analyzed a total of 474 transcriptomes from nine studies (189 females with MS, 109 control females, 82 males with MS, 94 control males). Given that MS presents neurodegenerative and autoimmune features, the systematic review included studies focused on brain tissue and blood. The latter could reflect the sex-differential autoimmune profiles offering a means of discovering non-invasive biomarkers in blood. We would like to point out that we have analyzed MS disease without distinguishing subtypes since this information is not provided in many of the studies. The biomarkers thus found may have a diagnostic value to discriminate the pathological state vs physiological state in both sexes. Comparisons between males and females highlighted genes involved in MS that were common to both sexes, as well as those MS-associated genes that are unique to each sex. Subsequent functional enrichment analyses of the brain tissue data provided vital information about MS-associated biologic processes in males and females (mainly immune features) that allowed a better understanding of sex differences in MS. The identified molecular mechanisms may have important clinical implications for the prevention and treatment of MS and could potentially favor the development of novel, personalized diagnostic and therapeutic applications.

## 4.3 Materials and methods

### 4.3.1 Literature Review and Study Selection

The systematic review was performed by searching for the keyword "multiple sclerosis" on the GEO and ArrayExpress [11, 12] transcriptomic databases. The reviewed period was from January 2002 (generation of the public repositories) to October 2020. Results were filtered by i) organism: "*Homo sapiens*", ii) study type: "expression profiling by array" or "expression profiling by high throughput sequencing," and iii) sample count: at least 12 samples. PRISMA



Statement guidelines were followed to ensure the quality and transparency of the systematic review [13].

The following exclusion criteria were applied: i) studies not based on MS, ii) experimental design other than patients vs. controls, and iii) the absence of information on patient sex. In addition, studies with patients receiving treatment or who were immunosuppressed were excluded.

#### ***4.3.2 Bioinformatic Analyses***

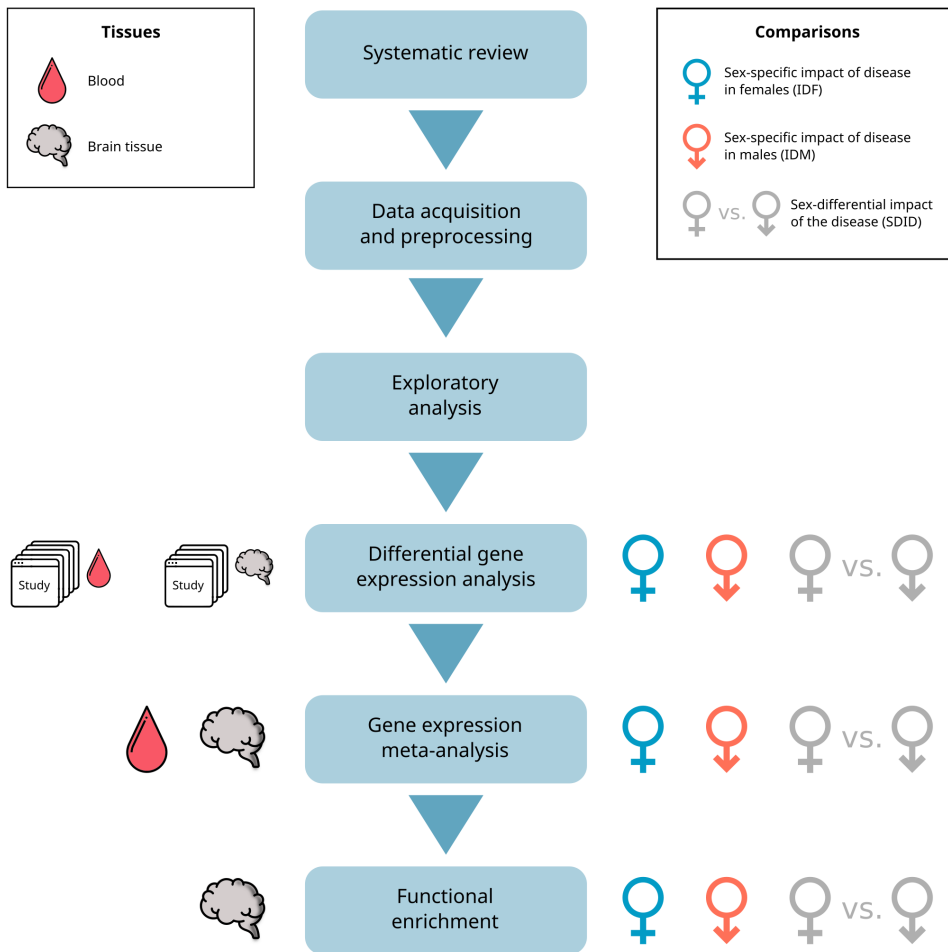
A defined analytical strategy was followed for each of the selected studies: i) data download and normalization, ii) exploratory analysis, iii) DGE analysis, iv) integration of DGE results using meta-analysis techniques (considering tissue of origin of samples to obtain robust results across all studies), and v) functional enrichment analysis of meta-analyses results (Fig. 4.1). Bioinformatics analysis was carried out using R 4.1.2 [14]; the packages used and their version number are described in Supplementary Table 1 and the following Zenodo repository (<http://doi.org/10.5281/zenodo.7738725>).

#### ***4.3.3 Data Acquisition and Preprocessing***

First, normalized data of selected studies were downloaded and standardized as follows: conversion and updating (if necessary) gene names to the symbols approved by the [15, 16], followed by the calculation of the median expression in the case of multiple values for a gene. The labels used for the sex of the patients were homogenized, and all individuals were grouped into healthy or affected by MS. Finally, the presence of possible batch effects or anomalous behavior of the data was analyzed by scanning unexplained groupings of the study samples using clustering and principal component analysis (PCA). The batch effects and paired samples present in the studies were incorporated into the model used in the differential expression.

#### ***4.3.4 Comparisons***

By creating a unique linear model, multiple comparisons were performed to detect sex differences in MS-associated gene expression and to discover altered functions. The impact of the disease in females (IDF) was analyzed by comparing female MS patients to healthy female controls (Female.MS - Female.Ctrl), while the impact of the disease in males (IDM) was analyzed by comparing male MS patients to healthy male controls (Male.MS - Male.Ctrl). Significant



**Fig. 4.1.** Flowchart describing the work performed, indicating tissues analyzed and comparisons performed in each step.

features in both comparisons represent potential disease biomarkers regardless of sex, while those significant in just one comparison can be considered sex-specific biomarkers of MS. Additionally, we analyzed the sex-differential impact of the disease (SDID) through an advanced comparison (Female.MS - Female.Ctrl) - (Male.MS - Male.Ctrl) to find differences in disease impact by sex (Fig. 4.1). As a measure of the effect size, the base 2 logarithm of the fold

change (LFC) was used for DGE analysis, and the logarithm of the odds ratio (LOR) was used for functional enrichment analysis.

#### ***4.3.5 Differential Gene Expression and Meta-analyses***

3 DGE analyses were performed for each study to detect differences in gene-level expression between groups following comparisons IDF, IDM, and SDID as described above. All comparisons were performed using the R package limma [17]. After calculating the differential expression statistics,  $p$ -values were adjusted using the Benjamini & Hochberg (BH) method [18].

To robustly integrate the DGE results of each study, the meta-analysis technique was performed for each comparison (IDF, IDM, and SDID) and available tissue of origin (blood and brain tissue), which led to a total of 6 meta-analyses. The meta-analyses were performed following the methodology previously described in García-García [19]. Briefly, the metafor R package [20] combined the DGE summary statistics from each of the individual studies, using a random-effects model for each gene, which is particularly suitable when there exist diverse sources of variability [21]. For the overall calculation of the LFC, this method gives greater weight to studies whose results were less variable. Subsequently, better integration of the selected studies and statistically more robust results were achieved by incorporating the variability between experiments in the random-effects model. The suitability of this method was evaluated by assessing the heterogeneity and influence of each study on the overall model and cross-validation techniques. The results of these analyses are available on the web platform and in the Zenodo repository.

The meta-analyses calculate a pooled measure of the expression effect across all studies for each gene (combined LFC and its 95% confidence interval) and provide a BH-adjusted  $p$ -value [18]. The significance threshold was set to 0.05. The Open Targets Platform [22] (release 21.09.5) was used to evaluate the prior known association of genes with MS.

#### ***4.3.6 Functional Enrichment Analysis***

To detect sex-specific and sex-differential functions or pathways associated with MS, a gene set analysis (GSA) [23] of brain tissue meta-analysis was performed for each of the 3 comparisons (IDF, IDM, and SDID). The GSA was performed only in brain tissue, given the high number of dysregulated genes found in this tissue. First, genes were ordered according to the  $p$ -value and sign of the statistic obtained in the DGE, and GSA was performed using the logistic regression

model implemented in the *mdgsa* R package [24]. The functional annotation required for GSA was obtained from the following databases: Gene Ontology (GO) [25] in the case of biological processes (BP) and Kyoto Encyclopedia of Genes and Genomes (KEGG) Pathway [26] for pathways. Due to their hierarchical structure, the gene annotations were propagated with GO terms applied in the *mdgsa* package to inherit the annotations of ancestor terms. Next, excessively specific or generic annotations were filtered out (blocks smaller than 10 or larger than 500 genes). Finally, functions with a Benjamini & Yekutieli (BY) adjusted  $p$ -value [27] under 0.05 were considered significant. To obtain a global view of sex-specific alterations in GO biological processes, the GSA results of each comparison were grouped into categories defined by hierarchically superior GO terms, counting the number of processes within each of them.

In order to deep in the unveiled biological sex-related alterations, the amounts of immune system cells in each sample of the brain tissue studies were estimated using the Cibersort deconvolution method (implemented in the *immunedeconv* R package) [28, 29]. Cell estimations were aggregated according to their lymphoid (B cell naive, B cell memory, B cell plasma, T cell CD8+, T cell CD4+ naive, T cell CD4+ memory resting, T cell CD4+ memory activated, T cell follicular helper, T cell regulatory, T cell gamma delta, NK cell resting and NK cell activated) or myeloid (Monocyte, Macrophage M0, Macrophage M1, Macrophage M2, Myeloid dendritic cell resting, Myeloid dendritic cell activated, Mast cell activated, Mast cell resting, Eosinophil and Neutrophil) lineage belonging. The mean cell amounts of each immune lineage, within each study, was obtained for each of the study groups (Female.MS, Female.Ctrl, Male.MS and Male.Ctrl), and within-study variation in these lineages was assessed by SDID comparison (Female.MS - Female.Ctrl) - (Male.MS - Male.Ctrl).

#### 4.3.7 *Metafun-MS Platform*

All data and results generated in the different steps of the meta-analysis are available on the Metafun-MS platform, which is freely accessible to any user and allows the confirmation of the results described in this manuscript and the exploration of other results of interest. The front-end was developed using the Angular Framework. All graphics used in this web resource have been implemented with Plot.ly [30] except for the exploratory analysis cluster plot generated with ggplot2 [31].

This easy-to-use resource is divided into 5 sections: i) the summary of analysis results in each phase, followed by detailed results for the ii) exploratory analysis, iii) DGE, iv) meta-analysis, and v) functional profiling for each study. The

user can interact with the web tool through graphics and tables and explore information associated with specific genes or biological functions. In addition, the detailed results of all analyses are available in the public repository Zenodo.

## 4.4 Results

### 4.4.1 *Systematic Review and Exploratory Analysis*

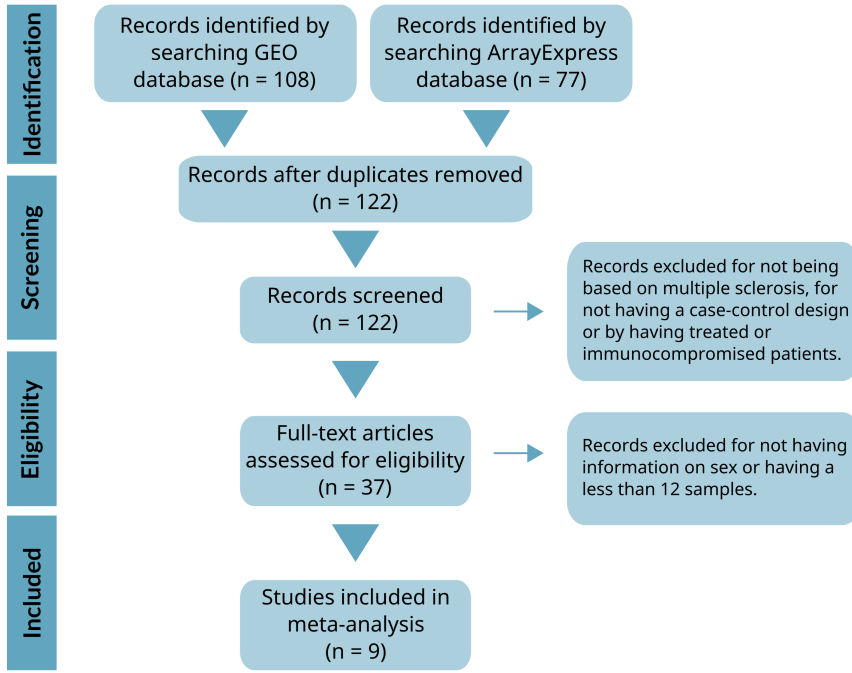
Our systematic review screened 122 publications, which rendered a total of 9 studies eligible for our analyses: 5 studies with blood samples and 4 studies with brain samples (Fig. 4.2 and Table 4.1). Notably, we discarded 35% of the studies evaluated for not including sex information or only including individuals of a single sex.

The 9 selected studies contained 474 samples - 271 MS patients and 203 healthy control patients (62.9% female samples and 37.1% male samples). Table 4.2 describes the distribution of the samples in detail. All the information available in GEO on the studies is available in the supplementary material.

Exploratory analyses revealed the absence of biases and anomalous behavior in the samples from all studies except for GSE62584, in which we detected a batch effect produced by the 4 cell types analyzed in that study (Supplementary Fig. 1). To control undesired variability coming from this covariate, we introduced it in the differential expression model as a blocking factor.

### 4.4.2 *Differential Gene Expression Analysis*

As described in the Materials and Methods section, we evaluated 3 comparisons for each study: IDF, IDM, and SDID. Positive values indicate that the variable has a higher mean in the first element of the comparison, while negative statistics indicate higher mean values in the second element. Positive LFC or LOR values indicate higher values in MS patients (MS females and MS males in IDF and IDM comparisons) and, therefore, gene upregulation or function overrepresentation in MS. Negative LFC or LOR indicates higher values in control samples (healthy control females and healthy control males), and therefore gene downregulation or function underrepresentation in MS. The SDID comparison, for which each element is a comparison itself, focuses on finding differences between IDF and IDM, indicating a sex difference in the genes whose expression is associated with MS. Thus, a positive statistic could indicate: i) an upregulation in IDF together with a downregulation in IDM,



**Fig. 4.2. PRISMA diagram of the systematic review.**

ii) an upregulation in IDF without significant changes in IDM, iii) an upregulation in IDF and IDM but higher in IDF, iv) a downregulation in IDM without a significant change in IDF, or v) a downregulation in IDF and IDM, but higher in IDM (Supplementary Fig. 2A-E). A negative statistic associated with the opposite of these cases (Supplementary Fig. 2F-J). Because of the combination of possibilities in SDID comparison, the particular behavior of each feature across the 4 groups must be assessed a posteriori, examining IDF and IDM comparisons. For simplicity, features with positive LFC or LOR will be referred to as features with an incremental change in females or increased in females, and features with a negative LFC or LOR as features with an incremental change in males or increased in males. Table 4.3 summarizes the number of significantly differentially-expressed genes, while the associated Zenodo repository and Metafun-MS web platform contain the full summary statistics for each comparison.

**Table 4.1. Description of selected studies.** Platform, tissue, and MS subtype for the samples from 9 studies. The characteristics of the studies are shown as available in the GEO database.

| GEO accession  | Platform                                              | Tissue                                                                       |
|----------------|-------------------------------------------------------|------------------------------------------------------------------------------|
| GSE37750 [32]  | Affymetrix Human Genome U133 Plus 2.0 Array           | Plasmacytoid dendritic cells                                                 |
| GSE41848 [33]  | Affymetrix Human Exon 1.0 ST Array                    | Whole blood                                                                  |
| GSE41849 [33]  | Affymetrix Human Exon 1.0 ST Array                    | Whole blood                                                                  |
| GSE41890 [34]  | Affymetrix Human Gene 1.0 ST Array                    | Peripheral blood leukocytes                                                  |
| GSE62584 [35]  | Affymetrix Human Genome U133A 2.0 Array               | CD4+ blood cells, CD8+ blood cells, CD14+ blood cells, and CD19+ blood cells |
| GSE108000 [36] | Agilent-026652 Whole Human Genome Microarray 4x44K v2 | Brain                                                                        |
| GSE111972 [37] | Illumina NextSeq 500                                  | Corpus callosum and occipital cortex                                         |
| GSE131281 [38] | Illumina HumanHT-12 V4.0 expression beadchip          | Grey matter                                                                  |
| GSE135511 [39] | Illumina HumanRef-8 v3.0 expression beadchip          | Motor cortex                                                                 |

**Table 4.2. Distribution of samples by sex and experimental group.** Tissue and number of samples of female, male, MS and control groups per experiment.

| GEO accession    | Tissue       | Females |         | Males |         | Total |
|------------------|--------------|---------|---------|-------|---------|-------|
|                  |              | MS      | Control | MS    | Control |       |
| <b>GSE37750</b>  | Blood        | 7       | 6       | 2     | 2       | 17    |
| <b>GSE41848</b>  | Blood        | 41      | 28      | 13    | 10      | 92    |
| <b>GSE41849</b>  | Blood        | 12      | 13      | 9     | 9       | 43    |
| <b>GSE41890</b>  | Blood        | 6       | 12      | 8     | 12      | 38    |
| <b>GSE62584</b>  | Blood        | 16      | 20      | 8     | 16      | 60    |
| <b>GSE108000</b> | Brain tissue | 22      | 7       | 8     | 3       | 40    |
| <b>GSE111972</b> | Brain tissue | 5       | 9       | 10    | 7       | 31    |
| <b>GSE131281</b> | Brain tissue | 58      | 11      | 6     | 31      | 106   |
| <b>GSE135511</b> | Brain tissue | 22      | 3       | 18    | 4       | 47    |
| <b>Total</b>     |              | 189     | 109     | 82    | 94      | 474   |

**Table 4.3. The number of significantly differentially-expressed genes by study (and tissue). Description of significant genes by comparisons: IDF (impact of the disease in females), IDM (impact of the disease in males), and SDID (sex-differential impact of the disease). Significant genes are separated according to the sign of their log fold-change (LFC). The final column reports the number of genes assessed in each study.**

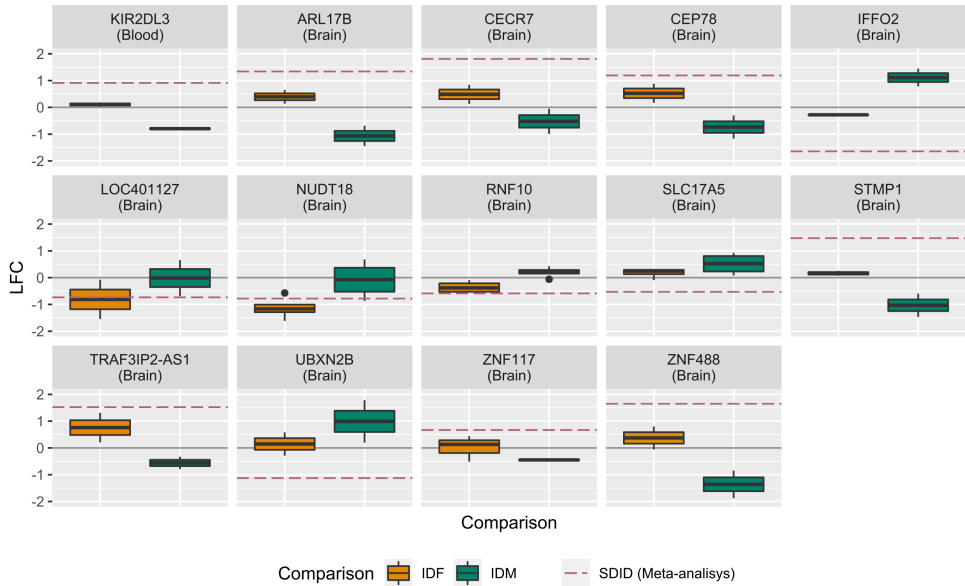
| GEO accession | Tissue       | COMPARISON |        |        |        |        |        |        |        | Analyzed genes |
|---------------|--------------|------------|--------|--------|--------|--------|--------|--------|--------|----------------|
|               |              | SDID       |        | IDF    |        | IDM    |        |        |        |                |
|               |              | LFC >0     | LFC <0 | LFC >0 | LFC <0 | LFC >0 | LFC <0 | LFC >0 | LFC <0 |                |
| GSE37750      | Blood        | 0          | 0      | 0      | 0      | 0      | 0      | 0      | 0      | 22,899         |
| GSE41848      | Blood        | 0          | 0      | 23     | 9      | 0      | 0      | 0      | 0      | 17,479         |
| GSE41849      | Blood        | 0          | 0      | 0      | 0      | 0      | 0      | 0      | 0      | 17,479         |
| GSE41890      | Blood        | 15         | 15     | 17     | 4      | 93     | 3      | 25     | 3      | 19,914         |
| GSE62584      | Blood        | 145        | 356    | 0      | 0      | 55     | 25     | 2209   | 0      | 13,277         |
| GSE108000     | Brain tissue | 154        | 230    | 2038   | 3114   | 1811   | 2209   |        |        | 21,586         |
| GSE111972     | Brain tissue | 0          | 0      | 0      | 0      | 0      | 0      | 0      | 0      | 20,846         |
| GSE131281     | Brain tissue | 4          | 0      | 15     | 8      | 0      | 3      |        |        | 20,923         |
| GSE135511     | Brain tissue | 0          | 0      | 12     | 0      | 100    | 114    |        |        | 15,434         |



#### 4.4.3 Meta-analysis of DGE Profiles

The SDID meta-analyses in the different tissues reported 14 genes that displayed significant alterations between sexes (*KIR2DL3* in blood; *ARL17B*, *CECR7*, *CEP78*, *IFFO2*, *LOC401127*, *NUDT18*, *RNF10*, *SLC17A5*, *STEMP1*, *TRAF3IP2-AS1*, *UBXN2B*, *ZNF117*, *ZNF488* in brain tissue). We also found that specific genes displayed significance for IDF and/or IDM comparisons, which allowed us to detect the change between sexes and the alterations between healthy and diseased individuals within each sex. We will discuss the alterations found in each tissue in the following sections.

We have made available the complete results of the meta-analyses in the Zenodo repository, the supplementary materials, and on the Metafun-MS web platform. In summary, Fig. 4.3 shows the pattern of LFC change between sexes (SDID, dashed line), in females (IDF, orange), and males (IDM, green). In addition, Table 4.4 indicates the number of significantly altered genes for each tissue and comparison.



**Fig. 4.3. Boxplot of LFC values for the DGE in the SDID comparison.** Each boxplot indicates the log fold-change (LFC) values of a gene in the individual differential gene expression analyses. The dashed purple line indicates the LFC value of the meta-analysis for the SDID comparison. The gray line (at a value of 0) indicates the absence of change. The tissue in which the gene is significant is indicated in parentheses under the gene name.

**Table 4.4. Results of meta-analyses by comparison and tissue.** Description of significant genes by comparisons: IDF (Impact of the disease in females), IDM (Impact of the disease in males), and SDID (Sex-differential impact of the disease). Significant genes are grouped according to the sign of their log fold-change (LFC).

| Tissue              | COMPARISON |        |        |        |        |        |
|---------------------|------------|--------|--------|--------|--------|--------|
|                     | SDID       |        | IDF    |        | IDM    |        |
|                     | LFC >0     | LFC <0 | LFC >0 | LFC <0 | LFC >0 | LFC <0 |
| <b>Blood</b>        | 1          | 0      | 5      | 4      | 27     | 5      |
| <b>Brain tissue</b> | 7          | 6      | 33     | 57     | 29     | 43     |

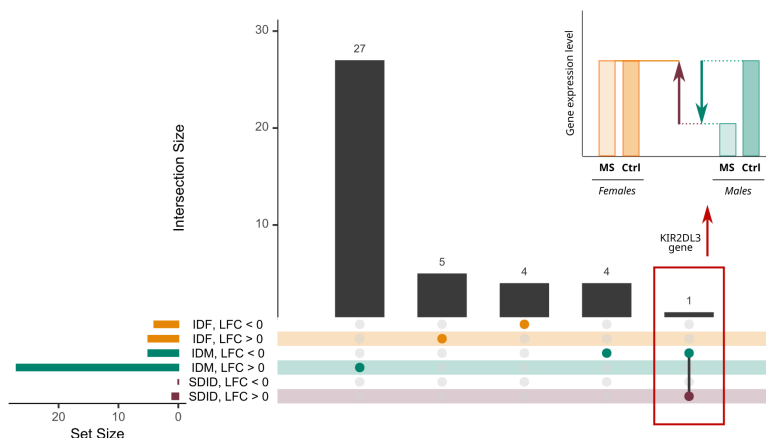
#### 4.4.4 Meta-analysis of DGE in Blood

In blood, the DGE meta-analysis only identified the *KIR2DL3* gene, which encodes for a transmembrane glycoprotein expressed by natural killer cells and subsets of T cells, as significantly altered in the SDID comparison. In particular, *KIR2DL3* displays increased expression in MS females relative to MS males (LFC > 0). Delving into its pattern for each sex, the IDM comparison revealed lower *KIR2DL3* expression in MS males than in healthy control males but did not show a significant change in the IDF comparison (Fig. 4.3 and Fig. 4.4, red box). This is a special case of Supplementary Fig. 2D.

The IDF and IDM comparisons reported 40 significant sex-specific DGE genes not significant in the SDID comparison (Fig. 4.4 and Table 4.5). Interestingly, some genes have known associations with MS in the Open Targets Platform (Table 4.5, in bold), while other genes, such as *PSMB2*, *SRP14*, *TESC*, and *KAT7*, have known associations with neurodegenerative diseases, the innate immune system, neuroprotective effects, and T cell development, respectively [40, 41, 42, 43].

#### 4.4.5 Meta-analysis of DGE in Brain Tissue

In brain tissue, our meta-analysis detected 13 significant genes in the SDID comparison (Table 4.6). Of these 13, 6 genes (*RNF10*, *IFFO2*, *SLC17A5*, *UBXN2B*, *ZNF117*, and *ZNF488*) have previous links to neurodegenerative diseases, while *IFFO2* participates in muscle control [45, 46, 47, 48, 49, 50] and 5 other genes (*CECR7*, *STMP1*, *TRAF31P2-AS1*, and *LOC401127*) [51, 52, 53, 54] have roles in inflammation, bone, and autoimmune diseases. The remaining two genes - *CEP78* and *NUDT18* - encode for a centrosomal protein and a hydrolase of the Nudix superfamily, respectively. Interestingly, the



**Fig. 4.4. Upset plot of meta-analysis results in blood studies.** The results of each comparison are separated according to the LFC sign. Horizontal bars indicate the number of significant genes in each comparison (1 specific color for each comparison). The vertical bars indicate the genes included in the intersection of the groups denoted with a colored dot underneath. A single colored dot under a bar indicates that the genes are specific to this group. The upper right corner bar plot indicates the case of the *KIR2DL3* gene, this plot represents gene expression levels in each analyzed group and, purple and green arrows indicate significant differences in SDID and IDM comparisons, respectively.

