

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

ESCOLA TÈCNICA SUPERIOR D'ENGINYERIA
AGRONÒMICA I DEL MEDI NATURAL (ETSIAMN)



Perfil consenso transcripcional de las diferencias de sexo en los accidentes cerebrovasculares

Grado en Biotecnología

Trabajo Final de Grado (TFG)

Autor: Irene Cánovas Cervera

Tutora UPV: Susana Barceló Cerdá

Tutores CIPF: Francisco García García, Pablo Malmierca Merlo

Curso Académico: 2020/2021

Valencia, Junio de 2021



Esta obra está bajo una licencia de Creative Commons Reconocimiento-NoComercial-CompartirIgual 4.0 Internacional

Título: Perfil consenso transcripcional de las diferencias de sexo en los accidentes cerebrovasculares

Resumen

El accidente cerebrovascular se define como la patología caracterizada por la interrupción o disminución del flujo sanguíneo cerebral, desencadenando con ello la muerte celular del tejido. Por tanto, su efecto es doble: como enfermedad cardiovascular y como enfermedad neurológica. El ictus es una de las enfermedades con mayor impacto en los pacientes, tanto por su alta mortalidad, como por el nivel de discapacidad producida tras sufrir este evento.

Sin embargo, existe una diferencia notable entre sexos. Aunque es la segunda causa de muerte a nivel nacional, es la primera en mujeres. Esta variación implica una diferencia molecular ligada al sexo cuya adecuada caracterización podría favorecer el correcto diagnóstico y tratamiento del ictus. No obstante, la mayoría de estudios relacionados con el accidente cerebrovascular no considera el sexo como factor importante y, por tanto, no es anotado para cada muestra.

En este trabajo se ha aplicado una estrategia bioinformática en la cual se estudiaron las diferencias debidas tanto a la naturaleza del ictus como al sexo. Para ello, primero se ha llevado a cabo una revisión sistemática en bases de datos donde se obtuvieron todos los datos de estudios relacionados con esta patología. En aquellos donde no se describe la información del sexo en las muestras, se realizó su imputación a partir de los datos de expresión génica. A continuación, para cada uno de los estudios seleccionados, se ha realizado tanto un análisis exploratorio como un análisis de expresión diferencial. Por último, se ha desarrollado un metaanálisis de genes con el fin de averiguar si el sexo influye como factor determinante en el ictus. Todos los resultados han sido integrados para identificar un perfil consenso transcriptómico de las diferencias de sexo en los accidentes cerebrovasculares.

El abordaje descrito identificó 7 biomarcadores específicos de esta patología por sexo: 5 genes expresados diferencialmente en mujeres, y 2 en hombres. Un análisis posterior de los 25 genes más expresados para cada sexo ha revelado tener implicaciones con genes *anti-stroke*. Estos genes, que previenen la aparición del ictus, estaban más expresados en hombres que en mujeres. Además, los hombres mostraron una regulación positiva de la respuesta neuroinflamatoria. Esto sugeriría una menor protección de las mujeres ante este tipo de accidentes. Sin embargo, se necesita realizar estudios más exhaustivos, y que además incluyan el sexo como factor necesario en su diseño y en las estrategias de análisis de los datos de estos estudios.

Palabras clave: ictus; accidente cerebrovascular; CVA; metaanálisis; diferencias de sexo

Autor: Irene Cánovas Cervera

Tutora UPV: Susana Barceló Cerdá

Tutores CIPF: Francisco García García, Pablo Malmierca Merlo

Valencia, Junio de 2021



Title: Consensus transcriptional profiling of sex differences in cerebrovascular accidents

Abstract

The cerebrovascular accident is defined as a pathology characterized by the interruption or decrease of cerebral blood flow, thus triggering tissue cell death. Therefore, its effect is twofold: as a cardiovascular disease and as a neurological disease. Stroke is one of the diseases with the most impact on patients, both due to its high mortality and the level of disability produced after suffering this event.