*CECR7* gene displayed significance in all 3 comparisons (Fig. 4.5, red box and upper right corner), being overexpressed in MS females and healthy males.

We detected 90 and 72 significant DGE genes in the IDF and IDM comparisons, respectively (Fig. 4.5). In the IDF and IDM comparisons, 9 genes displayed the same pattern of change, suggesting that they may be sex-independent disease markers (Fig. 4.5). Table 4.6 describes all significant genes in brain tissue, indicating those associated with MS according to the Open Targets Platform (in bold) and those common to the IDF and IDM comparisons (underlined). We display a functional enrichment analysis in the following sections, given the high number of genes involved in IDF and IDM comparisons.

**Table 4.5. Meta-analysis results of blood samples.** Significant genes for SDID (sex-differential impact of the disease), IDF (impact of the disease in females), and IDM (impact of the disease in males) comparisons. The results are separated into columns according to the LFC sign of the genes. Genes associated with MS in the Open Targets database are indicated in bold. \**KIR2DL3* was not associated with MS in Open Targets but was in a previous publication [44].

| Comparison | LFC >0                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               | LFC <0                                                                                                                                                                                        |
|------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| SDID       | <i>KIR2DL3</i> *                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |                                                                                                                                                                                               |
| IDF        | <i>DDX39A</i> , <i>LAPTM4B</i> , <i>PCK2</i> ,<br><b><i>RANGRF</i></b> , <b><i>SERPINF1</i></b><br><br><i>B2M</i> , <i>COMMD9</i> , <i>COX6B1</i> ,<br><i>COX8A</i> , <i>GLA</i> , <b><i>GRIK4</i></b> , <i>MIR6773</i> ,<br><i>MRPL23</i> , <i>MRPS16</i> , <i>NAA38</i> ,<br><b><i>NDUFA2</i></b> , <b><i>NDUFS7</i></b> , <i>OSER1</i> ,<br>IDM <i>PSMB2</i> , <i>RP11-79P5.2</i> , <i>RPS19</i> ,<br><b><i>SOCS3</i></b> , <i>SRP14</i> , <i>SSNA1</i> , <i>SSR4</i> ,<br><i>TESC</i> , <i>TMEM222</i> , <i>TMX4</i> ,<br><i>TXNL4A</i> , <i>UXT</i> , <i>WDR83OS</i> ,<br><i>ZNF816-ZNF321P</i> | <i>FNBP1</i> , <i>KAT7</i> , <i>KMT2C</i> ,<br><i>SETD5</i><br><br><br><br><br><br><br><i>KIR2DL3</i> *, <i>LINC00965</i> ,<br><i>LOC101930363</i> ,<br><i>LOC102725213</i> ,<br><i>TCAF1</i> |

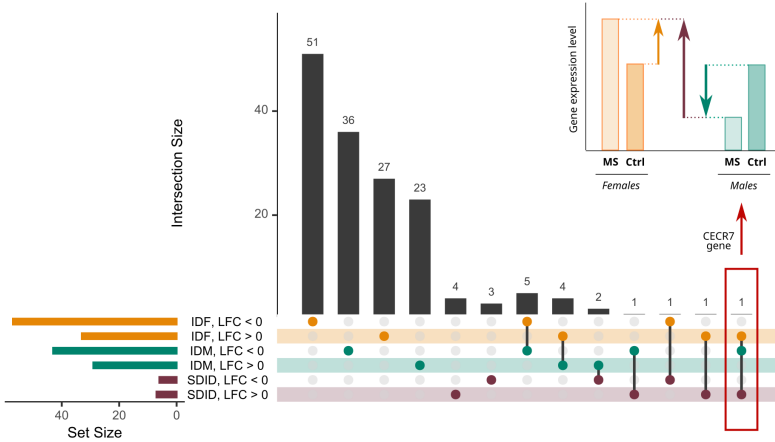
#### 4.4.6 Functional Enrichment of the Brain-Tissue DGE Meta-analysis

We performed a GSA on the genes altered in each comparison (IDF, IDM, and SDID) of the brain tissue meta-analyses to provide a functional and integrative perspective of these genes in Biological Processes GO terms and KEGG pathways. In the SDID comparison, we detected significant (adj.  $p$ -value = 0.008) alteration of the "*Staphylococcus aureus* infection" KEGG pathway (hsa05150), which includes genes such as MHC class II antigen and Interleukin 10. In the IDF and IDM comparisons, we found significant enrichment for both BP GO terms and KEGG pathways (Fig. 4.6). We detected sex-specific elements and elements common to both sexes in both ontologies. Detailed results of the functional enrichment can be found in the supplementary material, the Zenodo repository, and the metafun-MS web platform.

To summarize the high number of significant BP GO terms in the IDF and IDM comparisons, we classified them into 15 manually selected parental GO classes (hierarchically superior GO terms) (Fig. 4.7). We classified BP GO terms with no ancestor among the selected parental GO classes as "Other". Interestingly, the number of altered functions with LOR > 0 is notably higher than the number of altered functions with LOR < 0 in both IDM and IDF

**Table 4.6. Meta-analysis results from brain tissue samples.** Significant genes for SDID (sex-differential impact of the disease), IDF (impact of the disease in females), and IDM (impact of the disease in males) comparisons. The results are grouped into columns according to the LFC sign of the genes. Genes associated with MS in the Open Targets database are indicated in bold. Genes altered in the same direction in the IDF and IDM comparisons are underlined.

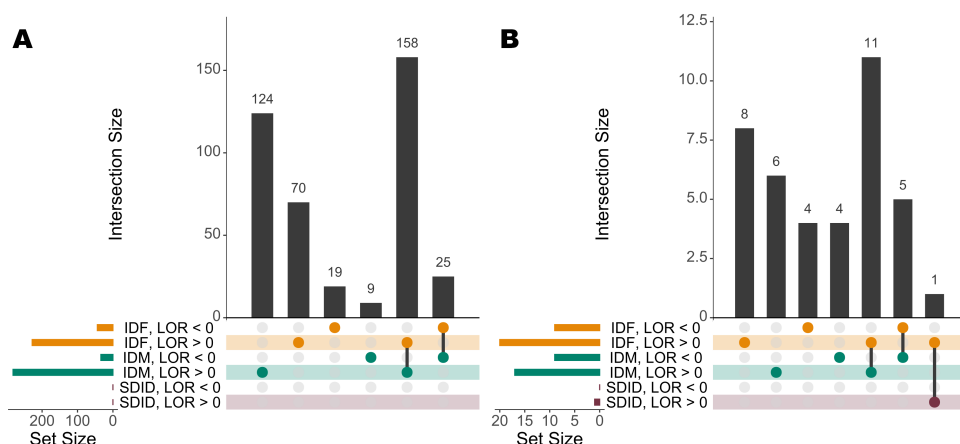
| Comparison | LFC >0                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                | LFC <0                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |
|------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| SDID       | <i>ARL17B</i> , <i>CECR7</i> , <i>CEP78</i> ,<br><i>STMP1</i> , <i>TRAF3IP2-AS1</i> ,<br><i>ZNF117</i> , <i>ZNF488</i>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                | <i>IFFO2</i> , <i>LOC401127</i> , <i>NUDT18</i> ,<br><i>RNF10</i> , <b><i>SLC17A5</i></b> , <i>UBXN2B</i>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |
| IDF        | <i>BMP6</i> , <b><i>CCNA2</i></b> , <i>CECR7</i> ,<br><i>CEP78</i> , <i>DNAJB5</i> , <i>ERI1</i> ,<br><i>G2E3</i> , <i>GAS2L3</i> , <i>HCG4</i> ,<br><b><i>HLA-DRB1</i></b> , <b><i>HLA-DRB5</i></b> ,<br><i>IGLL5</i> , <i>LDAH</i> , <i>LOC729970</i> ,<br><i>MGAM</i> , <i>OLFML2B</i> , <b><i>RTP4</i></b> ,<br><i>SFT2D2</i> , <i>STBD1</i> ,<br><i>STON1-GTF2A1L</i> , <i>STYXL1</i> ,<br><i>SUPT3H</i> , <i>TCP10L</i> , <i>TDRD3</i> ,<br><b><i>TLR8</i></b> , <i>TMEM140</i> , <i>TTC17</i> ,<br><i>TXLNB</i> , <b><i>UQCC2</i></b> , <i>XRCC4</i> ,<br><u><i>ZFAS1</i></u> , <i>ZNF615</i> , <i>ZSCAN29</i> | <i>ANKRD26P3</i> , <i>ATP1A2</i> , <b><i>BMPER</i></b> ,<br><i>C19orf47</i> , <i>C2CD4C</i> , <i>CATSPERZ</i> ,<br><i>CCDC57</i> , <i>CES2</i> , <i>CMYA5</i> , <i>DGCR9</i> ,<br><i>EBF4</i> , <i>ERCC2</i> , <b><i>FAM156B</i></b> , <i>FBXL6</i> ,<br><i>FGFBP3</i> , <i>GCAT</i> , <b><i>HBB</i></b> , <i>HOXA4</i> ,<br><i>HPDL</i> , <i>KRTAP10-9</i> , <b><i>KRTAP5-7</i></b> ,<br><i>LINGO3</i> , <i>LOC100130093</i> ,<br><i>LOC100130370</i> , <i>LOC100288911</i> ,<br><i>LOC285696</i> , <i>LOC339529</i> ,<br><i>LOC728743</i> , <i>LRFN1</i> , <i>MAGED4B</i> ,<br><b><i>MLXIPL</i></b> , <i>NLRP2</i> , <i>NOXA1</i> ,<br><i>NPIPB15</i> , <i>NUDT18</i> , <i>NUTM2F</i> ,<br><i>ODAD4</i> , <i>OVGP1</i> , <i>PET100</i> ,<br><i>PKI55</i> , <i>PMM2</i> , <i>PRDM12</i> ,<br><i>SAXO1</i> , <i>SBF1P1</i> , <i>SNHG20</i> ,<br><i>SNHG28</i> , <i>SP9</i> , <b><i>SREBF1</i></b> , <i>STKLD1</i> ,<br><i>TADA2B</i> , <i>TESPA1</i> , <i>TMC4</i> , <i>TPBG</i> ,<br><i>TRABD2A</i> , <i>VPREB3</i> , <i>XXYLT1</i> ,<br><i>ZNF839</i> |
| IDM        | <i>ATP13A3</i> , <i>B4GALT1</i> , <i>BAG2</i> ,<br><b><i>CASP7</i></b> , <i>CR1L</i> , <i>CTAGE4</i> , <i>EFNA1</i> ,<br><b><i>HBEGF</i></b> , <i>HS3ST3B1</i> , <i>HYAL1</i> ,<br><i>IFFO2</i> , <i>LIMS4</i> , <i>LOC100129617</i> ,<br><i>LOC100506136</i> , <i>LOC401127</i> ,<br><i>LOC729970</i> , <i>MIR100HG</i> , <b><i>MPZL2</i></b> ,<br><i>PRO0628</i> , <i>RBM14-RBM4</i> , <i>RIC1</i> ,<br><i>SNORA59B</i> , <i>ST7</i> , <i>STBD1</i> , <i>STON1</i> ,<br><i>TMEM140</i> , <i>TMEM163</i> , <b><i>TRPM8</i></b> ,<br><u><i>ZFAS1</i></u>                                                              | <i>ACTL6B</i> , <i>ARIH2OS</i> , <i>C2orf72</i> ,<br><i>C8orf44-SGK3</i> , <i>CA5BP1</i> , <i>CECR7</i> ,<br><i>COQ4</i> , <i>DEFA3</i> , <i>DGCR10</i> , <i>DGCR6L</i> ,<br><i>DGCR9</i> , <b><i>DHRS9</i></b> , <i>DOLK</i> , <i>EPHA7</i> ,<br><i>FAM66D</i> , <i>FLJ32063</i> , <i>GPR89A</i> ,<br><i>GTF2IRD2</i> , <i>JMJD7</i> , <i>LINC00685</i> ,<br><i>LOC100652758</i> , <i>LOC338963</i> ,<br><i>LOC644189</i> , <i>LOC730183</i> , <i>LOH12CR2</i> ,<br><i>LPP-AS2</i> , <i>MDP1</i> , <i>NIP7</i> , <i>OR7D2</i> ,<br><i>PHF24</i> , <i>PIPOX</i> , <i>PKI55</i> , <i>RWDD3</i> ,<br><i>SBF1P1</i> , <i>SERF1A</i> , <i>SNHG20</i> , <i>SYNE4</i> ,<br><i>TADA2B</i> , <i>TRIM66</i> , <i>TTC30A</i> ,<br><i>URB1</i> , <i>ZNF117</i> , <i>ZNF771</i>                                                                                                                                                                                                                                                                                     |



**Fig. 4.5. Upset plot of the meta-analysis results in brain tissue samples.** The results of each comparison are separated according to the LFC sign. Horizontal bars indicate the number of significant genes in each comparison (a specific color for each comparison). The vertical bars indicate the genes included in the intersection of the groups denoted with a colored dot underneath. A single colored dot under a bar indicates that the genes are specific to this group. The upper right corner bar plot indicates the case of the *CECR7* gene, this plot represents gene expression levels in each analyzed group and; purple, orange and green arrows indicate significant differences in SDID, IDF and IDM comparisons, respectively.

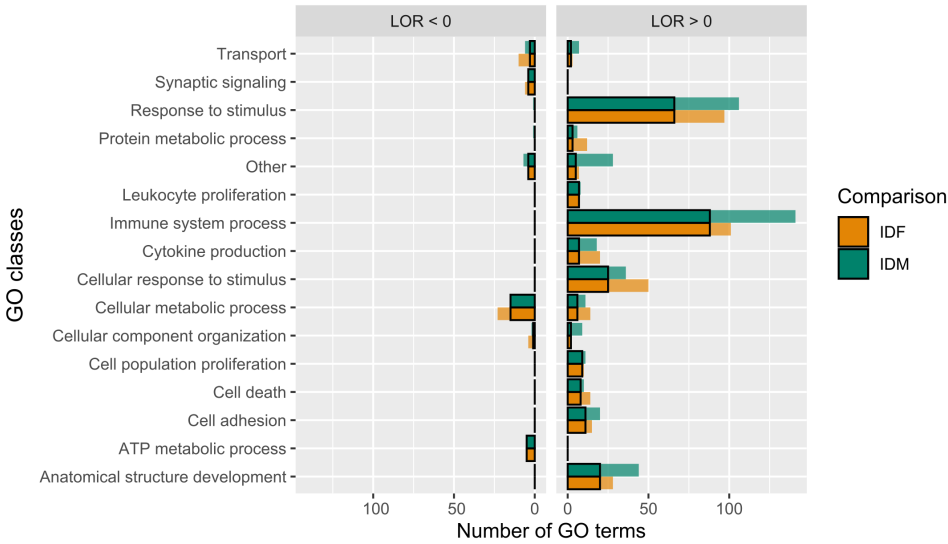
comparisons. In MS, the parental GO classes with the highest number of alterations were "Immune system process", "Response to stimulus" and, to a lesser extent, "Cellular response to stimulus" and "Anatomical structure development". Regarding the immune system, MS females specifically presented altered BP GO terms related to the myeloid lineage, regulation, and production of cytokines (IL-1, 2, 6, and 10), while MS males presented alterations in the regulation and signaling of the lymphocyte lineage, mainly (Fig. 4.8). MS females also presented altered mitochondrial respiratory chain complexes and purine and glutamate metabolism, while MS males presented alterations in amine and amino acid transport and stress response to metal ions (Fig. 4.8). The results of sex-related immune lineage variations in MS using Ciber-sort showed that in 3 of the 4 brain tissue studies, there was an increase in myeloid lineage cells in MS females while in MS males there was an increase in lymphoid lineage cells (Supplementary Table 9).

As a result of the KEGG analysis, in IDF comparison we identified altered signaling pathways via Toll-like receptor and NOD-like receptor (constitute two major forms of innate immune sensors), as well as osteoclast differen-



**Fig. 4.6.** Upset plots of the functional enrichment results in brain tissue for BP GO terms (A) and KEGG pathways (B). Horizontal bars indicate the number of significant BP GO terms or KEGG pathways in each comparison (a specific color for each comparison). Results are exhibited separately for significantly upregulated ( $LOR > 0$ ) and downregulated ( $LOR < 0$ ) altered functions. The vertical bars indicate the genes included in the intersection of the groups denoted with a colored dot underneath. A colored dot under a bar indicates that the genes are specific to this group.

tiation, the p53 pathway ( $LOR > 0$ ), and metabolism pathways related to biotin, alanine, aspartate and glutamate, xenobiotics, and drugs ( $LOR < 0$ ). KEGG pathway analysis of the IDF comparison also demonstrated alterations in terms related to diseases other than MS, including Chagas disease, prostate cancer, and *Staphylococcus aureus* infection ( $LOR > 0$ ). The IDM comparison showed alterations in KEGG pathways related to cytokine-cytokine receptor interactions, the JAK-STAT signaling pathway, protein processing in the endoplasmic reticulum, olfactory transduction, ( $LOR > 0$ ) the peroxisome, calcium signaling, metabolism related to valine, leucine and isoleucine degradation, butanoate ( $LOR < 0$ ), and diseases such as small cell lung cancer and autoimmune thyroid disease ( $LOR > 0$ ) (Supplementary Table 10).



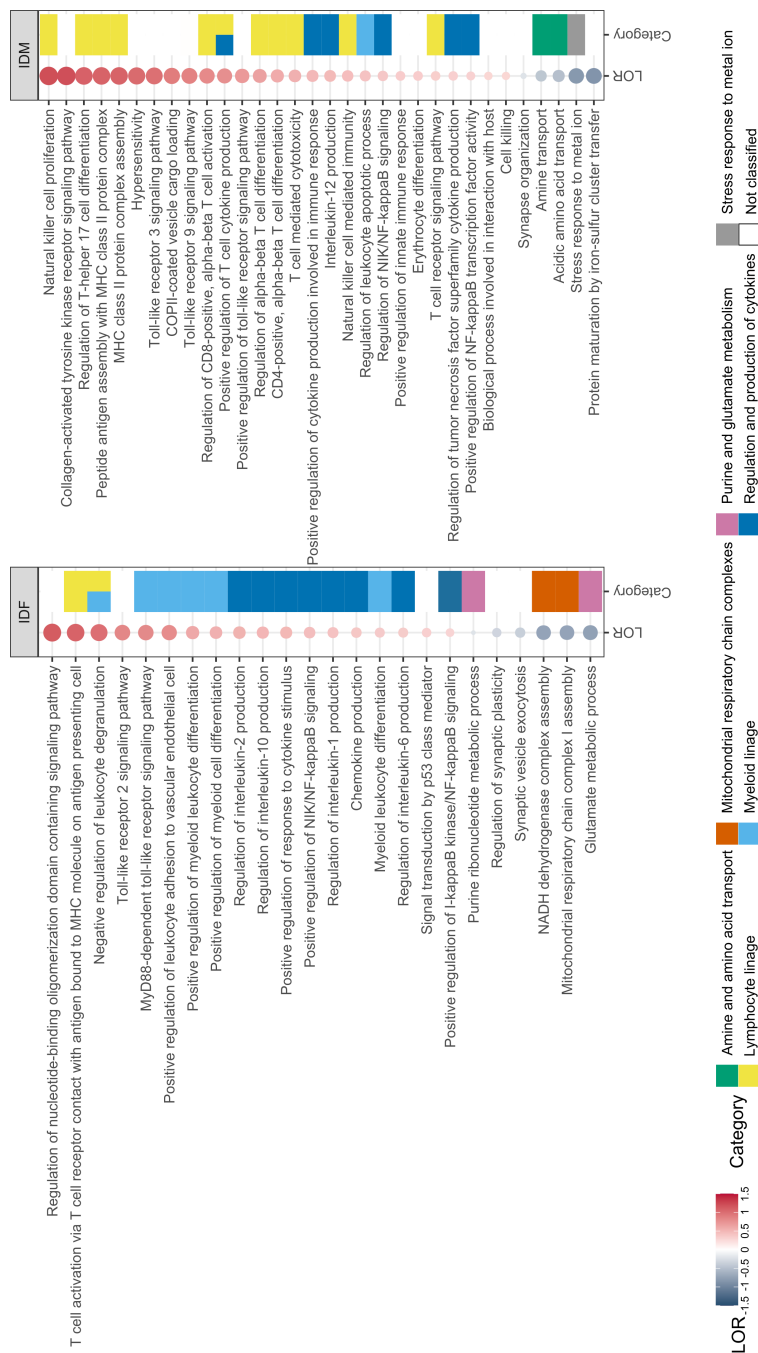
**Fig. 4.7. Bar chart of the number of significant BP GO terms in the IDF and IDM comparisons.** The horizontal bars show the number of BP GO terms, separated between higher expressed in MS patients ( $LOR > 0$ , right) and healthy patients ( $LOR < 0$ , left) in each parental GO class. The orange and green bars (IDF and IDM, respectively) show the number of GO terms within each parental GO class. Opaque bars surrounded by a black border indicate items common to the comparisons, while translucent bars indicate items unique to IDF or IDM.

#### 4.4.7 Metafun-MS platform

We have developed an open-source web platform, Metafun-MS (<http://bioinfo.cipf.es/metafun-MS>), containing a detailed description of the 9 datasets and 474 samples involved in this study. This resource includes detailed information on all genes, BP GO terms, and KEGG pathways that users can explore to identify transcriptomic profiles of interest for each study.

For each significant gene in a meta-analysis, Metafun-MS depicts the combined effect for all studies and each study's specific contribution using statistical metrics (LFC, confidence interval,  $p$ -value, and adjusted  $p$ -value) and graphical representations (forest and funnel plots). The platform also displays GSA results ( $LOR$ ,  $p$ -value, and adjusted  $p$ -value) of the GSA in brain tissue, for the GO biological processes and the KEGG pathways. This open resource aims to enable data sharing between researchers, the elaboration of innovative studies, and the discovery of new findings.





**Fig. 4.8. Dot plot of a representative selection of the specific significant BP GO terms in IDF and IDM. GO terms affected exclusively in IDF (left) or IDM (right). The color and size of the dots represent the sign and magnitude of the change (LOR). GO terms have been manually classified into functional categories, which are indicated in the "Category" column as colored squares.**

## 4.5 Discussion

A range of previous studies have integrated data from numerous sources to characterize sex-based differences in the human transcriptome [55, 56, 57, 58]; however, to our knowledge, this study represents the first meta-analysis of transcriptomic studies in MS, which provides a better understanding of the sex-related molecular mechanisms underlying MS. We have carried out an *in silico* approach using transcriptome data from male and female MS patients and healthy controls obtained from public repositories. This data acquisition strategy is intended as a rapid and cost-effective means to predict and validate the effects of relevant biological variables on human health and guide further clinical research.

Thus, we performed 6 meta-analyses using a robust random-effects model, which allows for the integration of information from individual studies to obtain greater statistical power and precision, revealing findings that cannot be easily obtained through the intersection or addition of results from individual studies. Table 4.3 depicts the diverse nature of each study and highlights the need to employ strategies such as meta-analysis to integrate information from individual studies and obtain an ultimate transcriptional consensus profile. Since MS is a chronic autoinflammatory disorder of the CNS, we performed separate meta-analyses in blood and brain tissue. Identifying sex-based differences represented our primary goal by studying genes differentially altered in the SDID comparison. Meanwhile, the IDF and IDM comparison allowed us to detect specific effects in each sex, understand the basis of the disease for each sex, and find common patterns in both sexes. The biomarkers that have been identified are discussed below, and their experimental validation could lead to their clinical use.

### 4.5.1 *Meta-analysis in blood*

Identifying biomarkers in blood will facilitate non-invasive liquid biopsy for MS diagnose, as opposed to cerebrospinal fluid analysis. Our meta-analysis in blood identified sex differences in *KIR2DL3*, a gene importantly involved in self-tolerance by regulating the activity of natural killer cells [59]. The absence of *KIR2DL3* has been linked to MS development [44, 60], and its protective role has been documented in other autoimmune diseases such as rheumatoid vasculitis [61]. In our study, *KIR2DL3* has been identified as a biomarker candidate of MS in males since it is significantly downregulated in male MS patients compared to MS controls. Its relevance in susceptibility and follow up has been described in other studies [62, 63, 64]. However, further experimen-

tal studies must be developed given its allelic diversity and that of its ligands that can importantly influence the disease [63, 65]. Unfortunately, identifying blood-based biomarkers remains challenging due to factors such as their concentration, sensitivity or specificity, the molecule availability in the blood or the patient disease stage to which should be added the study of sex variable. Additionally, the lack of standardization in analytical practices and sample handling between different laboratories could give rise to disparities. Together, these factors may explain why we only identified a single gene, *KIR2DL3*, in the SDID meta-analysis in blood.

The blood meta-analysis in IDF comparison revealed a gene dysregulation profile related to the control of immune system responses and symptoms associated with MS. [66, 67]. Among the identified genes, *RANGRF* has been related to cardiac arrhythmia in Brugada Syndrome [68]. The number of cardiac ventricular arrhythmias worsens the prognosis of MS patients by increasing risk of sudden death [69]. *SERPINF1* may influence immune and vascular injury [70]. Serpin family members have been identified in MS plaques and other neurodegenerative conditions such as Alzheimer’s disease [71] and have been related to inflammation, immunity and dementia. *FBNP1* may represent a potential biomarker, as the expression of this gene links to prognosis and level of immune infiltration in different cancers [72] and could play a similar role in MS. Interestingly, *KAT7* is crucial for the development, maintenance, and survival of T cells and is protective for auto-immune diseases [73]. This gene appeared downregulated in MS females relative to healthy females, being this study the first description of a sex association of *KAT7* with MS to the best of our understanding.

The IDM blood meta-analysis reflected the dysregulation of genes also related to inflammation, immune system, and autoimmune diseases (*B2M*, *COMMD9*, *GLA*, *PSMB2*, *SOCS3*, *SRP14*, *SSNA1*, *SSR4*, *TESC*, *UXT*, *RPS19* and *KIR2DL3*), genes related to redox activity/stress oxidative and mitochondrial function (*COX6B1*, *COX8A*, *MRPL23*, *MRPS16*, *NDUFA2*, *NDUFS7*, *OSER1*, *TMX4*, and *TXNL4A*) and genes related to neurodegenerative diseases and intellectual disabilities (*WDR83OS*, *GRIK4*, *NAA38*, *TMEM222*). Short and long non-coding (lnc)RNA genes (*MIR6773*, *LINC00965*, *LOC101930363*, and *LOC102725213*) altered in IDM could be critical for post-transcriptional regulation of other genes, while *TCAF1*, a transient receptor potential channel-associated factor, can regulate cell migration [74].

These results demonstrate that MS hallmarks such as inflammation and immunity occur in both females and males with MS, although through different

target genes in each case, which may be critical to guide medical decisions. Unfortunately, the blood meta-analyses failed to identify MS biomarkers common to males and females with MS, perhaps due to difficulties identifying stable markers in the blood, as discussed above.

#### 4.5.2 *Meta-analysis in brain tissue*

The brain is a complex and heterogeneous tissue and MS can injure different brain areas which is the cause of the variability and multiplicity of symptoms in MS. Therefore, the meta-analysis in this region included studies from different brain areas as indicated in Table 4.1. Analyses in brain tissue yielded a higher number of differentially expressed genes than in blood, reflecting the notable impact of MS in this tissue, the complexity of the brain but also the limitations to identify biomarkers in blood as commented before. The SDID brain tissue meta-analysis identified 13 differentially affected genes by sex, with specific genes related to the loss of mental functions and neural disorders, blood cell abnormalities, inflammation, and autoimmune disorders.

Note the dysregulation of *CECR7*, which has been previously implicated in immune infiltration in hepatocellular carcinoma [75] and regulates the expression of *CTLA4* [76]; meanwhile, *CEP78* is recognized as an auto-antigen in prostate cancer [77]. Thus, both genes represent potential candidates for immunotherapies. *STMP1* has a reported involvement in the NLRP3 inflammasome and Paget's disease of bone [52], while *TRAF3IP2-AS1* participates in IL-17 regulation. Interestingly, this interleukin possesses pro-inflammatory properties and contributes to immune dysregulation in other autoimmune diseases such as colitis and rheumatoid arthritis [78]. *IFFO2*, Intermediate Filament Family Orphan 2 may represent a novel IFAP (intermediate filament-associated protein) for males with MS given that the presence of IFAPs in MS plaques tissue has been previously documented [79]. The *LOC401127* lncRNA could play an essential role in MS regulation or even serve a prognostic value as other described lncRNAs in chronic musculoskeletal disorders [80]. We also highlight the protective role against oxidative stress and DNA damage of *NUDT18* (also known as *MTH3*) [81].

In the IDF and IDM meta-analyses in brain tissue, we identified genes that appeared dysregulated with the LFC in the same sense in both sexes and may represent potential MS biomarkers. Specific genes identified in the IDF and IDM comparisons have not been previously linked to MS and may represent interesting future research directions. In general, these genes are mainly involved in inflammatory and immune responses but with different connotations

in females and males. Autoimmune diseases are generally exacerbated in female patients and a more pro-inflammatory phenotype in XX mice has been demonstrated [82, 83]. In males, sex-based differences in immune responses are well-documented and influenced by multiple factors (sex chromosomes, reproductive organs, and sex hormone levels), which, in turn, influence the incidence and other aspects of autoimmune diseases [84].

Through the **functional enrichment analysis in brain tissue**, we found a single significantly dysregulated KEGG pathway in the SDID comparison: "*Staphylococcus aureus* infection" with important immune genes (MHC class II antigen and IL-10) that regulate polyclonal T cell activation, cytokine release and inflammation. This pathway also is upregulated in females with MS in the IDF comparison, further supporting the differential role of the immune system in MS between males and females. When we analyzed GO terms in the IDF and IDM comparisons, many biological processes related to inflammation and immunity displayed sex-specific dysregulation in MS patients. Specifically, the transcriptomes of females with MS represented alterations more related to myeloid lineage and pro-inflammatory environment, influenced by alterations in interleukin regulation and response (e.g., IL-1, IL-2, IL-6, IL-10, and interferon-gamma). These alterations are closely related to innate immune responses and have been previously linked to MS and autoimmune diseases [85, 86, 87, 88, 89]. In contrast, males with MS seem to exhibit transcriptome alterations related to immunomodulation of the lymphoid lineage (e.g., T cell subsets, natural killer cells, dendritic cells), affecting different features of these cells such activation, proliferation, or differentiation, which have a closer relation to adaptive immune responses [90, 91, 92]. Cibersort deconvolution analysis supported sex-based differences in functions related to the immune system, which could reflect a differential immune cell infiltration. Specifically, we observe an increase in myeloid lineage cells in females with MS and lymphoid lineage cells in males with MS. Identifying T cell clones associated with human autoimmunity remains challenging and could have critical connotations for MS risk and outcomes. Taken together, these MS-associated transcriptomic sex differences representing different components of the immune system could partly explain sex-based differences in susceptibility to MS. Pro-inflammatory cytokines, which are significantly altered in females with MS, have been implicated in autoimmune disorders such as rheumatoid arthritis [93]. Meanwhile, MS males suffer from a more rapid and worsened progression, which could be explained by more intensive participation of T cells in demyelination. The immune system is highly complex and much more so in a disease like MS, the innate and adaptive response are closely linked as well as their multiple components. In this context, our meta-analyses have found sex related differences,

but further studies must be carried out to clarify these hypotheses and help guide decisions in the design of personalized immunotherapies.

Other sex-related differences in GO terms suggested that females with MS exhibited alterations in mitochondrial respiratory chain complexes, purine, and glutamate metabolism, while males with MS exhibited alterations in amine and amino acid transport and stress response to metal ions. These biological processes have important connotations in MS - the glutamatergic system plays a vital role in neurotoxicity and MS pathogenesis [94], while amino acid metabolism affects immune homeostasis, as activated immune cells have a high demand for amino acids [94]. Remarking that the meta-analyses found many altered immune terms in the brain tissue, this result correlates with the disruption of the blood-brain barrier during MS and the presence of immune infiltrates in the brain.

KEGG pathway data revealed that females with MS displayed an upregulation in Toll-like receptor signaling pathway, NOD-like receptor signaling pathway, B cell receptor signaling pathway. Toll-like and Nod-like receptors are two major forms of innate immune sensors and provide immediate responses against pathogenic invasion or tissue injury [95, 96]. Activation of these sensors induces the recruitment of innate immune cells such as macrophages and neutrophils (myeloid lineage), which displayed dysregulation in females with MS, initiates tissue repair processes and results in adaptive immune activation. Abnormalities in these innate sensor-mediated processes may cause excessive inflammation due to hyper-responsive innate immune signaling or sustained compensatory adaptive immune activation [97]. Other KEGG pathways enriched in females with MS were those associated with chronic or autoimmune diseases (e.g., chronic myeloid leukemia, systemic lupus erythematosus, and rheumatoid arthritis), indicating that autoimmune diseases may share altered signaling pathways. We also discovered the upregulation of intestinal immune networks for IgA production and Chagas disease in MS females. Diet, microbiota gut, parasite, and viral infections are associated with chronic inflammatory diseases and could play a significant role in MS etiology [98]; meanwhile, intestinal disorders also represent common symptoms of MS [99, 100]. The KEGG pathways downregulated in females with MS include metabolism of xenobiotics by cytochrome P450, drug metabolism - cytochrome, cardiac muscle, and Alzheimer's disease, which could have significant implications in treatment responses and therapy development. In males with MS, upregulated KEGG pathways include RNA transport, cytokine-cytokine receptor interaction, protein processing in the endoplasmic reticulum, the phagosome, JAK-STAT signaling pathway, and olfactory transduction, while valine, leucine, and isoleucine

degradation represent the only downregulated pathway. RNA localization regulates gene expression, while neurological diseases such as amyotrophic lateral sclerosis and fragile X syndrome have been associated with dysregulated RNA localization in neurons [101]. Endoplasmic reticulum-associated protein modifications, folding, and degradation could mediate the pathogenesis of autoimmune disorders [102], while antigen processing and cross-presentation by dendritic cells require phagosome maturation [103]. Finally, metabolism and diet have relevant implications in MS, as in other autoimmune diseases [104], so amino acid degradation may impact MS at this level.

### 4.5.3 *Strengths and limitations*

Patient sex influences the observed pathophysiological characteristics of neurodegenerative and autoimmune diseases [9, 105], with sex differences described in MS risk and disease progression [8]. Unfortunately, sex is not frequently considered a biological variable during biomedical research; this became evident during our systematic review, where the absence of sex information prompted us to discard a substantial number of studies. Encouragingly, the sex perspective has gained recent relevance as a crucial aspect of personalized medicine and has been recognized as a necessary biological variable for inclusion in research by the National Institutes of Health [106] and the European Union [107, 108].

In spite of the impulse that is being given to the publication of open data under the FAIR principles (Findable, Accessible, Interoperable, Reusable) [109], their availability is still limited. This, in addition to the lack of homogeneity of the clinical and phenotypic data available from the individuals analyzed, has prevented us from integrating variables which could potentially contribute to disease variability, such as age, hormone concentration, or disease stage.

In the same line, lack of studies has limited the brain region specificity of the study. The brain is a heterogeneous tissue and different brain regions may have very different structure, functions and cell composition. Broadly, the brain is composed of white and gray matter. White matter is principally made of nerve fibers (axons) covered by myelin, while gray matter includes the majority of cells, such as neurons, microglia, astrocytes, and others. The gene expression in these two areas is expected to differ and also, MS may affect the brain regions differently. However, due to the lack of studies available for each region, we have been forced to perform our meta-analysis including samples from both white and gray matter. This may limit the region specificity of the results, but opens the door to a more general interpretation of the results, since the found

alterations will be common along the tissues. Also, transcriptomic studies of spinal cord tissue were not found in our systematic review, probably because access to postmortem spinal cord tissue is very complicated and not routinely performed at autopsy.

Despite all these limitations, the meta-analysis technique considers the variability between studies and is capable of robustly extracting gene expression changes across studies. However, further studies will be needed to validate sex differences found in our study.

Our meta-analyses have detected differentially altered genes and functions between sexes when comparing individuals with MS and healthy individuals, and could identify a set of biomarkers common to MS. Some of these findings have been previously associated - although not usually from a sex perspective - with MS, while others constitute a novelty. Gene dysregulations detected in blood samples could open the door to the study of these genes as sex-specific disease non-invasive biomarkers. At the functional level, results in brain tissue show alterations in different immune system components between females and males. These findings may also impact disease diagnosis and prognosis and the design of clinical trials. Furthermore, our results highlight the importance of recognizing and considering the sex factor in the study of MS.

## **4.6 Conclusions**

We have performed a gene and functional meta-analysis of 9 MS transcriptomic studies from a sex perspective, identifying sex-based differences and similarities at gene, function, and pathway levels. The results allow for a better understanding of those mechanisms crucially involved in MS and the identification of potential biomarkers, whose detailed study may improve personalized therapies. In addition, we highlight the importance of providing information on sex in studies and the need for open platforms for data sharing to allow for efficient advancements in research.



## 4.7 Additional files

Detailed results of all analyses and the versions of the R packages used can be found in the Zenodo repository: <https://zenodo.org/record/7738725>.

## 4.8 Availability of data and materials

The large volumes of data and results generated in this study are freely available through the Metafun-MS web tool: <http://bioinfo.cipf.es/metafun-MS> and the Zenodo repository: <http://doi.org/10.5281/zenodo.7738725>. This study analyzed transcriptomic data available in the Gene Expression Omnibus database with accession numbers GSE37750, GSE41848, GSE41849, GSE41890, GSE62584, GSE108000, GSE111972, GSE131281, and GSE135511.

## 4.9 Computer Code and Software

The code developed for the analyses described in this work and the software and its versions are publicly available at Zenodo: <https://doi.org/10.5281/zenodo.7738725>.

## 4.10 References

1. Dobson R, Giovannoni G. Multiple sclerosis - a review. *Eur J Neurol*. 2019;26:27-40. <https://doi.org/10.1111/ene.13819>.
2. Ford H. Clinical presentation and diagnosis of multiple sclerosis. *Clin Med (Lond)*. 2020;20:380-383. <https://doi.org/10.7861/clinmed.2020-0292>.
3. Muñoz-Culla M, Irizar H, Sáenz-Cuesta M, Castillo-Triviño T, Osorio-Querejeta I, Sepúlveda L, López de Munain A, Olascoaga J, Otaegui D. SncRNA (microRNA & snoRNA) opposite expression pattern found in multiple sclerosis relapse and remission is sex dependent. *Sci Rep*. 2016;6:20126. <https://doi.org/10.1038/srep20126>.
4. Bishop M, Rumrill PD. Multiple sclerosis: Etiology, symptoms, incidence and prevalence, and implications for community living and employment. *Work*. 2015;52:725-734. <https://doi.org/10.3233/WOR-152200>.
5. Voskuhl RR. The effect of sex on multiple sclerosis risk and disease progression. *Mult Scler*. 2020;26:554-560. <https://doi.org/10.1177/1352458519892491>.
6. Lasrado N, Jia T, Massilamany C, Franco R, Illes Z, Reddy J. Mechanisms of sex hormones in autoimmunity: focus on EAE. *Biol Sex Differ*. 2020;11:50. <https://doi.org/10.1186/s13293-020-00325-4>.
7. Thompson AJ, Baranzini SE, Geurts J, Hemmer B, Ciccarelli O. Multiple sclerosis. *Lancet*. 2018;391:1622-1636. [https://doi.org/10.1016/S0140-6736\(18\)30481-1](https://doi.org/10.1016/S0140-6736(18)30481-1).
8. Voskuhl RR, Sawalha AH, Itoh Y. Sex chromosome contributions to sex differences in multiple sclerosis susceptibility and progression. *Mult Scler*. 2018;24:22-31. <https://doi.org/10.1177/1352458517737394>.
9. Ortona E, Pierdominici M, Maselli A, Veroni C, Aloisi F, Shoenfeld Y. Sex-based differences in autoimmune diseases. *Ann Ist Super Sanita*. 2016;52:205-212. [https://doi.org/10.4415/ANN\\_16\\_02\\_12](https://doi.org/10.4415/ANN_16_02_12).
10. Pierdominici M, Maselli A, Colasanti T, Giammarioli AM, Delunardo F, Vacirca D, Sanchez M, Giovannetti A, Malorni W, Ortona E. Estrogen receptor profiles in human peripheral blood lymphocytes. *Immunol Lett*. 2010;132:79-85. <https://doi.org/10.1016/j.imlet.2010.06.003>.

11. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Holko M, Yefanov A, Lee H, Zhang N, Robertson CL, Serova N, Davis S, Soboleva A. NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res.* 2013;41:D991-D995. <https://doi.org/10.1093/nar/gks1193>.
12. Athar A, Füllgrabe A, George N, Iqbal H, Huerta L, Ali A, Snow C, Fonseca NA, Petryszak R, Papatheodorou I, Sarkans U, Brazma A. ArrayExpress update - from bulk to single-cell expression data. *Nucleic Acids Res.* 2019;47:D711-D715. <https://doi.org/10.1093/nar/gky964>.
13. Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLoS Med.* 2009;6:e1000097. <https://doi.org/10.1371/journal.pmed.1000097>.
14. R Core Team. R: A Language and environment for statistical computing. Vienna, Austria; 2021. <http://www.r-project.org/>
15. European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI). HGNC Database, HUGO Gene Nomenclature Committee (HGNC) [WWW Document]. URL: <https://www.genenames.org> (accessed 28/02/2022).
16. Tweedie S, Braschi B, Gray K, Jones TEM, Seal RL, Yates B, Bruford EA. Genenames.org: the HGNC and VGNC resources in 2021. *Nucleic Acids Res.* 2021;49:D939-D946. <https://doi.org/10.1093/nar/gkaa980>.
17. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 2015;43:e47. <https://doi.org/10.1093/nar/gkv007>.
18. 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)*. 1995;57:289-300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>.
19. García-García F. Métodos de análisis de enriquecimiento funcional en estudios genómicos. University of Valencia. 2016. <https://doi.org/10.5281/zenodo.7669523>.

20. Viechtbauer W. Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*. 2010;36:1-48. <https://doi.org/10.18637/jss.v036.i03>.
21. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials*. 1986;7:177-188. [https://doi.org/10.1016/0197-2456\(86\)90046-2](https://doi.org/10.1016/0197-2456(86)90046-2).
22. Ochoa D, Hercules A, Carmona M, Suveges D, Gonzalez-Uriarte A, Malangone C, Miranda A, Fumis L, Carvalho-Silva D, Spitzer M, Baker J, Ferrer J, Raies A, Razuvayevskaya O, Faulconbridge A, Petsalaki E, Mutowo P, Machlitt-Northern S, Peat G, McAuley E, Ong CK, Mountjoy E, Ghossaini M, Pierleoni A, Papa E, Pignatelli M, Koscielny G, Karim M, Schwartzentruber J, Hulcoop DG, Dunham I, McDonagh EM. Open Targets Platform: supporting systematic drug-target identification and prioritisation. *Nucleic Acids Res*. 2021;49:D1302-D1310. <https://doi.org/10.1093/nar/gkaa1027>.
23. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. 2005;102:15545-15550. <https://doi.org/10.1073/pnas.0506580102>.
24. Montaner D, Dopazo J. Multidimensional gene set analysis of genomic data. *PLoS One*. 2010;5:e10348. <https://doi.org/10.1371/journal.pone.0010348>.
25. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet*. 2000;25:25-9. <https://doi.org/10.1038/75556>.
26. Kanehisa M, Goto S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*. 2000;28:27-30. <https://doi.org/10.1093/nar/28.1.27>.
27. Benjamini Y, Yekutieli D. The control of the false discovery rate in multiple testing under dependency. *Ann. Statist*. 2001;29:1165-1188. <https://doi.org/10.1214/aos/1013699998>.
28. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, Hoang CD, Diehn M, Alizadeh AA. Robust enumeration of cell sub-

- sets from tissue expression profiles. *Nat Methods*. 2015;12:453-457. <https://doi.org/10.1038/nmeth.3337>.
29. Sturm G, Finotello F, Petitprez F, Zhang JD, Baumbach J, Fridman WH, List M, Aneichyk T. Comprehensive evaluation of transcriptome-based cell-type quantification methods for immuno-oncology. *Bioinformatics*. 2019;35:i436-i445. <https://doi.org/10.1093/bioinformatics/btz363>.
30. Sievert C. Interactive Web-Based Data Visualization with R, plotly, and shiny. 1st ed. Chapman and Hall/CRC; 2020. <https://doi.org/10.1201/9780429447273>.
31. Wickham H. ggplot2: Elegant Graphics for Data Analysis. 2nd ed. Springer Cham; 2016. <https://doi.org/10.1007/978-3-319-24277-4>.
32. Aung LL, Brooks A, Greenberg SA, Rosenberg ML, Dhib-Jalbut S, Balashov KE. Multiple sclerosis-linked and interferon-beta-regulated gene expression in plasmacytoid dendritic cells. *J Neuroimmunol*. 2012;250:99-105. <https://doi.org/10.1016/j.jneuroim.2012.05.013>.
33. Nickles D, Chen HP, Li MM, Khankhanian P, Madireddy L, Caillier SJ, Santaniello A, Cree BA, Pelletier D, Hauser SL, Oksenberg JR, Baranzini SE. Blood RNA profiling in a large cohort of multiple sclerosis patients and healthy controls. *Hum Mol Genet*. 2013;22:4194-4205. <https://doi.org/10.1093/hmg/ddt267>.
34. Irizar H, Muñoz-Culla M, Sepúlveda L, Sáenz-Cuesta M, Prada Á, Castillo-Triviño T, Zamora-López G, López de Munain A, Olascoaga J, Otaegui D. Transcriptomic profile reveals gender-specific molecular mechanisms driving multiple sclerosis progression. *PLoS One*. 2014;9:e90482. <https://doi.org/10.1371/journal.pone.0090482>.
35. Feldman A, Gurevich M, Huna-Baron R, Achiron A. The role of B cells in the early onset of the first demyelinating event of acute optic neuritis. *Invest Ophthalmol Vis Sci*. 2015;56:1349-1356. <https://doi.org/10.1167/iovs.14-15408>.
36. Hendrickx DAE, van Scheppingen J, van der Poel M, Bossers K, Schuurman KG, van Eden CG, Hol EM, Hamann J, Huitinga I. Gene Expression Profiling of Multiple Sclerosis Pathology Identifies Early Patterns of Demyelination Surrounding Chronic Active Lesions. *Front Immunol*. 2017;8:1810. <https://doi.org/10.3389/fimmu.2017.01810>.

37. van der Poel M, Ulas T, Mizze MR, Hsiao CC, Miedema SSM, Adelia, Schuurman KG, Helder B, Tas SW, Schultze JL, Hamann J, Huitinga I. Transcriptional profiling of human microglia reveals grey-white matter heterogeneity and multiple sclerosis-associated changes. *Nat Commun.* 2019;10:1139. <https://doi.org/10.1038/s41467-019-08976-7>.
38. Enz LS, Zeis T, Schmid D, Geier F, van der Meer F, Steiner G, Certa U, Binder TMC, Stadelmann C, Martin R, Schaeren-Wiemers N. Increased HLA-DR expression and cortical demyelination in MS links with HLA-DR15. *Neurol Neuroimmunol Neuroinflamm.* 2019;7:e656 <https://doi.org/10.1212/NXI.0000000000000656>.
39. Magliozzi R, Howell OW, Durrenberger P, Aricò E, James R, Cruciani C, Reeves C, Roncaroli F, Nicholas R, Reynolds R. Meningeal inflammation changes the balance of TNF signalling in cortical grey matter in multiple sclerosis. *J Neuroinflammation.* 2019;16:259. <https://doi.org/10.1186/s12974-019-1650-x>.
40. Heinlein M, Gandolfo LC, Zhao K, Teh CE, Nguyen N, Baell JB, Goldfarb Y, Abramson J, Wichmann J, Voss AK, Strasser A, Smyth GK, Thomas T, Gray DHD. The acetyltransferase KAT7 is required for thymic epithelial cell expansion, expression of AIRE target genes, and thymic tolerance. *Sci Immunol.* 2022;7:eabb6032. <https://doi.org/10.1126/sciimmunol.abb6032>.
41. Martinez B, Peplow PV. MicroRNAs in Parkinson's disease and emerging therapeutic targets. *Neural Regen Res.* 2017;12:1945-1959. <https://doi.org/10.4103/1673-5374.221147>.
42. SRP14 Gene - GeneCards [WWW Document]. URL <https://www.genecards.org/cgi-bin/carddisp.pl?gene=SRP14> (accessed 28/02/2022).
43. Takamatsu G, Katagiri C, Tomoyuki T, Shimizu-Okabe C, Nakamura W, Nakamura-Higa M, Hayakawa T, Wakabayashi S, Kondo T, Takayama C, Matsushita M. Tescalcin is a potential target of class I histone deacetylase inhibitors in neurons. *Biochem Biophys Res Commun.* 2017;482:1327-1333. <https://doi.org/10.1016/j.bbrc.2016.12.036>.
44. Jelčić I, Hsu KC, Kakalacheva K, Breiden P, Dupont B, Uhrberg M, Martin R, Münz C, Lünemann JD. Killer immunoglobulin-like receptor locus polymorphisms in multiple sclerosis. *Mult Scler.* 2012;18:951-958. <https://doi.org/10.1177/1352458511431726>.

- 
45. Hoshikawa S, Ogata T, Fujiwara S, Nakamura K, Tanaka S. A novel function of RING finger protein 10 in transcriptional regulation of the myelin-associated glycoprotein gene and myelin formation in Schwann cells. *PLoS One*. 2008;3:e3464. <https://doi.org/10.1371/journal.pone.0003464>.
  46. Kewitz S, Staeger MS. Expression and Regulation of the Endogenous Retrovirus 3 in Hodgkin's Lymphoma Cells. *Front Oncol*. 2013;3:179. <https://doi.org/10.3389/fonc.2013.00179>.
  47. Robak LA, Jansen IE, van Rooij J, Uitterlinden AG, Kraaij R, Jankovic J; International Parkinson's Disease Genomics Consortium (IPDGC); Heutink P, Shulman JM. Excessive burden of lysosomal storage disorder gene variants in Parkinson's disease. *Brain*. 2017;140:3191-3203. <https://doi.org/10.1093/brain/awx285>.
  48. Santana MH, Gomes RC, Utsunomiya YT, Neves HH, Novais FJ, Bonin MN, Fukumasu H, Garcia JF, Alexandre PA, Oliveira Junior GA, Coutinho LL, Ferraz JB. Genome-wide association with residual body weight gain in *Bos indicus* cattle. *Genet Mol Res*. 2015;14:5229-5233. <https://doi.org/10.4238/2015.May.18.14>.
  49. Soundarapandian MM, Selvaraj V, Lo UG, Golub MS, Feldman DH, Pleasure DE, Deng W. Zfp488 promotes oligodendrocyte differentiation of neural progenitor cells in adult mice after demyelination. *Sci Rep*. 2011;1:2. <https://doi.org/10.1038/srep00002>.
  50. UBXN2B Gene - GeneCards [WWW Document]. URL <https://www.genecards.org/cgi-bin/carddisp.pl?gene=UBXN2B> (accessed 28/02/2022).
  51. He R, Wu S, Gao R, Chen J, Peng Q, Hu H, Zhu L, Du Y, Sun W, Ma X, Zhang H, Cui Z, Wang H, Martin BN, Wang Y, Zhang CJ, Wang C. Identification of a Long Noncoding RNA TRAF3IP2-AS1 as Key Regulator of IL-17 Signaling through the SRSF10-IRF1-Act1 Axis in Autoimmune Diseases. *J Immunol*. 2021;206:2353-2365. <https://doi.org/10.4049/jimmunol.2001223>.
  52. Mullin BH, Zhu K, Brown SJ, Mullin S, Tickner J, Pavlos NJ, Dudbridge F, Xu J, Walsh JP, Wilson SG. Genetic regulatory mechanisms in human osteoclasts suggest a role for the STMP1 and DCSTAMP genes in Paget's disease of bone. *Sci Rep*. 2019;9:1052. <https://doi.org/10.1038/s41598-018-37609-0>.

53. Subhi Y, Nielsen MK, Molbech CR, Liisborg C, Søndergaard HB, Sellebjerg F, Sørensen TL. The transcriptome of peripheral blood mononuclear cells in patients with clinical subtypes of late age-related macular degeneration. *Immun Ageing*. 2019;16:20. <https://doi.org/10.1186/s12979-019-0160-0>.
54. Varier KM, Dhandapani H, Liu W, Song J, Wang C, Hu A, Ben-David Y, Shen X, Li Y, Gajendran B. An immunotherapeutic approach to decipher the role of long non-coding RNAs in cancer progression, resistance and epigenetic regulation of immune cells. *J Exp Clin Cancer Res*. 2021;40:242. <https://doi.org/10.1186/s13046-021-01997-5>.
55. Casanova Ferrer F, Pascual M, Hidalgo MR, Malmierca-Merlo P, Guerri C, García-García F. Unveiling Sex-Based Differences in the Effects of Alcohol Abuse: A Comprehensive Functional Meta-Analysis of Transcriptomic Studies. *Genes (Basel)*. 2020;11:1106. <https://doi.org/10.3390/genes11091106>.
56. Català-Senent JF, Hidalgo MR, Berenguer M, Parthasarathy G, Malhi H, Malmierca-Merlo P, de la Iglesia-Vayá M, García-García F. Hepatic steatosis and steatohepatitis: a functional meta-analysis of sex-based differences in transcriptomic studies. *Biol Sex Differ*. 2021;12:29. <https://doi.org/10.1186/s13293-021-00368-1>.
57. López-Cerdán A, Andreu Z, Hidalgo MR, Grillo-Risco R, Català-Senent JF, Soler-Sáez I, Neva-Alejo A, Gordillo F, de la Iglesia-Vayá M, García-García F. Unveiling sex-based differences in Parkinson's disease: a comprehensive meta-analysis of transcriptomic studies. *Biol Sex Differ*. 2022;13:68. <https://doi.org/10.1186/s13293-022-00477-5>.
58. Pérez-Díez I, Hidalgo MR, Malmierca-Merlo P, Andreu Z, Romera-Giner S, Farràs R, de la Iglesia-Vayá M, Provencio M, Romero A, García-García F. Functional Signatures in Non-Small-Cell Lung Cancer: A Systematic Review and Meta-Analysis of Sex-Based Differences in Transcriptomic Studies. *Cancers (Basel)*. 2021;13:143. <https://doi.org/10.3390/cancers13010143>.
59. Moradi S, Stankovic S, O'Connor GM, Pymm P, MacLachlan BJ, Faoro C, Retière C, Sullivan LC, Saunders PM, Widjaja J, Cox-Livingstone S, Rossjohn J, Brooks AG, Vivian JP. Structural plasticity of KIR2DL2 and KIR2DL3 enables altered docking geometries atop HLA-C. *Nat Commun*. 2021;12:2173. <https://doi.org/10.1038/s41467-021-22359-x>.