However, there is a notable difference between the sexes. Although it is the second leading cause of death nationally, it is the first in women. This variation implies a molecular difference linked to sex whose adequate characterization could favor the correct diagnosis and treatment of stroke. Nevertheless, most studies related to the cerebrovascular accident do not consider sex as an important factor and, therefore, it is not recorded for each sample.

In this project, a bioinformatics strategy has been applied in which differences due to both the nature of the stroke and gender were studied. To do this, a systematic review was first carried out in databases where all the data from studies related to this pathology were obtained. In those where the sex information is not described in the samples, its imputation was made from the gene expression data. Then, for each of the selected studies, both an exploratory analysis and a differential expression analysis were performed. Finally, a gene meta-analysis has been developed to find out if sex influences stroke as a determining factor. All results have been integrated to identify a transcriptomic consensus profile of sex differences in strokes.

The described approach identified 7 specific biomarkers of this pathology by sex: 5 genes expressed differentially in women and 2 in men. A subsequent analysis of the 25 most expressed genes for each sex has revealed to have implications with anti-stroke genes. These genes, which prevent the onset of stroke, were more expressed in men than in women. In addition, the men showed a positive regulation of the neuroinflammatory response. This would suggest less protection for women against these types of accidents. However, more exhaustive studies are needed, and that also includes sex as a necessary factor in the design and the data analysis strategies of these studies.

Key words: stroke; cerebrovascular accident; CVA; meta-analysis; sex differences

Author: Irene Cánovas Cervera

UPV tutor: Susana Barceló Cerdá

CIPF tutors: Francisco García García, Pablo Malmierca Merlo

Valencia, June 2021



Agradecimientos

Existen ciertos momentos en la vida donde parece que cambias de rumbo de repente. Estos instantes no suelen ser muy importantes en el momento que ocurren, pero conforme pasa el tiempo te das cuenta de las implicaciones que las pequeñas decisiones tienen sobre tu vida.

Uno de los momentos más importantes que han resultado de este trabajo final de grado ocurrió al final de una práctica de estadística. En ese momento, mi tutora Susana me ayudó a ver lo importante de la carrera: las notas importan, pero solo hasta el momento donde te surgen oportunidades. Justo en ese instante, hay que aprovecharlas y disfrutar de ellas con total libertad. Susana me animó a realizar mi ERASMUS en París, y también me animó a realizar este trabajo en el CIPF. Me ha ofrecido un sinfín de oportunidades además de apoyo en todo lo que me he propuesto. Estas líneas no pueden describir lo agradecida que estoy.

Sin embargo, ha habido otro momento esencial en el desarrollo de este trabajo: mis prácticas durante el verano en el CIPF, junto con Paco y todos los componentes del laboratorio. Admito que al principio la bioinformática me parecía un tema demasiado complejo. Pero poco a poco, con la ayuda de Paco, Pau y el resto de bioinformáticos, aprendí las entrañas de esta parte de la ciencia que a veces pasa desapercibida. Agradeceremos todo el esfuerzo que habéis realizado para formarme y tener la suficiente destreza como para desarrollar este trabajo.

Agradecer también todos los momentos que he podido compartir con todas estas amistades que han surgido en estos cuatro años. No creo que hubiera podido sobrevivir sin vosotros, sin las cenas, y sin las risas de las carreras a la centrifuga... por si se rompía. También agradecer a mis mejores amigas Yasmina y Lea, que me han dado el apoyo moral y emocional tanto estos cuatro años como desde el momento que nos conocimos.

Por último, y no por ello menos importante, quería agradecer todos los momentos que he podido compartir con las personas más especiales de mi vida: mis padres. Gracias por apoyar todas mis aventuras científicas durante los veranos, ya sea yendo a Alemania, a Francia o a Irlanda. Gracias por ayudarme a descubrir mi pasión por la ciencia, y por motivarme a perseguir mis metas.