- 
60. Lorentzen AR, Karlsen TH, Olsson M, Smestad C, Mero IL, Woldseth B, Sun JY, Senitzer D, Celius EG, Thorsby E, Spurkland A, Lie BA, Harbo HF. Killer immunoglobulin-like receptor ligand HLA-Bw4 protects against multiple sclerosis. *Ann Neurol.* 2009;65:658-66. <https://doi.org/10.1002/ana.21695>.
  61. Nishimura WE, Sachetto Z, Costallat LT, Yazbek MA, Londe AC, Guariento EG, Marques SB, Bertolo MB. The role of KIR2DL3/HLA-C\*0802 in Brazilian patients with rheumatoid vasculitis. *Clinics (Sao Paulo).* 2015;70:408-412. [https://doi.org/10.6061/clinics/2015\(06\)04](https://doi.org/10.6061/clinics/2015(06)04).
  62. Hou Y, Zhang C, Xu D, Sun H. Association of killer cell immunoglobulin-like receptor and human leucocyte antigen-Cw gene combinations with systemic lupus erythematosus. *Clin Exp Immunol.* 2015;180:250-254. <https://doi.org/10.1111/cei.12582>.
  63. Jones DC, Edgar RS, Ahmad T, Cummings JR, Jewell DP, Trowsdale J, Young NT. Killer Ig-like receptor (KIR) genotype and HLA ligand combinations in ulcerative colitis susceptibility. *Genes Immun.* 2006;7:576-582. <https://doi.org/10.1038/sj.gene.6364333>.
  64. Poszepczynska-Guigné E, Schiavon V, D'Incan M, Echchakir H, Musette P, Ortonne N, Boumsell L, Moretta A, Bensussan A, Bagot M. CD158k/KIR3DL2 is a new phenotypic marker of Sezary cells: relevance for the diagnosis and follow-up of Sezary syndrome. *J Invest Dermatol.* 2004;122:820-823. <https://doi.org/10.1111/j.0022-202X.2004.22326.x>.
  65. Frazier WR, Steiner N, Hou L, Dakshanamurthy S, Hurley CK. Allelic variation in KIR2DL3 generates a KIR2DL2-like receptor with increased binding to its HLA-C ligand. *J Immunol.* 2013;190:6198-6208. <https://doi.org/10.4049/jimmunol.1300464>.
  66. Al-Mossawi H, Yager N, Taylor CA, Lau E, Danielli S, de Wit J, Gilchrist J, Nassiri I, Mahe EA, Lee W, Rizvi L, Makino S, Cheeseman J, Neville M, Knight JC, Bowness P, Fairfax BP. Context-specific regulation of surface and soluble IL7R expression by an autoimmune risk allele. *Nat Commun.* 2019;10:4575. <https://doi.org/10.1038/s41467-019-12393-1>.
  67. Huygens C, Liénart S, Dedobbeleer O, Stockis J, Gauthy E, Coulie PG, Lucas S. Lysosomal-associated Transmembrane Protein 4B (LAPTM4B) Decreases Transforming Growth Factor  $\beta$ 1 (TGF- $\beta$ 1) Production in Human Regulatory T Cells. *J Biol Chem.* 2015;290:20105-20116. <https://doi.org/10.1074/jbc.M115.655340>.

68. Kattygnarath D, Maugendre S, Neyroud N, Balse E, Ichai C, Denjoy I, Dilanian G, Martins RP, Fressart V, Berthet M, Schott JJ, Leenhardt A, Probst V, Le Marec H, Hainque B, Coulombe A, Hatem SN, Guicheney P. MOG1: a new susceptibility gene for Brugada syndrome. *Circ Cardiovasc Genet.* 2011;4:261-268. <https://doi.org/10.1161/CIRCGENETICS.110.959130>.
69. Espinosa Collado A, Ramos Ortiz C, Burillo Gómez F, Moreno Terribas G. Esclerosis múltiple y cardiopatía isquémica: nuevo factor de riesgo cardiovascular. *Med. Gen. Fam.* 2021;10:156-158. <https://doi.org/10.24038/mgyf.2021.013>.
70. Dai E, Guan H, Liu L, Little S, McFadden G, Vaziri S, Cao H, Ivanova IA, Bocksch L, Lucas A. Serp-1, a viral anti-inflammatory serpin, regulates cellular serine proteinase and serpin responses to vascular injury. *J Biol Chem.* 2003;278:18563-18572. <https://doi.org/10.1074/jbc.M209683200>.
71. Heit C, Jackson BC, McAndrews M, Wright MW, Thompson DC, Silverman GA, Nebert DW, Vasiliou V. Update of the human and mouse SERPIN gene superfamily. *Hum Genomics.* 2013;7:22. <https://doi.org/10.1186/1479-7364-7-22>.
72. Wang Z, Tian Z, Song X, Zhang J. Membrane tension sensing molecule-FNBP1 is a prognostic biomarker related to immune infiltration in BRCA, LUAD and STAD. *BMC Immunol.* 2022;23:1. <https://doi.org/10.1186/s12865-021-00475-z>.
73. Newman DM, Voss AK, Thomas T, Allan RS. Essential role for the histone acetyltransferase KAT7 in T cell development, fitness, and survival. *J Leukoc Biol.* 2017;101:887-892. <https://doi.org/10.1189/jlb.1MA0816-338R>.
74. Gkika D, Lemonnier L, Shapovalov G, Gordienko D, Poux C, Bernardini M, Bokhobza A, Bidaux G, Degerny C, Verreman K, Guarmit B, Benahmed M, de Launoit Y, Bindels RJ, Fiorio Pla A, Prevarskaya N. TRP channel-associated factors are a novel protein family that regulates TRPM8 trafficking and activity. *J Cell Biol.* 2015;208:89-107. <https://doi.org/10.1083/jcb.201402076>.
75. Yin L, Zhou L, Xu R. Identification of Tumor Mutation Burden and Immune Infiltrates in Hepatocellular Carcinoma Based

- on Multi-Omics Analysis. *Front Mol Biosci.* 2021;7:599142. <https://doi.org/10.3389/fmolb.2020.599142>.
76. Yu WD, Wang H, He QF, Xu Y, Wang XC. Long noncoding RNAs in cancer-immunity cycle. *J Cell Physiol.* 2018;233:6518-6523. <https://doi.org/10.1002/jcp.26568>.
77. Nesslinger NJ, Sahota RA, Stone B, Johnson K, Chima N, King C, Rasmussen D, Bishop D, Rennie PS, Gleave M, Blood P, Pai H, Ludgate C, Nelson BH. Standard treatments induce antigen-specific immune responses in prostate cancer. *Clin Cancer Res.* 2007;13:1493-1502. <https://doi.org/10.1158/1078-0432.CCR-06-1772>.
78. Swaidani S, Liu C, Zhao J, Bulek K, Li X. TRAF Regulation of IL-17 Cytokine Signaling. *Front Immunol.* 2019;10:1293. <https://doi.org/10.3389/fimmu.2019.01293>.
79. Jeffcoat S, Givens L, Bolanos S, Malhotra S, Koke J. A novel intermediate filament-associated protein: further characterization of the G.3.5 antigen. *Cytobios.* 1995;82:81-99. PMID: 7587376.
80. Huang H, Xing D, Zhang Q, Li H, Lin J, He Z, Lin J. LncRNAs as a new regulator of chronic musculoskeletal disorder. *Cell Prolif.* 2021;54:e13113. <https://doi.org/10.1111/cpr.13113>.
81. Hashiguchi K, Hayashi M, Sekiguchi M, Umezu K. The roles of human MTH1, MTH2 and MTH3 proteins in maintaining genome stability under oxidative stress. *Mutat Res.* 2018;808:10-19. <https://doi.org/10.1016/j.mrfmmm.2018.01.002>.
82. Sasidhar MV, Itoh N, Gold SM, Lawson GW, Voskuhl RR. The XX sex chromosome complement in mice is associated with increased spontaneous lupus compared with XY. *Ann Rheum Dis.* 2012;71:1418-1422. <https://doi.org/10.1136/annrheumdis-2011-201246>.
83. Smith-Bouvier DL, Divekar AA, Sasidhar M, Du S, Tiwari-Woodruff SK, King JK, Arnold AP, Singh RR, Voskuhl RR. A role for sex chromosome complement in the female bias in autoimmune disease. *J Exp Med.* 2008;205:1099-1108. <https://doi.org/10.1084/jem.20070850>.
84. Klein SL, Flanagan KL. Sex differences in immune responses. *Nat Rev Immunol.* 2016;16:626-638. <https://doi.org/10.1038/nri.2016.90>.

85. Bargiela D, Chinnery PF. Mitochondria in neuroinflammation - Multiple sclerosis (MS), leber hereditary optic neuropathy (LHON) and LHON-MS. *Neurosci Lett.* 2019;710:132932. <https://doi.org/10.1016/j.neulet.2017.06.051>.
86. Bartosik-Psujek H, Belniak E, Mitosek-Szewczyk K, Dobosz B, Stelmasiak Z. Interleukin-8 and RANTES levels in patients with relapsing-remitting multiple sclerosis (RR-MS) treated with cladribine. *Acta Neurol Scand.* 2004;109:390-392. <https://doi.org/10.1111/j.1600-0404.2004.00259.x>.
87. Campbell G, Mahad DJ. Mitochondrial dysfunction and axon degeneration in progressive multiple sclerosis. *FEBS Lett.* 2018;592:1113-1121. <https://doi.org/10.1002/1873-3468.13013>.
88. Lu CZ, Jensen MA, Arnason BG. Interferon gamma- and interleukin-4-secreting cells in multiple sclerosis. *J Neuroimmunol.* 1993;46:123-128. [https://doi.org/10.1016/0165-5728\(93\)90241-p](https://doi.org/10.1016/0165-5728(93)90241-p).
89. Matejčíková Z, Mareš J, Příkrylová Vranová H, Klosová J, Sládková V, Doláková J, Zapletalová J, Kaňovský P. Cerebrospinal fluid inflammatory markers in patients with multiple sclerosis: a pilot study. *J Neural Transm (Vienna).* 2015;122:273-277. <https://doi.org/10.1007/s00702-014-1244-9>.
90. Cuevas-García CF, Segura-Méndez NH, Herrera-Sánchez DA. Actualidades en la inmunopatología de la esclerosis múltiple. *Gac Med Mex.* 2018;154:588-597. <https://doi.org/10.24875/GMM.18003407>.
91. Kunkl M, Frasca S, Amormino C, Volpe E, Tuosto L. T Helper Cells: The Modulators of Inflammation in Multiple Sclerosis. *Cells.* 2020;9:482. <https://doi.org/10.3390/cells9020482>.
92. Saligrama N, Zhao F, Sikora MJ, Serratelli WS, Fernandes RA, Louis DM, Yao W, Ji X, Idoyaga J, Mahajan VB, Steinmetz LM, Chien YH, Hauser SL, Oksenberg JR, Garcia KC, Davis MM. Opposing T cell responses in experimental autoimmune encephalomyelitis. *Nature.* 2019;572:481-487. <https://doi.org/10.1038/s41586-019-1467-x>.
93. Ridgley LA, Anderson AE, Pratt AG. What are the dominant cytokines in early rheumatoid arthritis? *Curr Opin Rheumatol.* 2018;30:207-214. <https://doi.org/10.1097/BOR.0000000000000470>.

- 
94. Correale J. Immunosuppressive Amino-Acid Catabolizing Enzymes in Multiple Sclerosis. *Front Immunol.* 2021;11:600428. <https://doi.org/10.3389/fimmu.2020.600428>.
  95. Fitzgerald KA, Kagan JC. Toll-like Receptors and the Control of Immunity. *Cell.* 2020;180:1044-1066. <https://doi.org/10.1016/j.cell.2020.02.041>.
  96. Vijay K. Toll-like receptors in immunity and inflammatory diseases: Past, present, and future. *Int Immunopharmacol.* 2018;59:391-412. <https://doi.org/10.1016/j.intimp.2018.03.002>.
  97. Kim YK, Shin JS, Nahm MH. NOD-Like Receptors in Infection, Immunity, and Diseases. *Yonsei Med J.* 2016;57:5-14. <https://doi.org/10.3349/ymj.2016.57.1.5>.
  98. Haase S, Haghikia A, Wilck N, Müller DN, Linker RA. Impacts of microbiome metabolites on immune regulation and autoimmunity. *Immunology.* 2018;154:230-238. <https://doi.org/10.1111/imm.12933>.
  99. Martínez-Altarriba MC, Ramos-Campoy O, Luna-Calcaño IM, Arrieta-Antón E. Revisión de la esclerosis múltiple (2). Diagnóstico y tratamiento. *Semergen.* 2015;41:324-328. <https://doi.org/10.1016/j.semerg.2014.07.011>.
  100. Torre-Fuentes L, Moreno-Jiménez L, Pytel V, Matías-Guiu JA, Gómez-Pinedo U, Matías-Guiu J. Experimental models of demyelination and remyelination. *Neurologia (Engl Ed).* 2020;35:32-39. <https://doi.org/10.1016/j.nrl.2017.07.002>.
  101. Engel KL, Arora A, Goering R, Lo HG, Taliaferro JM. Mechanisms and consequences of subcellular RNA localization across diverse cell types. *Traffic.* 2020;21:404-418. <https://doi.org/10.1111/tra.12730>.
  102. Junjappa RP, Patil P, Bhattarai KR, Kim HR, Chae HJ. IRE1 $\alpha$  Implications in Endoplasmic Reticulum Stress-Mediated Development and Pathogenesis of Autoimmune Diseases. *Front Immunol.* 2018;9:1289. <https://doi.org/10.3389/fimmu.2018.01289>.
  103. Cruz FM, Colbert JD, Rock KL. The GTPase Rab39a promotes phagosome maturation into MHC-I antigen-presenting compartments. *EMBO J.* 2020;39:e102020. <https://doi.org/10.15252/embj.2019102020>.

104. Riccio P, Rossano R. Diet, Gut Microbiota, and Vitamins D + A in Multiple Sclerosis. *Neurotherapeutics*. 2018;15:75-91. <https://doi.org/10.1007/s13311-017-0581-4>.
105. Yanguas-Casás N. Sex Differences in Neurodegenerative Diseases. *SM J Neurol Disord Stroke*. 2017;3:1014.
106. Clayton JA, Collins FS. Policy: NIH to balance sex in cell and animal studies. *Nature*. 2014;509:282-283. <https://doi.org/10.1038/509282a>.
107. Accounting for sex and gender makes for better science. *Nature*. 2020;588:196. <https://doi.org/10.1038/d41586-020-03459-y>.
108. European Commission, Directorate General for Justice and Consumers, Fondazione Giacomo Brodolini. Gender Equality and Health in the EU Publications Office, LU. 2021. <https://doi.org/10.2838/991480>.
109. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten JW, da Silva Santos LB, Bourne PE, Bouwman J, Brookes AJ, Clark T, Crosas M, Dillo I, Dumon O, Edmunds S, Evelo CT, Finkers R, Gonzalez-Beltran A, Gray AJ, Groth P, Goble C, Grethe JS, Heringa J, 't Hoen PA, Hooft R, Kuhn T, Kok R, Kok J, Lusher SJ, Martone ME, Mons A, Packer AL, Persson B, Rocca-Serra P, Roos M, van Schaik R, Sansone SA, Schultes E, Sengstag T, Slater T, Strawn G, Swertz MA, Thompson M, van der Lei J, van Mulligen E, Velterop J, Waagmeester A, Wittenburg P, Wolstencroft K, Zhao J, Mons B. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data*. 2016;3:160018. <https://doi.org/10.1038/sdata.2016.18>.

## **4.11 Acknowledgements**

The authors thank the Principe Felipe Research Center (CIPF) for providing access to the cluster, co-funded by European Regional Development Funds (FEDER) in Valencian Community 2014-2020. Irene Soler-Sáez thanks the Spanish Ministry of Universities for her predoctoral grant FPU20/03544. The authors also thank Stuart P. Atkinson for reviewing the manuscript.

## 4.12 Funding

This research was supported by and partially funded by the Institute of Health Carlos III (project IMPaCT-Data, exp. IMP/00019), co-funded by the European Union, European Regional Development Fund (ERDF, “A way to make Europe”), PID2021-124430OA-I00 funded by MCIN/AEI/10.13039/501100011033/ FEDER, UE (“A way to make Europe”).

## 4.13 CRediT authorship contribution statement

**José Francisco Català-Senent:** Data curation, Formal analysis, Investigation, Methodology, Software, Visualization, Writing – original draft, Writing – review & editing. **Zoraida Andreu:** Investigation, Validation, Writing – original draft, Writing – review & editing. **Marta R. Hidalgo:** Investigation, Software, Visualization, Writing – review & editing. **Irene Soler-Sáez:** Data curation, Visualization, Writing – review & editing. **Francisco José Roig:** Data curation, Formal analysis, Writing – review & editing. **Natalia Yanguas-Casás:** Investigation, Validation, Writing – review & editing. **Almudena Neva-Alejo:** Data curation, Writing – review & editing. **Adolfo López-Cerdán:** Writing – review & editing. **María de la Iglesia-Vayá:** Writing – review & editing. **Barbara E. Stranger:** Investigation, Writing – review & editing. **Francisco García-García:** Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Supervision, Visualization, Writing – original draft, Writing – review & editing.

## 4.14 Author details

<sup>1</sup>Bioinformatics and Biostatistics Unit, Principe Felipe Research Center (CIPF), 46012, Valencia, Spain.

<sup>2</sup>Foundation Valencian Institute of Oncology (FIVO), 46009, Valencia, Spain.

<sup>3</sup>Faculty of Health Sciences. San Jorge University, 50830, Zaragoza, Spain.

<sup>4</sup>Instituto de Investigación Sanitaria Puerta de Hierro-Segovia de Arana (IDIPHISA), Grupo de Investigación en Linfomas, C/Joaquín Rodrigo 2, Majadahonda, 28222 Madrid, Spain.

<sup>5</sup>Biomedical Imaging Unit FISABIO-CIPF, Fundación para el Fomento de la Investigación Sanitaria y Biomédica de la Comunidad Valenciana, 46012, Valencia, Spain.

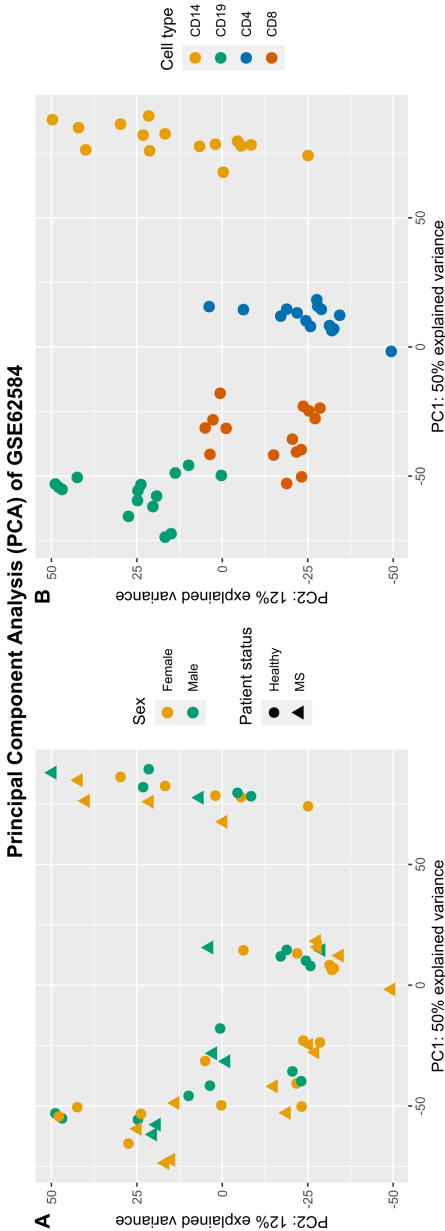
<sup>6</sup>Department of Pharmacology, Center for Genetic Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL, USA

<sup>†</sup>José Francisco Català-Senent and Zoraida Andreu contributed equally to this study.

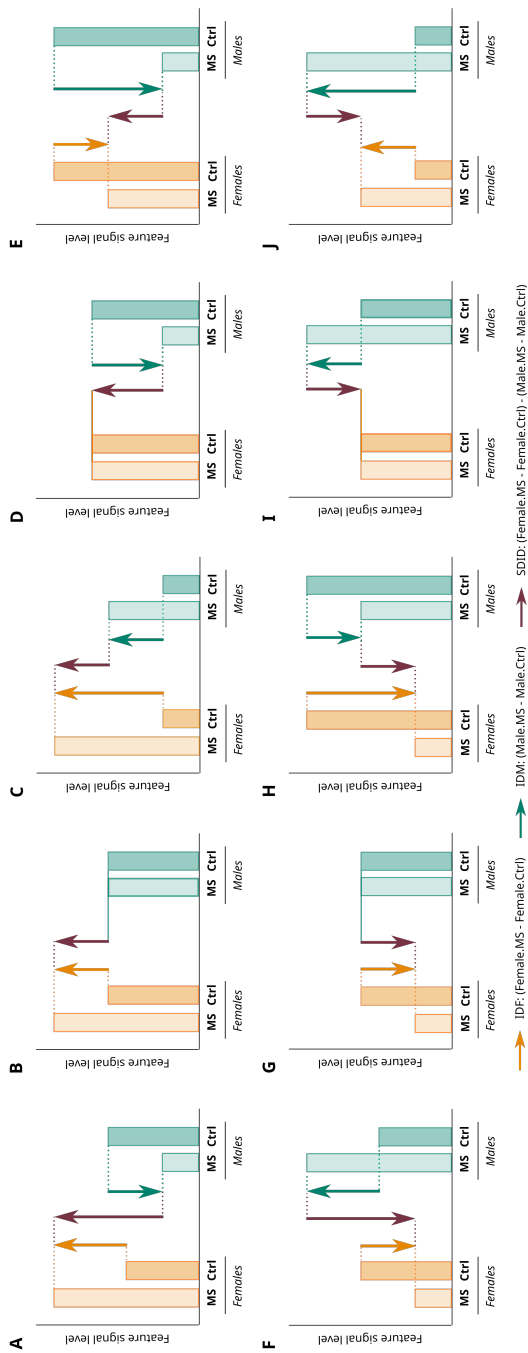
\*Corresponding author: fgarcia@cipf.es (Francisco García-García)



4.15 Supplementary material



**Fig. 4.S1. Principal component analysis plot of study GSE62584.** A) The sex of the individuals is indicated in green and orange while the disease status is indicated in rounded and triangular shapes. B) The colors indicate the 4 cell types included in the study, showing the batch effect caused by this factor.



**Fig. 4.S2. Bar chart** representing the possible causes of a positive (A-E) or negative (F-J) LFC/LOR in the SDID comparison, depending on the results of the IDf or IDm comparisons. The Y-axis represents the level of the feature signal (gene expression in the case of DGE and level of enrichment in functional enrichment). On the X-axis the values for diseased females (light orange), healthy females (dark orange), diseased males (light green) and healthy males (dark green) are shown as vertical bars. The orange arrow represents the magnitude of the IDf comparison (females), the green arrow the magnitude of the IDm comparison (males) and the purple arrow indicates the magnitude of the SDID comparison (IDf vs IDm).

All supplementary figures and tables of this article are available both on the article's website (<https://doi.org/10.1016/j.nbd.2023.106113>) and in the public repository Zenodo (<https://doi.org/10.5281/zenodo.7738725>) for better visualization.



## Chapter 5

# General discussion

### 5.1 Sex differences in health and research

The existence of sex differences in health and research is a fact that is currently taken into account and is beginning to be widely recognized. Since the first article in 1991 focused on this problem [10], hundreds of thousands of scientific papers have appeared on this subject [147]. As we have seen in Chapter 1 of this thesis, the efforts made in recent years in research and medical practice are helping to resolve the bias against women that has traditionally existed. However, there is still work to be done. In addition, including the biological differences between sexes in medicine allows for a more precise clinical practice, leading to treatments that are more effective and have fewer side effects [5, 11, 12].

The number of diseases in which sex differences are known to exist in some aspects (onset, development, diagnosis, treatment, etc.) is enormous. The list includes cardiac, neurodegenerative, autoimmune, metabolic, inflammatory, and cancer diseases [3, 64, 71, 148] (Table 1.1). As introduced in Section 1.1.4, various molecular mechanisms contribute to sex differences, including genetic, epigenetic, and hormonal factors. These mechanisms lead to distinct alterations in inflammatory, immune, and other physiological processes. [3, 27, 54, 64, 71]. In particular, Kronzer, Bridges, and Davis [47] propose that differences

in the immune systems of males and females, specifically the higher levels of antibodies found in females, could have an evolutionary basis. They suggest that, despite the risk of autoimmune diseases for mothers, the increased protection against infections during the vulnerable infancy period likely enhanced the survival rates of their offspring. As a result, the higher survival rates of their offspring may have led to the transmission of this propensity for autoimmunity to future generations, in line with classical Darwinian principles.

In this thesis, we aimed to elucidate the molecular mechanisms that cause sex differences in two pathologies: non-alcoholic fatty liver disease (NAFLD) and multiple sclerosis (MS). NAFLD comprises a spectrum of liver disorders that begin with fat accumulation in hepatocytes and can lead to inflammation, cirrhosis, and even liver cancer. The disease is more prevalent in males than females (3:2) [50], with estrogens providing a protective effect in females that diminishes after menopause [36]. On the other hand, MS is a chronic autoimmune, inflammatory, and degenerative disease of the central nervous system. Women are at a higher risk of developing MS, while men experience worse and more rapid progression of the disease [149, 150, 151].

Although NAFLD and MS are diseases with significantly different origins, this thesis has revealed common sex-associated alterations in both pathologies. In our investigation of NAFLD, we conducted an assessment of sex-related disparities in the transition from the NAFL stage to NASH. In contrast, our MS study identified sex-associated alterations in brain tissue utilizing three distinct contrasts (SDID, IDF, and IDM). Because of the different bioinformatics approaches used in the previous chapters (Figs. 3.1 and 4.1), the functional results using the biological processes of the Gene Ontology (GO) [118] and Kyoto Encyclopedia of Genes and Genomes (KEGG) [119] databases are the comparable elements in common between the two meta-analyses.

The primary common difference observed in both diseases is the sex-dependent variation of 2 cytokines: interleukin-2 and interleukin-6. In the NAFLD study, 2 GO terms involving in their descriptions IL-2 and 1 involving IL-6 were over-represented in females and males, respectively (Tables 3.3 and 5.1). On the other hand, in the IDF contrast of the MS study, we found 3 IL-6-associated and 2 IL-2-associated GO terms significantly overexpressed in females with the disease concerning healthy females (Table 5.1). These findings would support the existence of different behaviors of the immune system against these diseases, depending on the sex of the individual.

Interleukins are vital for the activation and differentiation of immune cells and their proliferation, maturation, migration, and adhesion. They have both

**Table 5.1.** Compilation of significant GO terms associated with IL-2 or IL-6 in all comparisons in both studies.

| Disease | Comparison       | GO term    |                                        | LOR    |
|---------|------------------|------------|----------------------------------------|--------|
| NAFLD   | <sup>1</sup>     | GO:0070669 | Response to IL-2                       | 0.446  |
| NAFLD   | <sup>1</sup>     | GO:0071352 | Cellular response to IL-2              | 0.467  |
| NAFLD   | <sup>1</sup>     | GO:1900165 | Negative regulation of IL-6 secretion  | -0.762 |
| MS      | IDF <sup>2</sup> | GO:0032623 | IL-2 production                        | 0.529  |
| MS      | IDF <sup>2</sup> | GO:0032635 | IL-6 production                        | 0.358  |
| MS      | IDF <sup>2</sup> | GO:0032663 | Regulation of IL-2 production          | 0.529  |
| MS      | IDF <sup>2</sup> | GO:0032675 | Regulation of IL-6 production          | 0.358  |
| MS      | IDF <sup>2</sup> | GO:0032755 | Positive regulation of IL-6 production | 0.480  |

<sup>1</sup> (NASH.W – NAFL.W) vs. (NASH.M – NAFL.M), where *NASH.W*, *NAFL.W*, *NASH.M*, and *NAFL.M* corresponded to NASH-affected premenopausal females, NAFL-affected premenopausal females, NASH-affected males, and NAFL-affected males, respectively.

<sup>2</sup> Impact of the disease in females (IDF): (Female.MS - Female.Ctrl), where *Female.MS* and *Female.Ctrl* correspond to females with MS and control females, respectively.

pro-inflammatory and anti-inflammatory properties. The primary role of interleukins is to regulate growth, differentiation, and activation during inflammatory and immune responses, facilitating communication and coordination among immune cells [56, 57, 152].

IL-2 is a pleiotropic interleukin essential in the maintenance of immune homeostasis [153, 154]. It serves as a significant expansion factor for most, if not all, types of activated T cells. Although other cytokines may provide some redundancy with IL-2 in this function, IL-2 is crucial for determining both the magnitude and duration of primary and memory immune responses [153]. Moreover, IL-2 can play a vital role in modulating immune responses; its absence can lead to severe autoimmunity due to the inability to eliminate activated T cells [62, 63, 154]. However, in the case of MS, how IL-2 levels participate in the disease is still unclear [155]. Distinct behaviors in MS with respect to other autoimmune diseases have been described, such as a more modest and delayed response to low-dose interleukin-2 treatments [156]. Interestingly, a recent study proposes that elevated levels of IL-2 in cerebrospinal fluid ( $\geq 1.23$  pg/ml) are linked to shorter relapse-free intervals in patients with MS [157]. In the liver, deficient IL-2 signaling has been linked to several chronic liver diseases [158]. In the context of NAFLD, elevated interleukin-2 receptor alpha (IL2RA) levels have been linked to the NASH stage [159, 160]. These findings reveal the influence of IL-2 in liver diseases.

As mentioned in the introduction, sex differences in IL-2 have been previously described. Research by Giron-Gonzalez et al. [58] found that, under normal conditions, IL-2 levels are higher in males than in females. These differences

may be partly attributed to estrogen. Indeed, Dodd and Menon [64] observed that increases in the concentration of this hormone, such as those occurring in pregnancy, lead to a decrease in IL-2 levels. Our research on multiple sclerosis found that diseased females produce higher levels of IL-2 compared to healthy females (Table 5.1). These results align with the study mentioned in the previous paragraph in which Petržalka et al. [157] found that high levels of IL-2 in cerebrospinal fluid are associated with a more severe MS stage. While this study did not identify any sex differences in this marker, it is possible that the unbalanced sex ratio (43 females to 15 males) may have influenced the findings. Further research is needed to confirm this. In NAFLD, in the progression from non-alcoholic fatty liver disease (NAFL) to non-alcoholic steatohepatitis (NASH), premenopausal women exhibit a stronger response to IL-2 (Table 5.1), which aligns with conclusions drawn from previous studies that associate high levels of IL2RA with the inflammatory stage of the disease [159, 160]. In conclusion of all these results, IL-2 plays a critical role in influencing the sex differences observed in both NAFLD and MS.

IL-6 is also a pleiotropic cytokine involved in a wide range of biological functions, including the inflammatory response, the immune system, hematopoiesis, and the maintenance of bone and skeletal muscle mass. It also plays a role in central nervous system activity and metabolism [161, 162, 163, 164, 165, 166]. Various cell types, both immune and non-immune, can produce IL-6 [56]. In muscles, IL-6 is released into the bloodstream during contractions. This release helps mobilize energy and increases body temperature since IL-6 can cross the blood-brain barrier [59, 162]. In the liver, IL-6 initiates lipolysis, which mobilizes energy-storing adipose tissue [40, 165], and it is involved in the acute phase response, playing a crucial role in the body's inflammatory response and liver regeneration [164, 167]. Moreover, elevated levels of IL-6 have been associated with the severity of metabolic pathologies such as insulin resistance or obesity [40, 164, 167, 168] and, may promote adipose tissue inflammation [40, 163]. In the central nervous system, IL-6 can activate glial cells and disrupt the blood-brain barrier, leading to increased neuroinflammation, while also promoting neuronal survival and neurogenesis [166, 161]. Dysregulation of IL-6 signaling in brain regions such as the hippocampus, cortex, striatum, and substantia nigra can significantly impact neurological function and contribute to various symptoms seen in neurological disorders [166, 169].

This broad spectrum of functions throughout the body means that IL-6 plays a role in a wide range of pathologies, as mentioned above. In the case of NAFLD, previous studies such as Wieckowska et al. [168] correlate IL-6 levels with disease severity, which is supported by a recent meta-analysis with



36074 NAFLD patients [170]. Our results (Table 5.1) show a negative regulation of IL-6 in disease progression in men concerning premenopausal women, leading us to hypothesize that premenopausal women may have a decreased immune response to lipotoxicity. Additionally, the differential regulation of IL-6 may also influence the protective effect of subcutaneous adipose tissue (SAT) fat accumulation against metabolic diseases, which is more prevalent in premenopausal women and is regulated by IL-6 [171]. Concerning multiple sclerosis, increased IL-6 expression in the brain is linked to neuropathologies such as Alzheimer’s disease, Parkinson’s disease, and multiple sclerosis [161]. Although its critical role in host defense and glial/neuronal homeostasis is known, more research is necessary to clarify the specific cellular effects of IL-6 signaling within the CNS [161]. Our meta-analysis revealed an increase in IL-6 levels in affected females, a pattern not observed in males meta-analysis (Table 5.1). Notably, this finding aligns with a study on mice that had an IL-6 deficiency in astrocytes [172]. In that study, female mutant mice exhibited slightly lower clinical scores of autoimmune encephalomyelitis and less demyelination compared to their wild-type counterparts, while male mice did not show similar effects. These results suggest that further research into mimicking this IL-6 deficiency effect for females with MS could be valuable.

At a more general level, the research presented in this doctoral thesis highlights various sex-associated immune system behaviors in both non-alcoholic fatty liver disease and multiple sclerosis. Other authors have previously described the existence of sex differences in the immune system [11, 27, 29, 45, 54, 64, 71], however, the findings of this work represent an advance in the knowledge of the behavior of the immune system in the pathologies studied. In our research on NAFLD, we identified 13 significantly altered GO terms (Table 3.3). Notably, 3 of these terms (23%) are associated with IL-2 and IL-6, underscoring the vital role of the modulation of immune system in this pathology. A previous study conducted by Lonardo et al. [29] demonstrated that the innate immune cells in male mice are activated in a manner that promotes liver inflammation and fibrosis. In contrast, the macrophages in female mice exhibit either a regulatory phenotype or a more protective and antifibrotic response. Our research on sex differences in MS highlights significant distinctions between females and males affected by the condition (Fig. 4.7). In females, we identified changes primarily associated with myeloid cells and a pro-inflammatory environment, influenced by alterations in interleukin regulation and response (e.g., IL-1, IL-2, IL-6, IL-10, and interferon-gamma) (Fig. 4.8). Conversely, in males, our analysis indicates alterations related to lymphoid cells. These changes influence key aspects such as cell activation, proliferation, and differentiation (Fig. 4.8),

which are essential components of the adaptive immune response [173, 174, 175].

The findings and hypotheses presented in this doctoral thesis contribute to addressing the disparities in sex differences related to NAFLD and MS. This work paves the way for new research opportunities aimed at developing more precise therapies for both men and women in the management and treatment of these conditions.

## 5.2 Research synthesis and Open Science

### 5.2.1 *Research synthesis*

In recent decades, the number of scientific papers and works has increased exponentially [176, 177]. It is now common for several studies to have addressed the same scientific question, either due to ignorance of prior research, skepticism about earlier findings, or a desire to go further than the previous findings. [178]. However, it is rare for studies on the same issue to produce identical results [137, 178]. This variability in outcomes raises the challenge of determining which conclusions are correct. In response to this challenge, research synthesis has emerged, aiming to gather all available evidence on a scientific problem to integrate the results and provide a more robust answer to the posed question.

During the latter half of the twentieth century, the research synthesis was considered a second-class job, and was even regarded as a "parasitic recycling of the work of those engaged in the real business of science" [137]. However, this vision has changed in the last two decades, driven by the evidence of the benefits of integrative methods and in the heat of movements for a more collaborative science such as Open Science. Proof of this can be seen in the fact that the number of publications using meta-analysis techniques has increased dramatically since the beginning of the 21st century [143].

As the principal research synthesis methodology, systematic review and meta-analysis are widely used in clinical research [143, 179]. When properly conducted, a meta-analysis is considered decisive evidence on a clinical issue, placing it at the highest level in the hierarchy of evidence [180]. As remarked in the introduction of this thesis, this technique not only draws combined information from studies that posed a common scientific question but also extends previous knowledge by answering new questions from existing data [131]. This

approach fosters the inclusion of fresh perspectives in existing research. For example, this doctoral thesis incorporates the sex variable into the comprehensive meta-analyses of publicly available transcriptomic studies on non-alcoholic fatty liver disease and multiple sclerosis, enriching our understanding of these conditions. The systematic review and meta-analysis methodology has been extensively employed by the Computational Biomedicine Laboratory at the Centro de Investigación Príncipe Felipe as a powerful approach for uncovering sex-specific molecular alterations in a range of diseases, including cancers [181, 182], neurodegenerative disorders [31, 183, 184], and psychiatric conditions [185], among others [186, 187]; inspiring advancements in our understanding and treatment of these pathologies.

Nowadays, the field of research synthesis is on the edge of a revolutionary transformation. Advances in artificial intelligence (AI) have led to the development of numerous tools that can assist with various stages of systematic reviews, such as screening and data extraction [143, 188]. These tools enable a more efficient processing of large volumes of information. However, AI in research synthesis is still in the beginning phase and should be considered a supplement to human work. Further research is necessary to evaluate the effectiveness, accuracy, validity, and limitations of these AI tools, as well as their impact on the quality and reliability of research synthesis [188].

### **5.2.2 Open Science**

Like many systematic reviews and meta-analyses, one of the primary limitations of this doctoral thesis was the scarcity of available transcriptomic studies in public databases containing complete information about individuals. The systematic review conducted in the NAFLD study identified only seven datasets that met the necessary criteria (Table 3.1). In contrast, the MS systematic review found five datasets related to blood samples and four related to brain tissue samples (Table 4.2).

Although the number of publicly accessible studies is continuously increasing [123], it remains relatively limited. Additionally, the metadata (information on samples such as sex, disease, and tissue source site [189]) available from these studies is heterogeneous, complicating the application of integrative techniques and preventing the possibility of integrating other variables of interest. In this regard, it is noteworthy that a significant number of studies meeting the inclusion criteria did not include the sex variable in their available sample metadata. For the NAFLD study, 44.5% of the studies did not specify the sex of the participants, while for the MS study, the percentage was 42.2%. This

highlights the importance of not only depositing data in open repositories but also ensuring that all relevant metadata is included and considering sex as one of them.

To address these issues, one potential solution is Open Science. Open Science is a transformative movement that aims to make scientific research, data, and methodologies accessible to everyone [124, 125, 126]. It includes various initiatives such as open access publishing, data sharing, and the development of open-source software [190]. This movement is supported by important international institutions, such as UNESCO, the European Union, and the National Academies of Sciences, Engineering, and Medicine (US) [124, 128, 130], since the principles it promotes are highly beneficial both for science itself and for the general public. First, the open publication of data enhances reproducibility, leading to more robust and high-quality scientific research. Second, allowing data to be reused accelerates innovation and optimizes the use of public funds, enabling researchers to address questions that were not initially considered in the original studies. Additionally, open publications receive greater visibility, resulting in a higher number of citations [189]. Finally, Open Science aims to make knowledge accessible to the public, contributing to the democratization of information and combating misinformation [124].

Within the Open Science movement and open data sharing, a key aspect is the FAIR principles: Findable, Accessible, Interpretable, Reusable. These principles, outlined in Wilkinson et al. [127], provide clear standards to ensure that others can easily find and access your data. Once users access the data, the principles also ensure that it is straightforward for them to understand and reuse it [189], allowing for the effective use of this valuable resource. Applying the FAIR principles, ensuring datasets with complete and standardized metadata, including variables like sex and age, would have increased the number of eligible studies in our systematic reviews. This, in turn, would have enhanced the statistical power of our meta-analysis.

Other essential aspects of Open Science include collaboration and open software. In this regard, web tools serve as vital resources, enabling less experienced researchers to conduct bioinformatic analyses. The range of web tools is extensive, from tools for graphical representation of results to platforms for much more complex analysis and pipelines [191, 192, 193]. Concerning functional meta-analyses from a sex perspective, the *MetaFun* tool [194] has recently been published by our research group, allowing this analysis to be performed without the need for bioinformatics knowledge. This web tool is available on the web at <https://bioinfo.cipf.es/metafun/> and its source code is open to any user at <https://gitlab.com/ubb-cipf/metafunweb>. Web plat-

forms also provide a user-friendly means of visualizing and accessing results. In this thesis, we have created a dedicated website for each chapter, which collects the results and graphical representations generated within those chapters. These platforms are accessible in <http://bioinfo.cipf.es/metafun-NAFLD/> and <http://bioinfo.cipf.es/metafun-MS/> for NAFLD and MS works, respectively.

The full results of this doctoral thesis are publicly available in the Zenodo repository (see Appendix I), which is supported by the OpenAIRE program [195]. This initiative was commissioned by the European Commission to promote open access and open data movements in Europe.



## Chapter 6

# Conclusions

The conclusions obtained in this thesis are as follows:

1. The functional meta-analysis of NAFLD identified significant differences in interleukins 2 and 6 between males and females during the progression from NAFL to NASH. These differences point to a different immune system modulation between males and females in disease progression.
2. Functional sex-associated changes in the progression of NAFLD may be linked to apoptosis or cellular repair pathways.
3. Meta-analyses of genes in MS blood samples found sex-specific differences in the *KIR2DL3* gene, the absence of which has been linked with the development of the disease, identifying it as a candidate gene for MS detection in liquid biopsy in males.
4. Analysis of blood and brain tissue from MS patients shows that inflammation and immune responses are present in both females and males, though they involve different target genes in each sex.
5. The functional study of the transcriptomes from brain tissue in males and females revealed distinct patterns of immunomodulation in MS, with alterations more related to the myeloid lineage and the pro-inflammatory

environment in females and more associated with the lymphoid lineage in males.

6. Interleukins 2 and 6 show different behaviors between males and females in the pathologies studied, being potential targets for future treatments of these diseases depending on the patient's sex.
7. The joint results of this thesis highlight how sex differences in the immune system significantly influence the development of nonalcoholic fatty liver disease and multiple sclerosis.
8. The set of sex-associated alterations found in this doctoral thesis, associated with nonalcoholic fatty liver disease and multiple sclerosis, represent an advance in reducing the existing sex bias in these diseases. However, further studies are still needed to translate these findings into clinical practice.
9. Systematic reviews and meta-analyses have proven effective methodologies for reducing sex bias in health and medicine, identifying changes at the genetic, functional, and metabolic pathway levels, even with the limitation of having a small number of eligible studies available in publicly accessible databases.
10. Openly accessible data with standardized patient information is still limited, which is a constraint for systematic reviews and meta-analyses.



# References

- [1] Susan P Phillips. “Defining and measuring gender: A social determinant of health whose time has come”. In: *International Journal for Equity in Health* 4.1 (2005). DOI: 10.1186/1475-9276-4-11.
- [2] María Teresa Ruiz Cantero. *Perspectiva de género en medicina*. Fundación Dr. Antoni Esteve, 2019.
- [3] Franck Mauvais-Jarvis et al. “Sex and gender: modifiers of health, disease, and medicine”. In: *The Lancet* 396.10250 (2020), pp. 565–582. ISSN: 0140-6736. DOI: [https://doi.org/10.1016/S0140-6736\(20\)31561-0](https://doi.org/10.1016/S0140-6736(20)31561-0).
- [4] Sarah Hawkes, Fariha Haseen, and Hajer Aounallah-Skhiri. “Measurement and meaning: reporting sex in health research”. In: *The Lancet* 393.10171 (2019), pp. 497–499. DOI: 10.1016/s0140-6736(19)30283-1.
- [5] Nina S. Stachenfeld and Carolyn M. Mazure. “Precision medicine requires understanding how both sex and gender influence health”. In: *Cell* 185.10 (May 2022), pp. 1619–1622. ISSN: 0092-8674. DOI: 10.1016/j.cell.2022.04.012.
- [6] Robert Mayhew. *The Female in Aristotle's Biology*. University of Chicago Press, 2004. DOI: 10.7208/chicago/9780226512020.001.0001.

- [7] Galen and Margaret Tallmadge May. *Galen on the usefulness of the parts of the body: Peri chreias moriōn [romanized form] : De usu partium*. Cornell publications in the history of science. Ithaca, N.Y.: Cornell University Press, 1968.
- [8] Susan C. Lawrence and Kae Bendixen. “His and hers: Male and female anatomy in anatomy texts for U.S. medical students, 1890–1989”. In: *Social Science & Medicine* 35.7 (1992), pp. 925–934. DOI: 10.1016/0277-9536(92)90107-2.
- [9] Yonina R. Murciano-Goroff. “Differences in the Percentage of Illustrations Showing Males Versus Females in General Medicine and General Surgery Textbooks”. In: *Medical Science Educator* 25.2 (2015), pp. 123–126. DOI: 10.1007/s40670-015-0116-8.
- [10] John Z. Ayanian and Arnold M. Epstein. “Differences in the Use of Procedures between Women and Men Hospitalized for Coronary Heart Disease”. In: *New England Journal of Medicine* 325.4 (1991), pp. 221–225. DOI: 10.1056/nejm199107253250401.
- [11] Flavia Franconi et al. “Gender differences in drug responses”. In: *Pharmacological Research* 55.2 (2007), pp. 81–95. DOI: 10.1016/j.phrs.2006.11.001.
- [12] US General Accounting Office. *Drug safety: most drugs withdrawn in recent years had greater health risks for women*. Accessed on June 29, 2022. Jan. 2001. URL: <https://www.gao.gov/assets/100/90642.pdf>.
- [13] Cassidy R Sugimoto et al. “Factors affecting sex-related reporting in medical research: a cross-disciplinary bibliometric analysis”. In: *The Lancet* 393.10171 (2019), pp. 550–559. DOI: 10.1016/s0140-6736(18)32995-7.
- [14] Cara Tannenbaum et al. “Sex and gender analysis improves science and engineering”. In: *Nature* 575.7781 (2019), pp. 137–146. DOI: 10.1038/s41586-019-1657-6.
- [15] Jessica Elaine Young, Minjie Wu, and Holly C. Hunsberger. “Editorial: Sex and gender differences in neurodegenerative diseases”. In: *Frontiers*

- in Neuroscience* 17 (Mar. 2023). ISSN: 1662-453X. DOI: 10.3389/fnins.2023.1175674.
- [16] Jill B. Becker, Brian J. Prendergast, and Jing W. Liang. “Female rats are not more variable than male rats: a meta-analysis of neuroscience studies”. In: *Biology of Sex Differences* 7.1 (2016). DOI: 10.1186/s13293-016-0087-5.
- [17] Brian J. Prendergast, Kenneth G. Onishi, and Irving Zucker. “Female mice liberated for inclusion in neuroscience and biomedical research”. In: *Neuroscience & Biobehavioral Reviews* 40 (2014), pp. 1–5. DOI: 10.1016/j.neubiorev.2014.01.001.
- [18] Margaret M. McCarthy. “Incorporating Sex as a Variable in Preclinical Neuropsychiatric Research”. In: *Schizophrenia Bulletin* 41.5 (June 2015), pp. 1016–1020. ISSN: 1745-1701. DOI: 10.1093/schbul/sbv077.
- [19] Rebecca M. Shansky and Catherine S. Woolley. “Considering Sex as a Biological Variable Will Be Valuable for Neuroscience Research”. In: *The Journal of Neuroscience* 36.47 (Nov. 2016), pp. 11817–11822. ISSN: 1529-2401. DOI: 10.1523/jneurosci.1390-16.2016.
- [20] Thorsten Buch et al. “Benefits of a factorial design focusing on inclusion of female and male animals in one experiment”. In: *Journal of Molecular Medicine* 97.6 (2019), pp. 871–877. DOI: 10.1007/s00109-019-01774-0.
- [21] Leah R. Miller et al. “Considering sex as a biological variable in pre-clinical research”. In: *The FASEB Journal* 31.1 (2016), pp. 29–34. DOI: 10.1096/fj.201600781r.
- [22] Janine A. Clayton and Francis S. Collins. “Policy: NIH to balance sex in cell and animal studies”. In: *Nature* 509.7500 (2014), pp. 282–283. DOI: 10.1038/509282a.
- [23] “Accounting for sex and gender makes for better science”. In: *Nature* 588.7837 (2020), pp. 196–196. DOI: 10.1038/d41586-020-03459-y.

- [24] Paula Franklin, Clare Bambraand, and Viviana Albani. *Gender equality and health in the EU*. Publications Office, 2021. DOI: 10.2838/991480.
- [25] Jamie White et al. “The Integration of Sex and Gender Considerations Into Biomedical Research: Lessons From International Funding Agencies”. In: *The Journal of clinical endocrinology and metabolism* 106.10 (2021), pp. 3034–3048. DOI: 10.1210/clinem/dgab434.
- [26] European Commission Decision C(2022)2975 of 10 May 2022. *Horizon Europe. Work Programme 2021-2022. 13. General Annexes*. Accessed on June 29, 2022. 2022. URL: [https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/wp-call/2021-2022/wp-13-general-annexes\\_horizon-2021-2022\\_en.pdf](https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/wp-call/2021-2022/wp-13-general-annexes_horizon-2021-2022_en.pdf).
- [27] Arthur P. Arnold. “A general theory of sexual differentiation”. In: *Journal of Neuroscience Research* 95.1-2 (2017), pp. 291–300. DOI: <https://doi.org/10.1002/jnr.23884>.
- [28] P. Koopman. “The delicate balance between male and female sex determining pathways: potential for disruption of early steps in sexual development”. In: *International Journal of Andrology* 33.2 (2010), pp. 252–258. DOI: 10.1111/j.1365-2605.2009.01001.x.
- [29] Amedeo Lonardo et al. “Sex Differences in Nonalcoholic Fatty Liver Disease: State of the Art and Identification of Research Gaps”. In: *Hepatology* 70.4 (Sept. 2019), pp. 1457–1469. ISSN: 1527-3350. DOI: 10.1002/hep.30626.
- [30] José Francisco Català-Senent et al. “A deep transcriptome meta-analysis reveals sex differences in multiple sclerosis”. In: *Neurobiology of Disease* 181 (June 2023), p. 106113. ISSN: 0969-9961. DOI: 10.1016/j.nbd.2023.106113.
- [31] Adolfo López-Cerdán et al. “Unveiling sex-based differences in Parkinson’s disease: a comprehensive meta-analysis of transcriptomic studies”. In: *Biology of Sex Differences* 13.1 (Nov. 2022). ISSN: 2042-6410. DOI: 10.1186/s13293-022-00477-5.
- [32] Karthickeyan Chella Krishnan, Margarete Mehrabian, and Aldons J. Lusis. “Sex differences in metabolism and cardiometabolic disorders”.

- In: *Current Opinion in Lipidology* 29.5 (Oct. 2018), pp. 404–410. ISSN: 1473-6535. DOI: 10.1097/mol.0000000000000536.
- [33] Natasha A. Karp et al. “Prevalence of sexual dimorphism in mammalian phenotypic traits”. In: *Nature Communications* 8.1 (June 2017). ISSN: 2041-1723. DOI: 10.1038/ncomms15475.
- [34] Mita Varghese, Jianrui Song, and Kanakadurga Singer. “Age and Sex: Impact on adipose tissue metabolism and inflammation”. In: *Mechanisms of Ageing and Development* 199 (Oct. 2021), p. 111563. ISSN: 0047-6374. DOI: 10.1016/j.mad.2021.111563.
- [35] Atila van Nas et al. “Elucidating the Role of Gonadal Hormones in Sexually Dimorphic Gene Coexpression Networks”. In: *Endocrinology* 150.3 (Mar. 2009), pp. 1235–1249. ISSN: 1945-7170. DOI: 10.1210/en.2008-0563.
- [36] Stefano Ballestri 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”. In: *Advances in Therapy* 34.6 (May 2017), pp. 1291–1326. ISSN: 1865-8652. DOI: 10.1007/s12325-017-0556-1.
- [37] Michela Traglia et al. “Genetic Mechanisms Leading to Sex Differences Across Common Diseases and Anthropometric Traits”. In: *Genetics* 205.2 (Feb. 2017), pp. 979–992. ISSN: 1943-2631. DOI: 10.1534/genetics.116.193623.
- [38] Cene Skubic, Živa Drakulić, and Damjana Rozman. “Personalized therapy when tackling nonalcoholic fatty liver disease: a focus on sex, genes, and drugs”. In: *Expert Opinion on Drug Metabolism and Toxicology* 14.8 (July 2018), pp. 831–841. ISSN: 1744-7607. DOI: 10.1080/17425255.2018.1492552.
- [39] Daoshan Zheng et al. “Genomics of sex hormone receptor signaling in hepatic sexual dimorphism”. In: *Molecular and Cellular Endocrinology* 471 (Aug. 2018), pp. 33–41. ISSN: 0303-7207. DOI: 10.1016/j.mce.2017.05.025.