Estas palabras se quedan cortas de representar todo lo que significáis para mí y todos los momentos que me han llevado a ser la persona que soy hoy. Por todo ello, gracias a todos por todo y espero poder encontrar nuevos momentos que me lleven más cerca de mis metas.

Índice general

1. Introducción	1
1.1. Accidentes cerebrovasculares.....	1
1.1.1. Tipos de accidente cerebrovascular.....	3
1.1.2. Diferencias debidas al sexo.....	4
1.2. El papel de la bioinformática en el metaanálisis.....	5
1.2.1. Transcriptómica.....	5
1.2.2. La revisión sistemática y el metaanálisis.....	7
2. Objetivos	9
3. Materiales y métodos	10
3.1. Revisión sistemática.....	10
3.2. Análisis individual de los estudios.....	11
3.2.1. Imputación del sexo.....	11
3.2.2. Análisis exploratorio.....	12
3.2.3. Expresión diferencial.....	12
3.3. Metaanálisis de genes.....	13
4. Resultados	14
4.1. Revisión sistemática.....	14
4.2. Análisis individual de los estudios.....	15
4.2.1. Imputación del sexo.....	15
4.2.2. Análisis exploratorio.....	16
4.2.3. Expresión diferencial.....	18
4.3. Metaanálisis de genes.....	19
5. Discusión	24
6. Conclusiones	27
7. Bibliografía	28
Anexos	33
Figuras adicionales.....	33

Índice de figuras

Figura 1.1. Porcentaje de mortalidad dependiendo del tipo de enfermedad cardiovascular (datos adaptados de Joseph <i>et al.</i> , 2017).....	1
Figura 1.2. Impacto de las enfermedades neurológicas en 2015, en porcentajes, en relación a a) los años de vida ajustados por discapacidad (AVAD) y b) las muertes globales (GBD 2015 Neurological Disorders Collaborator Group, 2017).....	2
Figura 1.3. Número de defunciones totales por enfermedad en 2018, a nivel nacional para ambos sexos y ordenados de mayor a menor impacto (INE, 2018).....	2
Figura 1.4. Número de defunciones totales por enfermedad en 2018, a nivel nacional para mujeres y ordenados de mayor a menor impacto (INE, 2018).....	4
Figura 1.5. Esquematización de la cascada ómica y los diferentes elementos estudiados en cada subtipo (Bedia, 2018).....	6
Figura 1.6. Calidad de evidencia, ordenado de mayor a menor según el tipo de estudio realizado (adaptado de Ahn y Khan, 2018).....	7
Figura 3.1. Revisión sistemática siguiendo el estándar <i>Preferred Reporting Items for Systematic reviews and Meta-Analyses</i> (PRISMA).....	10
Figura 4.1. Resultado de la revisión sistemática siguiendo el estándar <i>Preferred Reporting Items for Systematic reviews and Meta-Analyses</i> (PRISMA).....	14
Figura 4.2. Diagrama de cajas y bigotes correspondiente al estudio GSE13353, por muestra. Las muestras rosa oscuro corresponden a mujeres patológicas; rosa claro, a mujeres control; azul oscuro, a hombres patológicos; y azul claro, a hombres control.....	16
Figura 4.3. PCA correspondiente al estudio GSE16561. Las muestras naranjas corresponden a mujeres patológicas; gris, a mujeres control; verde, a hombres patológicos; y azul, a hombres control.....	17
Figura 4.4. Análisis de grupos correspondiente al estudio GSE22255. Las muestras naranjas corresponden a mujeres patológicas; gris, a mujeres control; azul, a hombres patológicos; y verde, a hombres control.....	17
Figura 4.5. Representación en <i>Search Tool for the Retrieval of Interacting Genes/Proteins</i> (STRING) a) de los 25 genes más sobrerrepresentados en mujeres, b) marcando en rojo aquellos genes con funciones relacionadas con el sistema inmunológico y en azul aquellos con funciones relacionadas con la respuesta inmune.....	22
Figura 4.6. Representación en <i>Search Tool for the Retrieval of Interacting Genes/Proteins</i> (STRING) a) de los 25 genes más sobrerrepresentados en hombres, b) marcando en rojo aquellos genes con funciones relacionadas con la respuesta inflamatoria; en azul aquellos relacionados con quimiocina; y en verde aquellos relacionados con citoquina.....	23
Figura S1. PCA correspondiente al estudio GSE13353, por muestra. Las muestras naranjas corresponden a mujeres patológicas; gris, a mujeres control; verde, a hombres patológicos; y azul, a hombres control.....	33
Figura S2. Análisis de grupos correspondiente al estudio GSE13353. Las muestras naranjas corresponden a mujeres patológicas; gris, a mujeres control; azul, a hombres patológicos; y verde, a hombres control.....	33