- [40] Marisa Coelho, Teresa Oliveira, and Ruben Fernandes. “Biochemistry of adipose tissue: an endocrine organ”. In: *Archives of Medical Science* 2 (2013), pp. 191–200. ISSN: 1734-1922. DOI: 10.5114/aoms.2013.33181.
- [41] S Lemieux et al. “Sex differences in the relation of visceral adipose tissue accumulation to total body fatness”. In: *The American Journal of Clinical Nutrition* 58.4 (Oct. 1993), pp. 463–467. ISSN: 0002-9165. DOI: 10.1093/ajcn/58.4.463.
- [42] Susanne B. Votruba and Michael D. Jensen. “Sex differences in abdominal, gluteal, and thigh LPL activity”. In: *American Journal of Physiology-Endocrinology and Metabolism* 292.6 (June 2007), E1823–E1828. ISSN: 1522-1555. DOI: 10.1152/ajpendo.00601.2006.
- [43] Kathleen M. Gavin and Daniel H. Bessesen. “Sex Differences in Adipose Tissue Function”. In: *Endocrinology and Metabolism Clinics of North America* 49.2 (June 2020), pp. 215–228. ISSN: 0889-8529. DOI: 10.1016/j.ec1.2020.02.008.
- [44] Franck Mauvais-Jarvis. “Gender differences in glucose homeostasis and diabetes”. In: *Physiology & Behavior* 187 (Apr. 2018), pp. 20–23. ISSN: 0031-9384. DOI: 10.1016/j.physbeh.2017.08.016.
- [45] Elena Ortona et al. “Sex-based differences in autoimmune diseases”. In: *Annali dell’Istituto Superiore di Sanità* 52 (2016). DOI: 10.4415/ANN\_16\_02\_12.
- [46] Lifeng Wang, Fu-Sheng Wang, and M. Eric Gershwin. “Human autoimmune diseases: a comprehensive update”. In: *Journal of Internal Medicine* 278.4 (2015), pp. 369–395. DOI: 10.1111/joim.12395.
- [47] Vanessa L. Kronzer, Stanley Louis Bridges, and John M. Davis. “Why women have more autoimmune diseases than men: An evolutionary perspective”. In: *Evolutionary Applications* 14.3 (2020), pp. 629–633. DOI: 10.1111/eva.13167.
- [48] “Autoimmune diseases: a leading cause of death among young and middle-aged women in the United States”. In: *American Journal of Public Health* 90.9 (2000), pp. 1463–1466. DOI: 10.2105/ajph.90.9.1463.

- 
- [49] Damian N. Di Florio et al. “Sex differences in inflammation, redox biology, mitochondria and autoimmunity”. In: *Redox Biology* 31 (2020). Sexual Dimorphisms in Redox Biology, p. 101482. ISSN: 2213-2317. DOI: <https://doi.org/10.1016/j.redox.2020.101482>.
- [50] Margaret LP Teng et al. “Global incidence and prevalence of nonalcoholic fatty liver disease”. In: *Clinical and Molecular Hepatology* 29.Suppl (Feb. 2023), S32–S42. ISSN: 2287-285X. DOI: 10.3350/cmh.2022.0365.
- [51] M. J. Rowley and I. R. Mackay. “Measurement of antibody-producing capacity in man. I. The normal response to flagellin from *Salmonella adelaide*.” In: *Clinical and experimental immunology* 5 (4 Oct. 1969), pp. 407–418. ISSN: 0009-9104. ppublish.
- [52] Howard C. Crisp and James M. Quinn. “Quantitative immunoglobulins in adulthood”. In: *Allergy and Asthma Proceedings* 30.6 (2009), pp. 649–654. DOI: 10.2500/aap.2009.30.3292.
- [53] Charles F. Dillon, Michael H. Weisman, and Frederick W. Miller. “Population-based estimates of humoral autoimmunity from the U.S. National Health and Nutrition Examination Surveys, 1960–2014”. In: *PLOS ONE* 15.1 (2020). Ed. by Jaymie Meliker, e0226516. DOI: 10.1371/journal.pone.0226516.
- [54] Sabra L. Klein and Katie L. Flanagan. “Sex differences in immune responses”. In: *Nature Reviews Immunology* 16.10 (Aug. 2016), pp. 626–638. ISSN: 1474-1741. DOI: 10.1038/nri.2016.90.
- [55] Caleb E. Finch. “Evolution of the human lifespan and diseases of aging: Roles of infection, inflammation, and nutrition”. In: *Proceedings of the National Academy of Sciences* 107.suppl\_1 (2009), pp. 1718–1724. DOI: 10.1073/pnas.0909606106.
- [56] Mübeccel Akdis et al. “Interleukins, from 1 to 37, and interferon- $\gamma$ : Receptors, functions, and roles in diseases”. In: *Journal of Allergy and Clinical Immunology* 127.3 (Mar. 2011), 701–721.e70. ISSN: 0091-6749. DOI: 10.1016/j.jaci.2010.11.050.
- [57] Angel A. Justiz Vaillant and Ahmad Qurie. *Interleukin*. StatPearls Publishing, 2024.

- [58] JA Giron-Gonzalez et al. “Consistent production of a higher TH1:TH2 cytokine ratio by stimulated T cells in men compared with women”. In: *European Journal of Endocrinology* (July 2000), pp. 31–36. ISSN: 1479-683X. DOI: 10.1530/eje.0.1430031.
- [59] Joana Guerra et al. “Muscular interleukin-6 differentially regulates skeletal muscle adaptation to high-fat diet in a sex-dependent manner”. In: *Cytokine* 74.1 (July 2015), pp. 145–151. ISSN: 1043-4666. DOI: 10.1016/j.cyto.2015.04.018.
- [60] Andrea M. Spiker et al. “Interleukin-6 and Interleukin-8 Gene Expressions Differ Between Male and Female Patients at Time of Hip Arthroscopy for Femoroacetabular Impingement Syndrome”. In: *Arthroscopy, Sports Medicine, and Rehabilitation* (July 2024), p. 100985. ISSN: 2666-061X. DOI: 10.1016/j.asmr.2024.100985.
- [61] Anna Zychlinsky Scharff et al. “Sex differences in IL-17 contribute to chronicity in male versus female urinary tract infection”. In: *JCI Insight* 4.13 (July 2019). ISSN: 2379-3708. DOI: 10.1172/jci.insight.122998.
- [62] Hubert Schorle et al. “Development and function of T cells in mice rendered interleukin-2 deficient by gene targeting”. In: *Nature* 352.6336 (Aug. 1991), pp. 621–624. ISSN: 1476-4687. DOI: 10.1038/352621a0.
- [63] Thomas M. Kündig et al. “Immune Responses in Interleukin-2-Deficient Mice”. In: *Science* 262.5136 (Nov. 1993), pp. 1059–1061. ISSN: 1095-9203. DOI: 10.1126/science.8235625.
- [64] Katherine C. Dodd and Madhvi Menon. “Sex bias in lymphocytes: Implications for autoimmune diseases”. In: *Frontiers in Immunology* 13 (Nov. 2022). ISSN: 1664-3224. DOI: 10.3389/fimmu.2022.945762.
- [65] Claude Libert, Lien Dejager, and Iris Pinheiro. “The X chromosome in immune functions: when a chromosome makes the difference”. In: *Nature Reviews Immunology* 10.8 (Aug. 2010), pp. 594–604. ISSN: 1474-1741. DOI: 10.1038/nri2815.
- [66] Laura Carrel and Huntington F. Willard. “X-inactivation profile reveals extensive variability in X-linked gene expression in females”. In: *Nature*



- 
- 434.7031 (Mar. 2005), pp. 400–404. ISSN: 1476-4687. DOI: 10.1038/nature03479.
- [67] Shahnaz Hosseinzadeh et al. “The role of genetics and gender specific differences in neurodegenerative disorders: Insights from molecular and immune landscape”. In: *Journal of Neuroimmunology* 384 (Nov. 2023), p. 578206. ISSN: 0165-5728. DOI: 10.1016/j.jneuroim.2023.578206.
  - [68] Iain L Campbell et al. “Structural and functional neuropathology in transgenic mice with CNS expression of IFN- $\alpha$ ”. In: *Brain Research* 835.1 (July 1999), pp. 46–61. ISSN: 0006-8993. DOI: 10.1016/S0006-8993(99)01328-1.
  - [69] Pamela A. McCombe and Robert D. Henderson. “Effects of gender in amyotrophic lateral sclerosis”. In: *Gender Medicine* 7.6 (Dec. 2010), pp. 557–570. ISSN: 1550-8579. DOI: 10.1016/j.genm.2010.11.010.
  - [70] Alzheimer’s association report. “2023 Alzheimer’s disease facts and figures”. In: *Alzheimer’s & Dementia* 19.4 (Mar. 2023), pp. 1598–1695. ISSN: 1552-5279. DOI: 10.1002/alz.13016.
  - [71] Chloe Lopez-Lee, Lay Kodama, and Li Gan. “Sex Differences in Neurodegeneration: The Role of the Immune System in Humans”. In: *Biological Psychiatry* 91.1 (Jan. 2022), pp. 72–80. ISSN: 0006-3223. DOI: 10.1016/j.biopsych.2021.01.002.
  - [72] Timothy R. Hammond et al. “Single-Cell RNA Sequencing of Microglia throughout the Mouse Lifespan and in the Injured Brain Reveals Complex Cell-State Changes”. In: *Immunity* 50.1 (Jan. 2019), 253–271.e6. ISSN: 1074-7613. DOI: 10.1016/j.immuni.2018.11.004.
  - [73] Prapaporn Pisitkun et al. “Autoreactive B Cell Responses to RNA-Related Antigens Due to TLR7 Gene Duplication”. In: *Science* 312.5780 (June 2006), pp. 1669–1672. ISSN: 1095-9203. DOI: 10.1126/science.1124978.
  - [74] Antoine Louveau et al. “Structural and functional features of central nervous system lymphatic vessels”. In: *Nature* 523.7560 (June 2015), pp. 337–341. ISSN: 1476-4687. DOI: 10.1038/nature14432.

- [75] Marco Prinz and Josef Priller. “The role of peripheral immune cells in the CNS in steady state and disease”. In: *Nature Neuroscience* 20.2 (Jan. 2017), pp. 136–144. ISSN: 1546-1726. DOI: 10.1038/nn.4475.
- [76] Ilia D. Vainchtein and Anna V. Molofsky. “Astrocytes and Microglia: In Sickness and in Health”. In: *Trends in Neurosciences* 43.3 (Mar. 2020), pp. 144–154. ISSN: 0166-2236. DOI: 10.1016/j.tins.2020.01.003.
- [77] L.J. Lawson et al. “Heterogeneity in the distribution and morphology of microglia in the normal adult mouse brain”. In: *Neuroscience* 39.1 (Jan. 1990), pp. 151–170. ISSN: 0306-4522. DOI: 10.1016/0306-4522(90)90229-w.
- [78] Carole Sousa, Knut Biber, and Alessandro Michelucci. “Cellular and Molecular Characterization of Microglia: A Unique Immune Cell Population”. In: *Frontiers in Immunology* 8 (Mar. 2017). ISSN: 1664-3224. DOI: 10.3389/fimmu.2017.00198.
- [79] Alessandra Maria Casano and Francesca Peri. “Microglia: Multitasking Specialists of the Brain”. In: *Developmental Cell* 32.4 (Feb. 2015), pp. 469–477. ISSN: 1534-5807. DOI: 10.1016/j.devcel.2015.01.018.
- [80] Rosa C. Paolicelli et al. “Synaptic Pruning by Microglia Is Necessary for Normal Brain Development”. In: *Science* 333.6048 (Sept. 2011), pp. 1456–1458. ISSN: 1095-9203. DOI: 10.1126/science.1202529.
- [81] Dilansu Guneykaya et al. “Transcriptional and Translational Differences of Microglia from Male and Female Brains”. In: *Cell Reports* 24.10 (Sept. 2018), 2773–2783.e6. ISSN: 2211-1247. DOI: 10.1016/j.celrep.2018.08.001.
- [82] Yasmine V. Doust, Anna E. King, and Jenna M. Ziebell. “Implications for microglial sex differences in tau-related neurodegenerative diseases”. In: *Neurobiology of Aging* 105 (Sept. 2021), pp. 340–348. ISSN: 0197-4580. DOI: 10.1016/j.neurobiolaging.2021.03.010.
- [83] Rosalind E. Franklin and R. G. Gosling. “Molecular Configuration in Sodium Thymonucleate”. In: *Nature* 171.4356 (Apr. 1953), pp. 740–741. ISSN: 1476-4687. DOI: 10.1038/171740a0.

- 
- [84] J. D. Watson and F. H. C. Crick. “Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid”. In: *Nature* 171.4356 (Apr. 1953), pp. 737–738. ISSN: 1476-4687. DOI: 10.1038/171737a0.
- [85] K. Mullis et al. “Specific Enzymatic Amplification of DNA In Vitro: The Polymerase Chain Reaction”. In: *Cold Spring Harbor Symposia on Quantitative Biology* 51.0 (Jan. 1986), pp. 263–273. ISSN: 1943-4456. DOI: 10.1101/sqb.1986.051.01.032.
- [86] W. Min Jou et al. “Nucleotide Sequence of the Gene Coding for the Bacteriophage MS2 Coat Protein”. In: *Nature* 237.5350 (May 1972), pp. 82–88. ISSN: 1476-4687. DOI: 10.1038/237082a0.
- [87] F. Sanger, S. Nicklen, and A. R. Coulson. “DNA sequencing with chain-terminating inhibitors”. In: *Proceedings of the National Academy of Sciences* 74.12 (Dec. 1977), pp. 5463–5467. ISSN: 1091-6490. DOI: 10.1073/pnas.74.12.5463.
- [88] Daniel Nathans and Hamilton O. Smith. “Restriction Endonucleases in the Analysis and Restructuring of DNA Molecules”. In: *Annual Review of Biochemistry* 44.1 (June 1975), pp. 273–293. ISSN: 1545-4509. DOI: 10.1146/annurev.bi.44.070175.001421.
- [89] Francisco J. M. Mojica and Francisco Rodriguez-Valera. “The discovery of CRISPR in archaea and bacteria”. In: *The FEBS Journal* 283.17 (June 2016), pp. 3162–3169. ISSN: 1742-4658. DOI: 10.1111/febs.13766.
- [90] Jennifer A. Doudna and Emmanuelle Charpentier. “The new frontier of genome engineering with CRISPR-Cas9”. In: *Science* 346.6213 (Nov. 2014). ISSN: 1095-9203. DOI: 10.1126/science.1258096.
- [91] Mario Vailati-Riboni, Valentino Palombo, and Juan J. Llor. “What Are Omics Sciences?” In: *Periparturient Diseases of Dairy Cows*. Springer International Publishing, 2017, pp. 1–7. ISBN: 9783319430331. DOI: 10.1007/978-3-319-43033-1\_1.
- [92] Bob Kuska. “Beer, Bethesda, and Biology: How “Genomics” Came Into Being”. In: *JNCI: Journal of the National Cancer Institute* 90.2 (1998), pp. 93–93. ISSN: 0027-8874. DOI: 10.1093/jnci/90.2.93.

- [93] Francis S. Collins et al. “A vision for the future of genomics research”. In: *Nature* 422.6934 (Apr. 2003), pp. 835–847. ISSN: 1476-4687. DOI: 10.1038/nature01626.
- [94] Anjana Munshi and Vandana Sharma. “Omics and Edible Vaccines”. In: *Omics Technologies and Bio-Engineering*. Elsevier, 2018, pp. 129–141. DOI: 10.1016/b978-0-12-815870-8.00008-5.
- [95] Zhong Wang, Mark Gerstein, and Michael Snyder. “RNA-Seq: a revolutionary tool for transcriptomics”. In: *Nature Reviews Genetics* 10.1 (Jan. 2009), pp. 57–63. ISSN: 1471-0064. DOI: 10.1038/nrg2484.
- [96] Alexey I Nesvizhskii. “Proteogenomics: concepts, applications and computational strategies”. In: *Nature Methods* 11.11 (Oct. 2014), pp. 1114–1125. ISSN: 1548-7105. DOI: 10.1038/nmeth.3144.
- [97] Clary B. Clish. “Metabolomics: an emerging but powerful tool for precision medicine”. In: *Molecular Case Studies* 1.1 (Sept. 2015), a000588. ISSN: 2373-2873. DOI: 10.1101/mcs.a000588.
- [98] Michelle R. Rondon et al. “Cloning the Soil Metagenome: a Strategy for Accessing the Genetic and Functional Diversity of Uncultured Microorganisms”. In: *Applied and Environmental Microbiology* 66.6 (June 2000), pp. 2541–2547. ISSN: 1098-5336. DOI: 10.1128/aem.66.6.2541-2547.2000.
- [99] John C. Wooley, Adam Godzik, and Iddo Friedberg. “A Primer on Metagenomics”. In: *PLoS Computational Biology* 6.2 (Feb. 2010). Ed. by Philip E. Bourne, e1000667. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1000667.
- [100] Shankar Subramaniam et al. “Bioinformatics and Systems Biology of the Lipidome”. In: *Chemical Reviews* 111.10 (Sept. 2011), pp. 6452–6490. ISSN: 1520-6890. DOI: 10.1021/cr200295k.
- [101] Michal Krassowski et al. “State of the Field in Multi-Omics Research: From Computational Needs to Data Mining and Sharing”. In: *Frontiers in Genetics* 11 (Dec. 2020). ISSN: 1664-8021. DOI: 10.3389/fgene.2020.610798.

- 
- [102] Mark D. Adams et al. “Complementary DNA Sequencing: Expressed Sequence Tags and Human Genome Project”. In: *Science* 252.5013 (June 1991), pp. 1651–1656. ISSN: 1095-9203. DOI: 10.1126/science.2047873.
- [103] Marco A. Marra, Ladeana Hillier, and Robert H. Waterston. “Expressed sequence tags — ESTablishing bridges between genomes”. In: *Trends in Genetics* 14.1 (Jan. 1998), pp. 4–7. ISSN: 0168-9525. DOI: 10.1016/S0168-9525(97)01355-3.
- [104] Rohan Lowe et al. “Transcriptomics technologies”. In: *PLOS Computational Biology* 13.5 (May 2017), e1005457. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1005457.
- [105] Victor E Velculescu et al. “Characterization of the Yeast Transcriptome”. In: *Cell* 88.2 (Jan. 1997), pp. 243–251. ISSN: 0092-8674. DOI: 10.1016/S0092-8674(00)81845-0.
- [106] Victor E. Velculescu et al. “Serial Analysis of Gene Expression”. In: *Science* 270.5235 (Oct. 1995), pp. 484–487. ISSN: 1095-9203. DOI: 10.1126/science.270.5235.484.
- [107] Aruna Pal and A.K. Chakravarty. “Advanced genomic techniques for studying immune-response genes”. In: *Genetics and Breeding for Disease Resistance of Livestock*. Elsevier, 2020, pp. 209–234. ISBN: 9780128164068. DOI: 10.1016/B978-0-12-816406-8.00014-0.
- [108] Bo Wang et al. “Reviving the Transcriptome Studies: An Insight Into the Emergence of Single-Molecule Transcriptome Sequencing”. In: *Frontiers in Genetics* 10 (Apr. 2019). ISSN: 1664-8021. DOI: 10.3389/fgene.2019.00384.
- [109] David J. Lockhart et al. “Expression monitoring by hybridization to high-density oligonucleotide arrays”. In: *Nature Biotechnology* 14.13 (Dec. 1996), pp. 1675–1680. ISSN: 1546-1696. DOI: 10.1038/nbt1296-1675.
- [110] Michael B. Black et al. “Comparison of Microarrays and RNA-Seq for Gene Expression Analyses of Dose-Response Experiments”. In: *Toxico-*

- logical Sciences* 137.2 (Nov. 2013), pp. 385–403. ISSN: 1096-0929. DOI: 10.1093/toxsci/kft249.
- [111] Ana Conesa et al. “A survey of best practices for RNA-seq data analysis”. In: *Genome Biology* 17.1 (Jan. 2016). ISSN: 1474-760X. DOI: 10.1186/s13059-016-0881-8.
  - [112] Fuchou Tang et al. “mRNA-Seq whole-transcriptome analysis of a single cell”. In: *Nature Methods* 6.5 (Apr. 2009), pp. 377–382. ISSN: 1548-7105. DOI: 10.1038/nmeth.1315.
  - [113] Patrik L. Ståhl et al. “Visualization and analysis of gene expression in tissue sections by spatial transcriptomics”. In: *Science* 353.6294 (July 2016), pp. 78–82. ISSN: 1095-9203. DOI: 10.1126/science.aaf2403.
  - [114] Sophia K. Longo et al. “Integrating single-cell and spatial transcriptomics to elucidate intercellular tissue dynamics”. In: *Nature Reviews Genetics* 22.10 (June 2021), pp. 627–644. ISSN: 1471-0064. DOI: 10.1038/s41576-021-00370-8.
  - [115] Anton Petrov and Soheil Shams. “Microarray Image Processing and Quality Control”. In: *The Journal of VLSI Signal Processing-Systems for Signal, Image, and Video Technology* 38.3 (Nov. 2004), pp. 211–226. ISSN: 0922-5773. DOI: 10.1023/b:vlsi.0000042488.08307.ad.
  - [116] Xiangqin Cui and Gary A Churchill. “Statistical tests for differential expression in cDNA microarray experiments”. In: *Genome Biology* 4.4 (Apr. 2003). ISSN: 1474-760X. DOI: 10.1186/gb-2003-4-4-210.
  - [117] Dongmei Li et al. “An evaluation of RNA-seq differential analysis methods”. In: *PLOS ONE* 17.9 (Sept. 2022). Ed. by Dov Joseph Stekel, e0264246. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0264246.
  - [118] Michael Ashburner et al. “Gene Ontology: tool for the unification of biology”. In: *Nature Genetics* 25.1 (May 2000), pp. 25–29. ISSN: 1546-1718. DOI: 10.1038/75556.

- 
- [119] M. Kanehisa. “KEGG: Kyoto Encyclopedia of Genes and Genomes”. In: *Nucleic Acids Research* 28.1 (Jan. 2000), pp. 27–30. ISSN: 1362-4962. DOI: 10.1093/nar/28.1.27.
- [120] Marija Milacic et al. “The Reactome Pathway Knowledgebase 2024”. In: *Nucleic Acids Research* 52.D1 (Nov. 2023), pp. D672–D678. ISSN: 1362-4962. DOI: 10.1093/nar/gkad1025.
- [121] David Montaner and Joaquín Dopazo. “Multidimensional Gene Set Analysis of Genomic Data”. In: *PLoS ONE* 5.4 (Apr. 2010). Ed. by Jörg Hoheisel, e10348. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0010348.
- [122] Marta R. Hidalgo et al. “High throughput estimation of functional cell activities reveals disease mechanisms and predicts relevant clinical outcomes”. In: *Oncotarget* 8.3 (Dec. 2016), pp. 5160–5178. ISSN: 1949-2553. DOI: 10.18632/oncotarget.14107.
- [123] Emily Clough et al. “NCBI GEO: archive for gene expression and epigenomics data sets: 23-year update”. In: *Nucleic Acids Research* 52.D1 (Nov. 2023), pp. D138–D144. ISSN: 1362-4962. DOI: 10.1093/nar/gkad965.
- [124] *UNESCO Recommendation on Open Science*. 2021. DOI: 10.54677/mnmh8546.
- [125] B. A. Nosek et al. “Promoting an open research culture”. In: *Science* 348.6242 (June 2015), pp. 1422–1425. ISSN: 1095-9203. DOI: 10.1126/science.aab2374.
- [126] Kevin C. Elliott and David B. Resnik. “Making Open Science Work for Science and Society”. In: *Environmental Health Perspectives* 127.7 (July 2019). ISSN: 1552-9924. DOI: 10.1289/ehp4808.
- [127] Mark D. Wilkinson et al. “The FAIR Guiding Principles for scientific data management and stewardship”. In: *Scientific Data* 3.1 (2016). DOI: 10.1038/sdata.2016.18.

- [128] European Commission. Directorate General for Research and Innovation. *Horizon Europe, open science: early knowledge and data sharing, and open collaboration*. Publications Office, 2021. DOI: 10.2777/18252.
- [129] European Commission. Directorate General for Research and Innovation. *Open innovation, open science, open to the world: a vision for Europe*. Publications Office, 2016. DOI: 10.2777/061652.
- [130] *Open Science by Design: Realizing a Vision for 21st Century Research*. National Academies Press, Aug. 2018. ISBN: 9780309476249. DOI: 10.17226/25116.
- [131] J. E. Hunter and F. L. Schmidt. *Methods of meta-analysis: Correcting error and bias in research findings*. Sage Publications, Inc, 1990.
- [132] Alessandro Liberati et al. “The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration”. In: *Journal of Clinical Epidemiology* 62.10 (Oct. 2009), e1–e34. ISSN: 0895-4356. DOI: 10.1016/j.jclinepi.2009.06.006.
- [133] Guillaume Lame. “Systematic Literature Reviews: An Introduction”. In: *Proceedings of the Design Society: International Conference on Engineering Design* 1.1 (July 2019), pp. 1633–1642. ISSN: 2220-4342. DOI: 10.1017/dsi.2019.169.
- [134] Matthew J Page et al. “The PRISMA 2020 statement: an updated guideline for reporting systematic reviews”. In: *BMJ* (Mar. 2021), n71. ISSN: 1756-1833. DOI: 10.1136/bmj.n71.
- [135] James Lind. *A treatise on the scurvy, in three parts: containing an inquiry into the nature, causes, and cure, of that disease, together with a critical and chronological view of what has been published on the subject*. 1757. DOI: 10.5281/zenodo.12783573.
- [136] Quan Nha Hong and Pierre Pluye. “Systematic reviews: A brief historical overview”. In: *Education for Information* 34.4 (Dec. 2018), pp. 261–276. ISSN: 1875-8649. DOI: 10.3233/efi-180219.



- 
- [137] Iain Chalmers, Larry V. Hedges, and Harris Cooper. “A Brief History of Research Synthesis”. In: *Evaluation & the Health Professions* 25.1 (Mar. 2002), pp. 12–37. ISSN: 1552-3918. DOI: 10.1177/0163278702025001003.
- [138] David Moher et al. “Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement”. In: *PLoS Medicine* 6.7 (July 2009), e1000097. ISSN: 1549-1676. DOI: 10.1371/journal.pmed.1000097.
- [139] Matthew J Page et al. “PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews”. In: *BMJ* (Mar. 2021), n160. ISSN: 1756-1833. DOI: 10.1136/bmj.n160.
- [140] Henry S. Sacks et al. “Meta-Analyses of Randomized Controlled Trials”. In: *New England Journal of Medicine* 316.8 (Feb. 1987), pp. 450–455. ISSN: 1533-4406. DOI: 10.1056/nejm198702193160806.
- [141] Ferrán Catalá-López and Aurelio Tobías. “Metaanálisis de ensayos clínicos aleatorizados, heterogeneidad e intervalos de predicción”. In: *Medicina Clínica* 142.6 (Mar. 2014), pp. 270–274. ISSN: 0025-7753. DOI: 10.1016/j.medcli.2013.06.013.
- [142] Sharon-Lise T. Normand. “Meta-analysis: formulating, evaluating, combining, and reporting”. In: *Statistics in Medicine* 18.3 (1999), pp. 321–359. DOI: [https://doi.org/10.1002/\(SICI\)1097-0258\(19990215\)18:3<321::AID-SIM28>3.0.CO;2-P](https://doi.org/10.1002/(SICI)1097-0258(19990215)18:3<321::AID-SIM28>3.0.CO;2-P).
- [143] Jessica Gurevitch et al. “Meta-analysis and the science of research synthesis”. In: *Nature* 555.7695 (Mar. 2018), pp. 175–182. ISSN: 1476-4687. DOI: 10.1038/nature25753.
- [144] Gene V. Glass. “Primary, Secondary, and Meta-Analysis of Research”. In: *Educational Researcher* 5.10 (Nov. 1976), p. 3. ISSN: 0013-189X. DOI: 10.2307/1174772.
- [145] Joseph R. Dettoni, Daniel C. Norvell, and Jens R. Chapman. “Fixed-Effect vs Random-Effects Models for Meta-Analysis: 3 Points to Consider”. In: *Global Spine Journal* 12.7 (June 2022), pp. 1624–1626. ISSN: 2192-5690. DOI: 10.1177/21925682221110527.

- [146] Steve Kanters. “Fixed- and Random-Effects Models”. In: *Meta-Research*. Springer US, Sept. 2021, pp. 41–65. ISBN: 9781071615669. DOI: 10.1007/978-1-0716-1566-9\_3.
- [147] U.S. Department of Health and Human Services (HHS). *PubMed "sex differences" query*. Accessed on September 06, 2024.
- [148] Arthur P Arnold. “Promoting the understanding of sex differences to enhance equity and excellence in biomedical science”. In: *Biology of Sex Differences* 1.1 (2010), p. 1. ISSN: 2042-6410. DOI: 10.1186/2042-6410-1-1.
- [149] Malachy Bishop and Phillip D. Rumrill. “Multiple sclerosis: Etiology, symptoms, incidence and prevalence, and implications for community living and employment”. In: *Work* 52.4 (Nov. 2015). Ed. by Malachy Bishop and Phillip D. Rumrill, pp. 725–734. ISSN: 1875-9270. DOI: 10.3233/wor-152200.
- [150] Maider Muñoz-Culla et al. “SncRNA (microRNA & snoRNA) opposite expression pattern found in multiple sclerosis relapse and remission is sex dependent”. In: *Scientific Reports* 6.1 (Feb. 2016). ISSN: 2045-2322. DOI: 10.1038/srep20126.
- [151] Rhonda R Voskuhl. “The effect of sex on multiple sclerosis risk and disease progression”. In: *Multiple Sclerosis Journal* 26.5 (Jan. 2020), pp. 554–560. ISSN: 1477-0970. DOI: 10.1177/1352458519892491.
- [152] Mahmoud Samadi et al. “Cancer immunotherapy focusing on the role of interleukins: A comprehensive and updated study”. In: *Pathology - Research and Practice* 249 (Sept. 2023), p. 154732. ISSN: 0344-0338. DOI: 10.1016/j.prp.2023.154732.
- [153] Sarah L. Gaffen and Kathleen D. Liu. “Overview of interleukin-2 function, production and clinical applications”. In: *Cytokine* 28.3 (Nov. 2004), pp. 109–123. ISSN: 1043-4666. DOI: 10.1016/j.cyto.2004.06.010.
- [154] Thomas Hünig and Anneliese Schimpl. “The IL-2 Deficiency Syndrome”. In: *Cytokine Knockouts*. Ed. by Scott K. Durum and Kathrin Muegge.

- Totowa, NJ: Humana Press, 1998, pp. 1–19. ISBN: 978-1-4757-2753-1. DOI: 10.1007/978-1-4757-2753-1.
- [155] Daphne Peerlings, Max Mimpfen, and Jan Damoiseaux. “The IL-2 – IL-2 receptor pathway: Key to understanding multiple sclerosis”. In: *Journal of Translational Autoimmunity* 4 (2021), p. 100123. ISSN: 2589-9090. DOI: 10.1016/j.jtauto.2021.100123.
- [156] C. Louapre et al. “A randomized double-blind placebo-controlled trial of low-dose interleukin-2 in relapsing–remitting multiple sclerosis”. In: *Journal of Neurology* 270.9 (May 2023), pp. 4403–4414. ISSN: 1432-1459. DOI: 10.1007/s00415-023-11690-6.
- [157] Marko Petržalka et al. “IL-2, IL-6 and chitinase 3-like 2 might predict early relapse activity in multiple sclerosis”. In: *PLOS ONE* 17.6 (June 2022). Ed. by Carmen Infante-Duarte, e0270607. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0270607.
- [158] Santosh Saxena et al. “Interleukin 2 activity in chronic liver disease and the effect of in vitro alpha-interferon.” In: *Clinical & Experimental Immunology* 63.3 (1986), pp. 541–548.
- [159] Emily R. Perito et al. “Association between cytokines and liver histology in children with nonalcoholic fatty liver disease”. In: *Hepatology Communications* 1.7 (July 2017), pp. 609–622. ISSN: 2471-254X. DOI: 10.1002/hep4.1068.
- [160] Wei-Yu Kao et al. “Interleukin-2 receptor alpha as a biomarker for non-alcoholic fatty liver disease diagnosis”. In: *Journal of the Chinese Medical Association* 84.3 (Mar. 2021), pp. 261–266. ISSN: 1726-4901. DOI: 10.1097/jcma.0000000000000469.
- [161] Michelle Rothaug, Christoph Becker-Pauly, and Stefan Rose-John. “The role of interleukin-6 signaling in nervous tissue”. In: *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* 1863.6 (June 2016), pp. 1218–1227. ISSN: 0167-4889. DOI: 10.1016/j.bbamcr.2016.03.018.
- [162] Kai Zhu et al. “The interleukin-6 regulates the function of monocytes/macrophages (MO/M $\phi$ ) via the interleukin-6 receptor  $\beta$  in ayu

- (*Plecoglossus altivelis*)". In: *Fish & Shellfish Immunology* 93 (Oct. 2019), pp. 191–199. ISSN: 1050-4648. DOI: 10.1016/j.fsi.2019.07.049.
- [163] Timothy M. Kistner, Bente K. Pedersen, and Daniel E. Lieberman. "Interleukin 6 as an energy allocator in muscle tissue". In: *Nature Metabolism* 4.2 (Feb. 2022), pp. 170–179. ISSN: 2522-5812. DOI: 10.1038/s42255-022-00538-4.
- [164] Dirk Schmidt-Arras and Stefan Rose-John. "IL-6 pathway in the liver: From physiopathology to therapy". In: *Journal of Hepatology* 64.6 (June 2016), pp. 1403–1415. ISSN: 0168-8278. DOI: 10.1016/j.jhep.2016.02.004.
- [165] Weimin Lin et al. "Functional role of skeletal muscle-derived interleukin-6 and its effects on lipid metabolism". In: *Frontiers in Physiology* 14 (July 2023). ISSN: 1664-042X. DOI: 10.3389/fphys.2023.1110926.
- [166] Irina Kerkis, Álvaro Prieto da Silva, and Rodrigo Pinheiro Araldi. "The impact of interleukin-6 (IL-6) and mesenchymal stem cell-derived IL-6 on neurological conditions". In: *Frontiers in Immunology* 15 (June 2024). ISSN: 1664-3224. DOI: 10.3389/fimmu.2024.1400533.
- [167] Vincent Braunersreuther. "Role of cytokines and chemokines in non-alcoholic fatty liver disease". In: *World Journal of Gastroenterology* 18.8 (2012), p. 727. ISSN: 1007-9327. DOI: 10.3748/wjg.v18.i8.727.
- [168] Anna Wieckowska et al. "Increased Hepatic and Circulating Interleukin-6 Levels in Human Nonalcoholic Steatohepatitis". In: *The American Journal of Gastroenterology* 103.6 (June 2008), pp. 1372–1379. ISSN: 1572-0241. DOI: 10.1111/j.1572-0241.2007.01774.x.
- [169] Benedetta Vai et al. "Interleukin 6 associates with reduced grey matter volume and resting-state connectivity in the anterior cingulate cortex in bipolar patients". In: *Brain, Behavior, & Immunity - Health* 26 (Dec. 2022), p. 100522. ISSN: 2666-3546. DOI: 10.1016/j.bbih.2022.100522.
- [170] Yamei Duan et al. "Association of Inflammatory Cytokines With Non-Alcoholic Fatty Liver Disease". In: *Frontiers in Immunology* 13 (May 2022). ISSN: 1664-3224. DOI: 10.3389/fimmu.2022.880298.

- 
- [171] Xiaofang Zhang et al. “Interleukin-6 promotes visceral adipose tissue accumulation during aging via inhibiting fat lipolysis”. In: *International Immunopharmacology* 132 (May 2024), p. 111906. ISSN: 1567-5769. DOI: 10.1016/j.intimp.2024.111906.
- [172] Maria Erta et al. “Astrocytic IL-6 Influences the Clinical Symptoms of EAE in Mice”. In: *Brain Sciences* 6.2 (May 2016), p. 15. ISSN: 2076-3425. DOI: 10.3390/brainsci6020015.
- [173] Carlos Fredy Cuevas-García, Nora Hilda Segura-Méndez, and Diana Andrea Herrera-Sánchez. “Actualidades en la inmunopatología de la esclerosis múltiple”. In: *Gaceta Médica de México* 154.5 (Sept. 2018). ISSN: 0016-3813. DOI: 10.24875/gmm.18003407.
- [174] Martina Kunkl et al. “T Helper Cells: The Modulators of Inflammation in Multiple Sclerosis”. In: *Cells* 9.2 (Feb. 2020), p. 482. ISSN: 2073-4409. DOI: 10.3390/cells9020482.
- [175] Naresha Saligrama et al. “Opposing T cell responses in experimental autoimmune encephalomyelitis”. In: *Nature* 572.7770 (Aug. 2019), pp. 481–487. ISSN: 1476-4687. DOI: 10.1038/s41586-019-1467-x.
- [176] William J. Broad. “The Publishing Game: Getting More for Less: Meet the Least Publishable Unit, one way of squeezing more papers out of a research project”. In: *Science* 211.4487 (Mar. 1981), pp. 1137–1139. ISSN: 1095-9203. DOI: 10.1126/science.7008199.
- [177] Lutz Bornmann and Rüdiger Mutz. “Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references”. In: *Journal of the Association for Information Science and Technology* 66.11 (Apr. 2015), pp. 2215–2222. ISSN: 2330-1643. DOI: 10.1002/asi.23329.
- [178] Harris M. Cooper, Larry V. Hedges, and Jeff C. Valentine. *The handbook of research synthesis and meta-analysis*. Russell Sage Foundation, 2009, p. 615. ISBN: 9780871541635.
- [179] Anoop Kumar. “Introduction”. In: *Meta-analysis in Clinical Research: Principles and Procedures*. Springer Nature Singapore, 2023, pp. 1–6. ISBN: 9789819923700. DOI: 10.1007/978-981-99-2370-0\_1.

- [180] Gordon H. Guyatt. “Users’ Guides to the Medical Literature: IX. A Method for Grading Health Care Recommendations”. In: *JAMA* 274.22 (Dec. 1995), p. 1800. ISSN: 0098-7484. DOI: 10.1001/jama.1995.03530220066035.
- [181] Irene Pérez-Díez et al. “Functional Signatures in Non-Small-Cell Lung Cancer: A Systematic Review and Meta-Analysis of Sex-Based Differences in Transcriptomic Studies”. In: *Cancers* 13.1 (Jan. 2021), p. 143. ISSN: 2072-6694. DOI: 10.3390/cancers13010143.
- [182] Irene Pérez-Díez et al. “A Comprehensive Transcriptional Signature in Pancreatic Ductal Adenocarcinoma Reveals New Insights into the Immune and Desmoplastic Microenvironments”. In: *Cancers* 15.11 (May 2023). ISSN: 2072-6694. DOI: 10.3390/cancers15112887.
- [183] Jaime Llera-Oyola et al. “The role of microRNAs in understanding sex-based differences in Alzheimer’s disease”. In: *Biology of Sex Differences* 15.1 (Jan. 2024). ISSN: 2042-6410. DOI: 10.1186/s13293-024-00588-1.
- [184] Adolfo López-Cerdán et al. “An integrated approach to identifying sex-specific genes, transcription factors, and pathways relevant to Alzheimer’s disease”. In: *Neurobiology of Disease* 199 (Sept. 2024). ISSN: 0969-9961. DOI: 10.1016/j.nbd.2024.106605.
- [185] Hector Carceller et al. “The impact of sex on gene expression in the brain of schizophrenic patients: a systematic review and meta-analysis of transcriptomic studies”. In: *Biology of Sex Differences* 15.1 (July 2024). ISSN: 2042-6410. DOI: 10.1186/s13293-024-00635-x.
- [186] Franc Casanova Ferrer et al. “Unveiling Sex-Based Differences in the Effects of Alcohol Abuse: A Comprehensive Functional Meta-Analysis of Transcriptomic Studies”. In: *Genes* 11.9 (Sept. 2020), p. 1106. ISSN: 2073-4425. DOI: 10.3390/genes11091106.
- [187] Maria Guaita-Cespedes et al. “Deciphering the sex bias in housekeeping gene expression in adipose tissue: a comprehensive meta-analysis of transcriptomic studies”. In: *Biology of Sex Differences* 14.1 (Apr. 2023). ISSN: 2042-6410. DOI: 10.1186/s13293-023-00506-x.

- 
- [188] Nicholas Fabiano et al. “How to optimize the systematic review process using AI tools”. In: *JCPP Advances* 4.2 (Apr. 2024). ISSN: 2692-9384. DOI: 10.1002/jcv2.12234.
- [189] Samantha L. Wilson et al. “Sharing biological data: why, when, and how”. In: *FEBS Letters* 595.7 (Apr. 2021), pp. 847–863. ISSN: 1873-3468. DOI: 10.1002/1873-3468.14067.
- [190] Jeremy Y. Ng et al. “Open science practices in traditional, complementary, and integrative medicine research: A path to enhanced transparency and collaboration”. In: *Integrative Medicine Research* 13.2 (June 2024), p. 101047. ISSN: 2213-4220. DOI: 10.1016/j.imr.2024.101047.
- [191] F. Al-Shahrour et al. “BABELOMICS: a suite of web tools for functional annotation and analysis of groups of genes in high-throughput experiments”. In: *Nucleic Acids Research* 33.Web Server (July 2005), W460–W464. ISSN: 1362-4962. DOI: 10.1093/nar/gki456.
- [192] Tianyuan Liu et al. “PaintOmics 4: new tools for the integrative analysis of multi-omics datasets supported by multiple pathway databases”. In: *Nucleic Acids Research* 50.W1 (May 2022), W551–W559. ISSN: 1362-4962. DOI: 10.1093/nar/gkac352.
- [193] Kinza Rian et al. “Genome-scale mechanistic modeling of signaling pathways made easy: A bioconductor/cytoscape/web server framework for the analysis of omic data”. In: *Computational and Structural Biotechnology Journal* 19 (2021), pp. 2968–2978. ISSN: 2001-0370. DOI: 10.1016/j.csbj.2021.05.022.
- [194] Pablo Malmierca-Merlo et al. “MetaFun: unveiling sex-based differences in multiple transcriptomic studies through comprehensive functional meta-analysis”. In: *Biology of Sex Differences* 15.1 (Aug. 2024). ISSN: 2042-6410. DOI: 10.1186/s13293-024-00640-0.
- [195] European Organization For Nuclear Research and OpenAIRE. *Zenodo*. en. 2013. DOI: 10.25495/7GXX-RD71.
- [196] Francisco García García. “Métodos de análisis de enriquecimiento funcional en estudios genómicos”. es. PhD thesis. 2016. DOI: 10.5281/ZENODO.7669523.

- [197] Yoav Benjamini and Yosef Hochberg. “Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing”. In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 57.1 (Jan. 1995), pp. 289–300. ISSN: 1467-9868. DOI: 10.1111/j.2517-6161.1995.tb02031.x.
- [198] Yoav Benjamini and Daniel Yekutieli. “The control of the false discovery rate in multiple testing under dependency”. In: *The Annals of Statistics* 29.4 (Aug. 2001). ISSN: 0090-5364. DOI: 10.1214/aos/1013699998.
- [199] J. P T Higgins. “Commentary: Heterogeneity in meta-analysis should be expected and appropriately quantified”. In: *International Journal of Epidemiology* 37.5 (Oct. 2008), pp. 1158–1160. ISSN: 1464-3685. DOI: 10.1093/ije/dyn204.
- [200] Wolfgang Viechtbauer. “Conducting Meta-Analyses in R with the metafor Package”. In: *Journal of Statistical Software* 36.3 (2010). ISSN: 1548-7660. DOI: 10.18637/jss.v036.i03.
- [201] Rebecca DerSimonian and Nan Laird. “Meta-analysis in clinical trials”. In: *Controlled Clinical Trials* 7.3 (Sept. 1986), pp. 177–188. ISSN: 0197-2456. DOI: 10.1016/0197-2456(86)90046-2.
- [202] Julian P. T. Higgins and Simon G. Thompson. “Quantifying heterogeneity in a meta-analysis”. In: *Statistics in Medicine* 21.11 (May 2002), pp. 1539–1558. ISSN: 1097-0258. DOI: 10.1002/sim.1186.
- [203] Jonathan A.C Sterne and Matthias Egger. “Funnel plots for detecting bias in meta-analysis”. In: *Journal of Clinical Epidemiology* 54.10 (Oct. 2001), pp. 1046–1055. ISSN: 0895-4356. DOI: 10.1016/s0895-4356(01)00377-8.
- [204] Richard J Light and David B Pillemer. *Summing up: The science of reviewing research*. Harvard University Press, 1984, p. 191. ISBN: 9780674854314.
- [205] David A. Belsley, Edwin Kuh, and Roy E. Welsch. *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. Wiley, June 1980. ISBN: 9780471725152. DOI: 10.1002/0471725153.



- [206] R. Dennis Cook and Sanford Weisberg. “Criticism and Influence Analysis in Regression”. In: *Sociological Methodology* 13 (1982), p. 313. ISSN: 0081-1750. DOI: 10.2307/270724.



# Artificial intelligence statement

The book cover image of the printed version of this doctoral thesis has been generated using the artificial intelligence tool ImageFX (<https://labs.google/fx/tools/image-fx>), developed by Google.

The text of this manuscript has been proofread using Grammarly (<https://www.grammarly.com>), thanks to the educational license provided by the Universitat Politècnica de València.



## Appendix I

# Interpretation of meta-analysis results

As this work is a doctoral thesis composed of various papers, it does not have a dedicated material and methods section. However, providing an explanation of the results from the meta-analyses may be helpful.

As indicated within each chapter, the detailed results of each article are available on a web platform and on the public repository Zenodo:

- Chapter 3: *Hepatic steatosis and steatohepatitis: a functional meta-analysis of sex-based differences in transcriptomic studies.* Results available at <https://bioinfo.cipf.es/metafun-NAFLD/> & <https://zenodo.org/doi/10.5281/zenodo.14062682>
- Chapter 4: *A deep transcriptome meta-analysis reveals sex differences in multiple sclerosis.* Results available at <https://bioinfo.cipf.es/metafun-MS/> & <https://zenodo.org/doi/10.5281/zenodo.5211962>

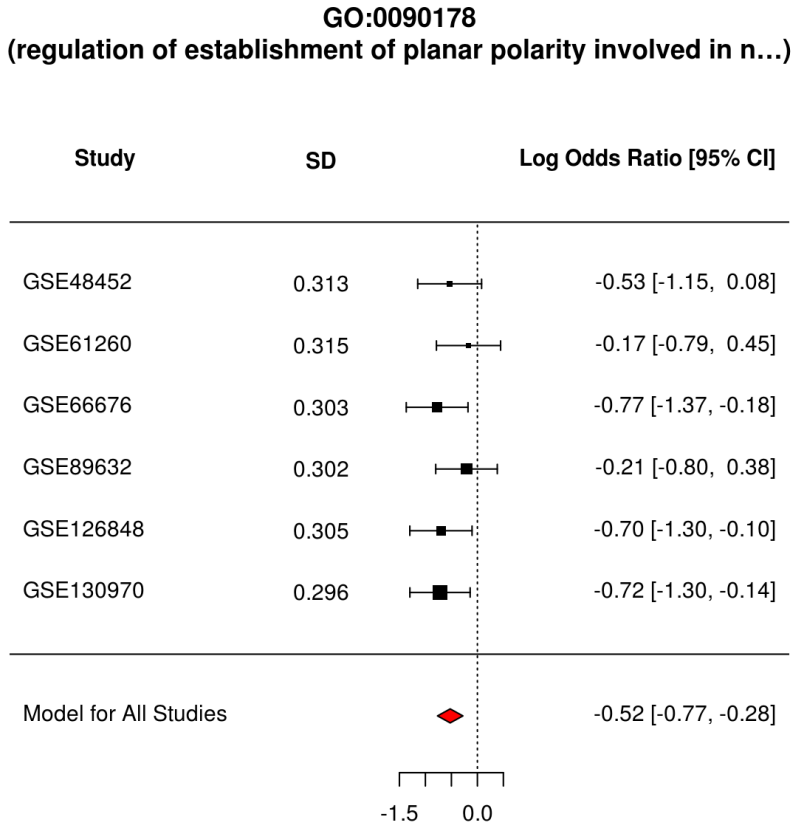
## I.1 Evaluation of the combined effect of the primary studies

For each gene, function, or pathway meta-analysis in this thesis, several statistics are calculated to provide information about the magnitude of the combined change [196]. These results are included both in data tables and visually through forest plots.

- An **estimator of the magnitude of the combined change** (logFC or LOR for genes and functions/pathways, respectively).
- The **95% confidence interval of the estimated combined effect**, which is constructed with the model's estimated variability. The non-inclusion of the value 0 in the interval would confirm the significance of the logFC/LOR.
- The **p-value**: the significance level for a null combined effect. These p-values are adjusted using the Benjamini & Hochberg [197] or Benjamini & Yekutieli [198] method, depending on the analysis.

Forest plots visually represent the results of meta-analyses. They consist of elements that are essential for accurate interpretation. These elements are detailed below and can be identified in Fig. I.1.

- On the left side of the plot, the studies included in the meta-analysis are listed along the y-axis.
- On the right is the value of the summary measure of each study and its confidence interval at 95%.
- In the center, there is a segment and a black square that represent the confidence interval and the effect measure value for each individual study. The size of the black square indicates the precision of the estimates.
- At the bottom of the plot, the combined effect of the meta-analysis is displayed, along with its 95% confidence interval. This is represented by a diamond shape that is centered on the effect value, and its width corresponds to the confidence interval for the effect.



**Fig. I.1:** Example of a forest plot showing the results of the meta-analysis of the GO:0090178 process for 6 primary studies. The image shows each study's effect measure (Log Odds Ratio), 95% confidence interval, and standard error (SD). Also, the combined effect measure and its 95% confidence interval resulted from the individual studies' meta-analysis.

## I.2 Evaluation of the heterogeneity and the influence of the primary studies

### *Evaluation of the heterogeneity of the primary studies*

The differences among the primary studies refer to the variability in the results obtained from each study. Variability in meta-analyses is unavoidable

[199], and various methods have been developed to evaluate and represent this variability. The methods used in this work are described below [141, 196, 200]:

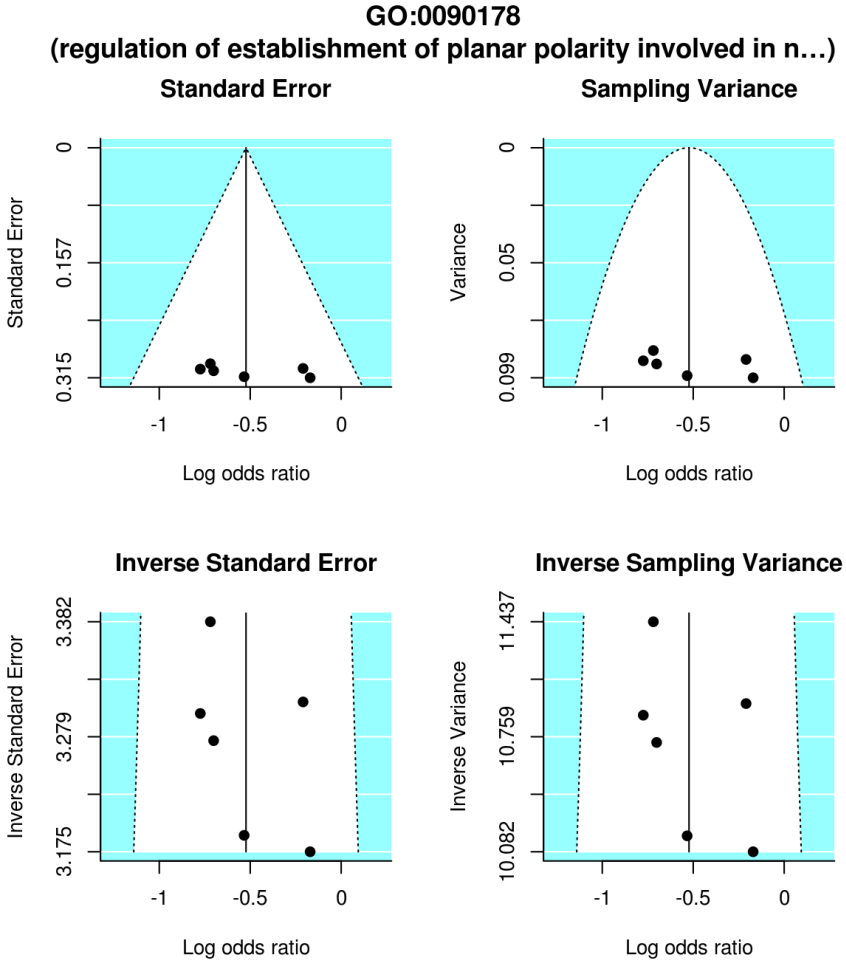
- $QE$  &  $QEp$  represent the contrast statistic and p-value, respectively, of the DerSimonian and Laird [201] test for detecting between-study heterogeneity. The null hypothesis of this test points to the absence of heterogeneity between studies.
- $\tau^2$ . Estimate of residual amount of heterogeneity.
- $I^2$ . Estimation, as a percentage, of the variability between studies concerning the overall variability, i.e., variability not due to chance. The 25%, 50%, and 75% values are usually used to delimit low, moderate, and high heterogeneity, respectively.
- $H^2$ . The ratio of global variability to sampling variability [202].

The DerSimonian and Laird test of heterogeneity between studies [201] has low statistical power when there are few primary studies, so it is beneficial to utilize graphical representations of heterogeneity. For this purpose, funnel plots have been used in this doctoral thesis. Funnel plots can be used to illustrate the variability of different studies and the presence of biases [203]. These representations show the magnitude of the effect on the X-axis versus a measure of precision (variance, standard error, etc.) on the Y-axis. The lines delimiting the funnel area represent the 95% confidence interval limits. In the absence of bias, a symmetrical graph is expected. Fig. I.2 shows an example funnel plot of the GO:0090178 process from Chapter 3.

### ***Evaluation of the influence of the primary studies***

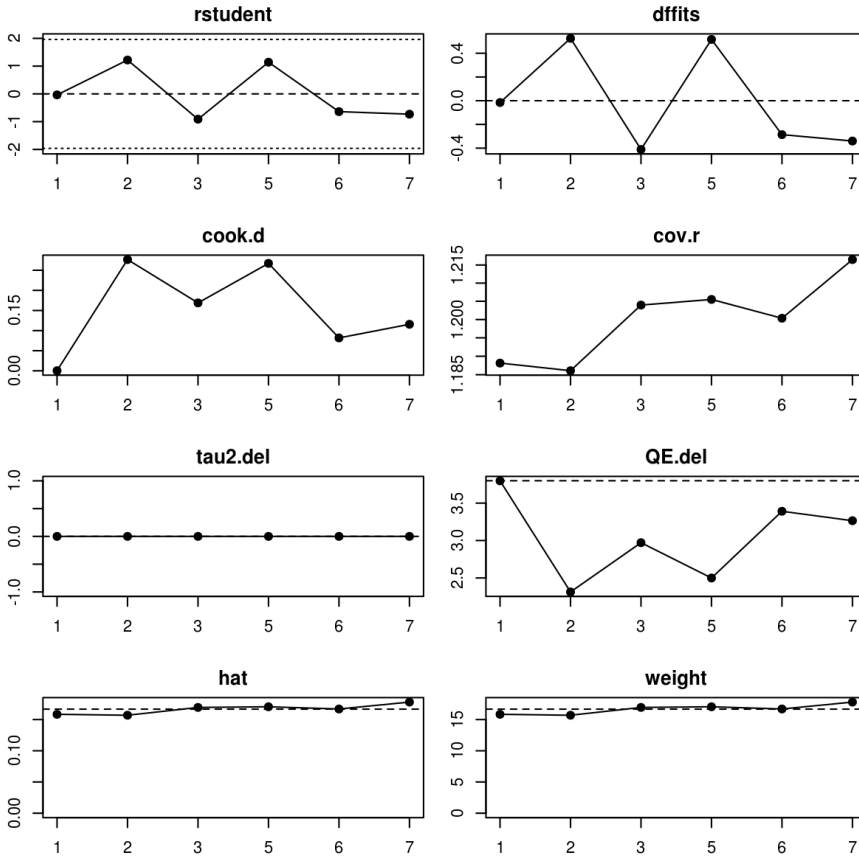
It is important to note that the presence of primary studies with significantly different magnitudes or variabilities compared to others can greatly affect the results of a meta-analysis [196]. Removing these studies would considerably reduce the heterogeneity and improve the precision of the average outcome estimated by the random-effects model. Nevertheless, rather than simply excluding these studies, it is essential to examine them to understand the reasons behind their unusual results. Outliers and influential cases may provide valuable insights into study characteristics that could be acting as potential moderators, revealing new patterns and leading to new insights [204]. The assessment of data influence analysis is already implemented in regression models [205, 206] and has been adapted to meta-analyses to detect studies with a superior influence on the model.





**Fig. I.2:** Example of a funnel plots showing the results of the meta-analysis of GO:0090178 process for 6 primary studies. The X-axis shows the variability estimators (standard error, sampling variance, inverse standard error, and inverse sampling variance), while the Y-axis shows the effect of magnitude as LOR. It can be seen how all four studies are within the confidence region defined by the white area.

Figure I.3 shows an example of the graphical representations of the influence diagnostics generated by the *metafor* R package [200], in which various statistics are evaluated for each primary study.



**Fig. I.3:** Representation of the influence diagnostics plots, performed by the “influence” function of the *metafor* R package [200], for the GO:0090178 process analyzing 6 primary studies. The figure shows a plot of the (1) externally standardized residuals, (2) DFFITS values, (3) Cook’s distances, (4) covariance ratios, (5) leave-one-out estimates of the amount of heterogeneity, (6) leave-one-out values of the test statistics for heterogeneity, (7) hat values, and (8) weights.

## Appendix II

# Scientific contributions during the period of this doctoral dissertation

Since I began this doctoral thesis in November 2019, I have contributed to various scientific projects, both directly related to this thesis and through collaborations with colleagues from the Bioinformatics and Biostatistics Unit (UBB/CBL) of the Prince Felipe Research Center as well as other research groups. Additionally, during this time, I have co-directed several final degree and master's theses. The details of my contributions are listed below:

### Journal papers

1. Irene Soler-Sáez; Alcida Karz; Marta R. Hidalgo; Borja Gómez-Cabañes; Adolfo López-Cerdán; **José F. Català-Senent**; Kylie Prutisto-Chang; Nicole M. Eskow; Benjamin Izar; Torben Redmer; Swaminathan Kumar; Michael A. Davies; María de la Iglesia-Vayá; Eva Hernando; Francisco García-García. *Unveiling common transcriptomic features between melanoma brain metastases and neurodegenerative diseases*. Journal of Investigative Dermatology. 24/09/2024. DOI: 10.1016/j.jid.2024.09.005

2. Pablo Malmierca-Merlo; Rubén Sánchez-García; Rubén Grillo-Risco; Irene Pérez-Díez; **José F. Català-Senent**; María de la Iglesia-Vayá; Marta R. Hidalgo; Francisco García-García. *MetaFun: unveiling sex-based differences in multiple transcriptomic studies through comprehensive functional meta-analysis*. Biology of Sex Differences. 27/08/2024. DOI: 10.1186/s13293-024-00640-0
3. Elena G. Biosca; Ricardo Delgado Santander; Félix Morán; Àngela Figàs-Segura; Rosa Vázquez; **José F. Català-Senent**; Belén Álvarez. *First european Erwinia amylovora lytic bacteriophage cocktails effective in the host: Characterization and prospects for fire blight biocontrol*. Biology. 08/03/2024. DOI: 10.3390/biology13030176
4. Isabel Fuster-Martínez; **José F. Català-Senent**; Marta R. Hidalgo; Francisco J. Roig; Juan V. Esplugues; Nadezda Apostolova; Francisco García-García; Ana Blas-García. *Integrated transcriptomic landscape of the effect of anti-steatotic treatments in high-fat diet mouse models of non-alcoholic fatty liver disease*. The Journal of Pathology. 05/02/2024. DOI: 10.1002/path.6242
5. Carla Perpiñá-Clérigues; Susana Mellado; **José F. Català-Senent**; Francesc Ibáñez; Pilar Costa; Miguel Marcos; Consuelo Guerri; Francisco García-García; Maria Pascual. *Lipidomic landscape of circulating extracellular vesicles isolated from adolescents exposed to ethanol intoxication: A sex difference study*. Biology of Sex Differences. 21/04/2023. DOI: 10.1186/s13293-023-00502-1
6. **José F. Català-Senent**; Zoraida Andreu; Marta R. Hidalgo; Irene Soler-Saez; Francisco J. Roig; Natalia Yanguas-Casas; Almudena Neva-Alejo; Adolfo López-Cerdán; María de la Iglesia-Vayá; Barbara E. Stranger; Francisco García-García. *A deep transcriptome meta-analysis reveals sex differences in multiple sclerosis*. Neurobiology of Disease. 05/04/2023. Chapter 4 of this thesis and DOI: 10.1016/j.nbd.2023.106113
7. Adolfo López-Cerdán; Zoraida Andreu; Marta R. Hidalgo; Rubén Grillo-Risco; **José F. Català-Senent**; Irene Soler-Sáez; Almudena Neva-Alejo; Fernando Gordillo; María de la Iglesia-Vayá; Francisco García-García. *Unveiling sex-based differences in Parkinson's disease: a comprehensive meta-analysis of transcriptomic studies*. Biology of Sex Differences. 22/11/2022. DOI: 10.1186/s13293-022-00477-5
8. Elena G. Biosca; **José F. Català-Senent**; Àngela Figàs-Segura; Edson Bertolini; María M. López; Belén Álvarez. *Genomic analysis of the first*

*european bacteriophages with depolymerase activity and biocontrol efficacy against the phytopathogen Ralstonia solanacearum*. 17/12/2021. DOI: 10.3390/v13122539

9. **José F. Català-Senent**; Marta R. Hidalgo; Marina Berenguer; Gopanandan Parthasarathy; Harmeet Malhi; Pablo Malmierca-Merlo; María de la Iglesia-Vayá; Francisco Garcia-Garcia. *Hepatic steatosis and steatohepatitis: A functional meta-analysis of sex-based differences in transcriptomic studies. Biology of Sex Differences*. 25/03/2021. Chapter 3 of this thesis and DOI: 10.1186/s13293-021-00368-1
10. Ricardo Delgado Santander; **José F. Català Senent**; Àngela Figàs Segura; Elena González Biosca. *From the roots to the stem: unveiling pear root colonization and infection pathways by Erwinia amylovora*. FEMS Microbiology Ecology. 08/10/2020. DOI: 10.1093/femsec/fiaa210

## National and international conferences

1. *Characterizing sex differences in neurodegenerative diseases through novel and integrative computational strategies: difneromics project*. 1er Congreso de la Sociedad Española de Bioinformática y Biología Computacional. 16/10/2024 - 18/10/2024. Francisco García García; Marta R. Hidalgo; Irene Soler-Sáez; Fernando Gordillo; Borja Gómez-Cabañes; Rubén Grillo-Risco; Carla Perpiñá-Clérigues; Cristina Galiana-Roselló; Edmond Geraud-Aguilar; Zoraida Andreu; Jaime Llera-Oyola; Beatriz Rosón; Beatriz Dolader-Rabinad; Natalia del Rey Díez; **José Francisco Català-Senent**; Adolfo López-Cerdán; Franca R. Guerini; Eva Hernando; Susana Barceló Cerdá; Vanja Tepavcevic; María de la Iglesia-Vayá; Amparo Galán; María Pascual; Amparo Oliver; Akiko Mizokami; Sara Gil-Perotin; Lucas Barea Moya; Jesus Alejandro Alzate-Grisales; Alejandro Mora-Rubio; Deborah J Burks.
2. *UMIvar: UMI-aware simulation of variants in alignment data for pipeline benchmarking*. 1er Congreso de la Sociedad Española de Bioinformática y Biología Computacional. 16/10/2024 - 18/10/2024. Daniel G. Camblor; Jorge Martín-Arana; Francisco Gimeno-Valiente; Blanca García-Micó; Belén Martínez-Castedo; **José F. Català-Senent**; Fabián Robledo-Yagüe; Francisco Martínez-Picó; Andrés Cervantes; Noelia Tarazona.
3. *Unveiling shared transcriptomic profiles of melanoma brain metastases and the neurodegenerative disorders alzheimer's, parkinson's and multiple*

- sclerosis*. 1er Congreso de la Sociedad Española de Bioinformática y Biología Computacional. 16/10/2024 - 18/10/2024. Irene Soler Sáez; Alcida Karz; Marta R. Hidalgo; Borja Gómez-Cabañes; Adolfo López-Cerdán; **José F. Català-Senent**; María de la Iglesia-Vayá; Eva Hernando; Francisco García-García.
4. *Unveiling shared transcriptomic profiles of melanoma brain metastases and the neurodegenerative disorders alzheimer's, parkinson's and multiple sclerosis*. IX Simposio nacional de estudiantes de Bioinformática. 16/10/2024. Irene Soler Sáez; Alcida Karz; Marta R. Hidalgo; Borja Gómez-Cabañes; Adolfo López-Cerdán; **José F. Català-Senent**; María de la Iglesia-Vayá; Eva Hernando; Francisco García-García.
  5. *Aislamiento y caracterización de bacteriófagos líticos frente a cepas clínicas de Escherichia coli*. VIII Reunión Fagoma. 02/10/2024 - 04/10/2024. Rosa Vázquez; **José F. Català-Senent**; Juan Frasset-Artes; Paula Ramírez-Galleymore; Elena G. Biosca.
  6. *Bacteriófagos líticos aislados del medio ambiente para el biocontrol de Xanthomonas arboricola pv. pruni*. VIII Reunión Fagoma. 02/10/2024 - 04/10/2024. Belén Álvarez; Félix Morán; Nerea García; Rosa Vázquez; Isabel Salas; **José F. Català-Senent**; Ana Palacio-Bielsa; Sergi Maicas; Elena G. Biosca.
  7. *Enseñanza activa y búsqueda colaborativa de bacteriófagos frente a las superbacterias (FAGO@VAL.2)*. VIII Reunión Fagoma. 02/10/2024 - 04/10/2024. Elena G. Biosca; Rosa Vázquez; Belén Fouz; Belén Fouz; Hortensia Rico; Jesús Zueco; Ana Pérez-Solsona; Belén Álvarez; Félix Morán; Isabel Salas; **José F. Català-Senent**; Juan Frasset-Artes; Luis Martínez-Dolz; José Luis Díez; Paula Ramírez-Galleymore.
  8. *COMPReS study: Multiomic profiling reveals organ-specific differences in metastases and identifies novel predictive biomarkers in relapsed localized colon cancer*. ESMO Congress 2024. 13/09/2024 - 17/09/2024. B. García-Mico; F. Gimeno-Valiente; J. Martín-Arana; B. Martínez-Castedo; D.G. Camblor; M. García-Bartolomé; B. Palomar; C. Martínez-Ciarpaglini; M. Cabeza Segura; V. Gambardella; **J.F. Català-Senent**; D. Moro; S. Garcia-Botello; L. Pérez; S. Rosello Keranen; T.C. Fleitas; M. Huerta; D. Roda Perez; A. Cervantes; N. Tarazona Llaverro.
  9. *Deciphering the molecular drivers behind locoregional progression, intratumoral heterogeneity, and clonal evolution in locally advanced head and neck cancer*. ESMO Congress 2024. 13/09/2024 - 17/09/2024. G.

- Bruixola; J. Martín-Arana; F. Gimeno-Valiente; **J.F. Català-Senent**; J.A. Carbonell-Asins; C. Alfaro-Cervello; D.G. Cambor; B. Martínez-Castedo; B. García-Mico; N. Grimalt Ferrer; V. Seguí-Manzanares; V. Escorihuela-García; J.L. León-Guijarro; M.E. Iglesias; D. Dualde; O. Pons; A. Cervantes; N. Tarazona Llaveró.
10. *UMIvar: Unveiling Insights in UMI Deduplication Software Benchmarking through UMI-aware Variant Simulation*. EACR 2024. 10/06/2024 - 13/06/2024. D. G. Cambor; J. Martín-Arana; F. Gimeno-Valiente; B. García-Micó; B. Martínez-Castedo; **J.F. Català-Senent**; F. Robledo-Yagüe; S. Zúñiga-Trejos; A. Cervantes; N. Tarazona.
  11. *ctDNA-based WES enhances actionable alterations detection in locally advanced head and neck* *ctDNA-based WES enhances actionable alterations detection in locally advanced head and neck cancer*. 9th International Congress on Innovative Approaches in Head and Neck Oncology. 21/03/2024 - 23/03/2024. Gema Bruixola; Jorge Martín-Arana; Francisco Gimeno-Valiente; Víctor Seguí; **Jose Francisco Català-Senent**; Juan Antonio Carbonell-Asins; Elena Duréndez; Nuria Grimalt; Delfina Dualde; Clara Alfaro; Olga Pons; Vicente Escorihuela-García; María Eugenia Iglesias; Andrés Cervantes; Noelia Tarazona.
  12. *Peptidylprolyl isomerase C in chronic liver disease: pathogenic or protective mechanism?*. 2nd International Workshop on Liver and Gut Fibrosis. 26/11/2023 - 27-11-2023. Isabel Fuster-Martínez; Ana M. Benedicto; **Jose F. Català-Senent**; Marta R. Hidalgo; Francisco García-García; Juan V. Esplugues; Nadezda Apostolova; Ana Blas-García.
  13. *Obtención y caracterización de bacteriófagos de Erwinia amylovora de ambientes mediterráneos con potencial de biocontrol*. XX Congreso de la Sociedad Española de Fitopatología. 22/10/2022 - 26/10/2022. À. Figàs-Segura; **J. F. Català-Senent**; P. Moreno; M. Orero; J.J. González-López; E. G. Biosca.
  14. *Ethanol-induced sex-based differences in the extracellular vesicles lipidome*. 21st European Conference on Computational Biology. 18/09/2022 - 21/09/2022. Carla Perpiñá-Clérigues; **José F. Català-Senent**; Susana Mellado; Consuelo Guerri; María Pascual; Francisco García-García.
  15. *Molecular and functional atlas of sex-differences in multiple sclerosis subtypes analysing single cell and single nucleus transcriptomic data*. 21st European Conference on Computational Biology. 18/09/2022 - 21/09/2022.

Irene Soler-Sáez; Zoraida Andreu; **José Francisco Català-Senent**; Rubén Grillo-Risco; Adolfo López-Cerdán; Almudena Neva-Alejo; Borja Gómez; Marta R. Hidalgo; Francisco García-García.

16. *Sex differences in the molecular basis of multiple sclerosis: meta-analysis of transcriptomic data*. 21st European Conference on Computational Biology. 18/09/2022 - 21/09/2022. **José Francisco Català-Senent**; Zoraida Andreu; Marta R. Hidalgo; Francisco José Roig; Natalia Yanguas-Casás; Almudena Neva-Alejo; Adolfo López-Cerdán; Irene Soler-Sáez; María de la Iglesia-Vayá; Francisco García-García.
17. *Isolation and characterization of Erwinia amylovora bacteriophages from Mediterranean environments with biocontrol potential*. Third International Symposium on Fire Blight of Rosaceous Plants. 06/09/2022 - 09/09/2022. À. Figàs-Segura; **José Francisco Català Senent**; P. Moreno-Zaragozà; J.J. González-López; R. D.Santander; E. G. Biosca.
18. *Atlas funcional de las diferencias de sexo en esclerosis múltiple mediante el análisis de datos transcriptómicos de célula única*. I Simposio de Estudiantes Hispanohablantes de Bioinformática y Biología Computacional. 02/06/2022 - 03/06/2022. Irene Soler-Sáez; Zoraida Andreu; **José Francisco Català-Senent**; Rubén Grillo-Risco; Adolfo López-Cerdán; Almudena Neva-Alejo; Héctor Carceller; María de la Iglesia-Vayá; Marta R. Hidalgo; Francisco García-García.
19. *Diferencias de sexo en el lipidoma de las vesículas extracelulares tras el consumo masivo de alcohol en la adolescencia*. I Simposio de Estudiantes Hispanohablantes de Bioinformática y Biología Computacional. 02/06/2022 - 03/06/2022. Carla Perpiñá-Clérigues; **José Francisco Català-Senent**; Francesc Ibáñez; Consuelo Guerri; María Pascual-Mora; Francisco García-García.
20. *Diferencias de sexo en la Enfermedad de Parkinson: metaanálisis de estudios transcriptómicos de tejido cerebral*. I Simposio de Estudiantes Hispanohablantes de Bioinformática y Biología Computacional. 02/06/2022 - 03/06/2022. Adolfo López-Cerdán; Zoraida Andreu; Marta R. Hidalgo; Rubén Grillo-Risco; **José Francisco Català-Senent**; Irene Soler-Sáez; Almudena Neva-Alejo; María de la Iglesia-Vayá; Francisco García-García.
21. *Estudio de las bases moleculares afectadas por el sexo en esclerosis múltiple a través del transcriptoma de tejido nervioso*. I Simposio de Estudiantes Hispanohablantes de Bioinformática y Biología Computacional. 02/06/2022 - 03/06/2022. **José Francisco Català-Senent**; Zoraida



- Andreu; Marta R. Hidalgo; Francisco José Roig; Natalia Yanguas-Casás; Almudena Neva-Alejo; Adolfo López-Cerdán; Irene Soler-Sáez; María de la Iglesia-Vayá; Francisco García-García.
22. *Reposicionamiento in silico de fármacos en la enfermedad de Alzheimer con perspectiva de sexo*. I Simposio de Estudiantes Hispanohablantes de Bioinformática y Biología Computacional. 02/06/2022 - 03/06/2022. Cristina Galiana-Roselló; Rubén Grillo-Risco; **José Francisco Català-Senent**; Adolfo López-Cerdán; Isabel del Pino; Francisco García-García.
  23. *Innovando en el control biológico de la bacteria fitopatógena *Ralstonia solanacearum* en agua de riego y en planta mediante el uso de bacteriófagos*. VI Reunión Fagoma. 20/10/2021 - 22/10/2021. Belén Álvarez; Àngela Figàs-Segura; **José F. Català-Senent**; María M. López; Elena G. Biosca.
  24. *A single-cell and single-nuclei transcriptomics approach reveals novel sex differences in multiple sclerosis*. VIII Bioinformatics Students Symposium. 18/10/2021 - 19/10/2021. Irene Soler-Sáez; Zoraida Andreu; Almudena Neva-Alejo; **José F. Català Senent**; Adolfo López Cerdán; María de la Iglesia-Vayá; Francisco García-García.
  25. *An integrated single-nucleus transcriptomics approach to identify sex molecular differences in Alzheimer's disease*. VIII Bioinformatics Students Symposium. 18/10/2021 - 19/10/2021. Almudena Neva-Alejo; Zoraida Andreu; Irene Soler-Sáez; **José F. Català Senent**; Adolfo López Cerdán; María de la Iglesia-Vayá; Francisco García-García
  26. *Lipidomic profile study of extracellular vesicles to identify differences in adolescents exposed to alcohol intoxication with sex perspective*. VIII Bioinformatics Students Symposium. 18/10/2021 - 19/10/2021. Carla Perpiñá-Clérigues; **José F. Català-Senent**; Francesc Ibáñez; Consuelo Guerri; María Pascual-Mora; Francisco García-García.
  27. *Sex differences in the lipidomic profile of extracellular vesicles of adolescents exposed to alcohol intoxication*. ESBRA 18th Conference. 07/10/2021 - 09/10/2021. María Pascual; Consuelo Guerri; Francesc Ibáñez; Carla Perpiñá-Clérigues; **José F. Català-Senent**; Francisco García-García.
  28. *Sex differences in Alzheimer's disease from single cell transcriptomics data analysis*. 2nd Women in Bioinformatics & Data Science LA Conference. 22/09/2021 - 24/09/2021. Almudena Neva-Alejo; Zoraida Andreu;

Irene Soler-Sáez; **José F. Català Senent**; Adolfo López Cerdán; María de la Iglesia-Vayá; Francisco García García.

29. *Unraveling sex differences in multiple sclerosis: an integrated single-cell transcriptomics approach*. 2nd Women in Bioinformatics & Data Science LA Conference. 22/09/2021 - 24/09/2021. Irene Soler-Sáez; Zoraida Andreu; Almudena Neva-Alejo; **José F. Català Senent**; Adolfo López Cerdán; María de la Iglesia-Vayá; Francisco García García.
30. *Transcriptomic signature of high fat diet murine models of NAFLD and its modulation by anti-steatotic treatments: a meta-analysis*. Digital NAFLD summit 2021. 16/09/2021 - 17/09/2021. Isabel Fuster-Martínez; **José F. Català-Senent**; Francisco J. Roig; Juan V. Esplugues; Nadezda Apostolova; Francisco García-García; Ana Blas-García.
31. *Análisis genómico y filogenético de bacteriófagos con actividad lítica y potencial de control biológico de la bacteria fitopatogena Ralstonia solanacearum*. XXVIII Congreso Nacional de Microbiología. 28/06/2021 - 02/07/2021. Belén Álvarez; Àngela Figàs-Segura; **José Francisco Català-Senent**; María M. López; Elena G. Biosca.
32. *Integrative study of the transcriptome signature of high fat diet murine models of MAFLD and its modulation by anti-steatotic treatments using a systematic review and meta-analysis*. 46º Congreso anual de la Asociación Española para el Estudio del Hígado. 14/06/2021 - 16/06/2021. Isabel Fuster-Martínez; **José F. Català-Senent**; Francisco J. Roig; Juan V. Esplugues; Nadezda Apostolova; Francisco García-García; Ana Blas-García.
33. *Functional signatures in Alzheimer disease: systematic review and meta-analysis of sex-based differences in transcriptomics studies*. OSSD 2021 Virtual annual meeting. 03/05/2021 - 06/05/2021. Adolfo López Cerdán; Zoraida Andreu; Irene Soler Sáez; Almudena Neva Alejo; **José Francisco Català Senent**; María de la Iglesia Vayá; Francisco García García.
34. *MetaFun: A web-based tool for comprehensive statistical meta-analysis of sex differences in omics studies*. OSSD 2021 Virtual annual meeting. 03/05/2021 - 06/05/2021. Pablo Malmierca Merlo; Rubén Sánchez García; Rubén Grillo Risco; Irene Pérez Díez; **José Francisco Català Senent**; M. Pérez Pérez; E. Prieto; Marta R. Hidalgo; María de la Iglesia Vayá; Francisco García García.

35. *Sex-based Differences in Multiple Sclerosis: A Systematic Review and Meta-Analysis of transcriptomic studies*. OSSD 2021 Virtual annual meeting. 03/05/2021 - 06/05/2021. **José Francisco Català Senent**; Zoraida Adreu; Almudena Neva Alejo; Irene Soler Sáez; Francisco L. Roig; Natalia Yanguas; Adolfo López Cerdán; María de la Iglesia Vayá; Francisco García García.
36. *Unveiling sex-based differences in Parkinson disease: a comprehensive functional meta-analysis of transcriptomic studies*. OSSD 2021 Virtual annual meeting. 03/05/2021 - 06/05/2021. Adolfo López Cerdán; Zoraida Adreu; Irene Soler Sáez; Almudena Neva Alejo; **José Francisco Català Senent**; María de la Iglesia Vayá; Francisco García García.
37. *Systematic review and gene expression meta-analysis of sex differences in multiple sclerosis*. IV Congreso Nacional Jóvenes Investigadores en Biomedicina. 04/11/2020 - 06/11/2020. **José F Català-Senent**; Zoraida Andreu; Almudena Neva Alejo; Irene Soler Sáez; Francisco J Roig; Natalia Yanguas; Francisco García-García.
38. *MetaFun: a web tool to unveil sex differences in multiple omics studies through comprehensive functional meta-analysis*. 6th European Student Council Symposium. 06/09/2020. Pablo Malmierca Merlo; Marta R. Hidalgo García; Irene Pérez Díez; Rubén Grillo Risco; **José Francisco Català Senent**; Francisco García García; Rubén Sánchez García.
39. *Sex differences in the progression from NAFL to NASH: a functional meta-analysis of transcriptomic studies*. 6th European Student Council Symposium. 06/09/2020. **José Francisco Català Senent**; Marta R. Hidalgo García; Marina Berenguer Haym; Harmeet Malhi; Pablo Malmierca Merlo; Francisco García García.
40. *Understanding sex differences in neurodegenerative diseases: A systematic review and meta-analysis of transcriptomic studies*. 6th European Student Council Symposium. 06/09/2020. Miriam Ferreiro Pantín; Violeta Lillo Collado; Alejandra Morcillo Nieto; Adolfo López Cerdán; **José Francisco Català Senent**; Maria de la Iglesia Vaya; Francisco García García.
41. *WODA: Web-based Omics Data Analysis course*. 6th European Student Council Symposium. 06/09/2020. Sergio Romera Giner; Marta R. Hidalgo García; **José Francisco Català Senent**; Sandra Alandes Esteve; Rubén Grillo Risco; Rubén Sánchez García; Irene Pérez Díez; Raúl Pérez

Moraga; Adolfo López Cerdán; Maria de la Iglesia Vaya; Francisco García García; Pablo Malmierca Merlo.

## Experience supervising master's thesis and/or final year projects

- 1. Project title:** Caracterización de las diferencias de sexo en esclerosis múltiple, mediante la revisión sistemática y metaanálisis de estudios transcriptómicos a nivel de microRNAs.  
**Type of project:** End of course project.  
**Entity:** Universidad Politécnica de Valencia.  
**Student:** Ester Muñoz del Campo  
**Co-director of project:** Francisco García-García & José Francisco Català Senent  
**Obtained qualification:** 10  
**Date of reading:** 27/06/2022
- 2. Project title:** Caracterización morfológica y molecular de bacteriófagos de *Erwinia amylovora*.  
**Type of project:** End of course project.  
**Entity:** Universitat de València.  
**Student:** Pilar Moreno Zaragoza  
**Co-director of project:** Elena González Biosca, Àngela Figàs Segura & José Francisco Català Senent  
**Obtained qualification:** 9.5  
**Date of reading:** 14/09/2021
- 3. Project title:** Caracterización de las diferencias de sexo en cáncer de páncreas ductal mediante el análisis masivo de datos transcriptómicos.  
**Type of project:** Master's thesis.  
**Entity:** Universitat de València.  
**Student:** Lucía Fantín  
**Co-director of project:** Francisco García-García, Rubén Grillo Risco & José Francisco Català Senent  
**Obtained qualification:** 8.9  
**Date of reading:** 20/07/2021