Figura S3. Diagrama de cajas y bigotes correspondiente al estudio GSE16561, por muestra. Las muestras rosa oscuro corresponden a mujeres patológicas; rosa claro, a mujeres control; azul oscuro, a hombres patológicos; y azul claro, a hombres control.....	34
Figura S4. Análisis de grupos correspondiente al estudio GSE16561. Las muestras naranjas corresponden a mujeres patológicas; gris, a mujeres control; azul, a hombres patológicos; y verde, a hombres control.....	34
Figura S5. Diagrama de cajas y bigotes correspondiente al estudio GSE22255, por muestra. Las muestras rosa oscuro corresponden a mujeres patológicas; rosa claro, a mujeres control; azul oscuro, a hombres patológicos; y azul claro, a hombres control.....	35
Figura S6. PCA correspondiente al estudio GSE22255. Las muestras naranjas corresponden a mujeres patológicas; gris, a mujeres control; verde, a hombres patológicos; y azul, a hombres control.....	35
Figura S7. Diagrama de cajas y bigotes correspondiente al estudio GSE58294, por muestra. Las muestras rosa oscuro corresponden a mujeres patológicas; rosa claro, a mujeres control; azul oscuro, a hombres patológicos; y azul claro, a hombres control.....	36
Figura S8. PCA correspondiente al estudio GSE58294. Las muestras naranjas corresponden a mujeres patológicas; gris, a mujeres control; verde, a hombres patológicos; y azul, a hombres control.....	36
Figura S9. Análisis de grupos correspondiente al estudio GSE58294. Las muestras naranjas corresponden a mujeres patológicas; gris, a mujeres control; azul, a hombres patológicos; y verde, a hombres control.....	37
Figura S10. Diagrama de cajas y bigotes correspondiente al estudio GSE66724, por muestra. Las muestras rosa oscuro corresponden a mujeres patológicas; rosa claro, a mujeres control; azul oscuro, a hombres patológicos; y azul claro, a hombres control.....	37
Figura S11. PCA correspondiente al estudio GSE66724. Las muestras naranjas corresponden a mujeres patológicas; gris, a mujeres control; verde, a hombres patológicos; y azul, a hombres control.....	38
Figura S12. Análisis de grupos correspondiente al estudio GSE66724. Las muestras naranjas corresponden a mujeres patológicas; gris, a mujeres control; azul, a hombres patológicos; y verde, a hombres control.....	38

Índice de tablas

Tabla 4.1. Estudios incluidos en el análisis tras la revisión sistemática en <i>Gene Expression Omnibus</i> (GEO), indicando la plataforma utilizada para la obtención de los datos y la referencia.....	15
Tabla 4.2. Resultados de la imputación de sexo, resaltando las diferencias en rojo. La información del sexo se obtuvo por <i>Gene Expression Omnibus</i> (GEO), estudios destacados en negrita, o mediante los datos proporcionados por el estudio.....	15
Tabla 4.3. Resumen de la expresión diferencial por estudio. Se indica el número de genes totales, la comparación realizada, y si el <i>log fold change</i> es positivo (<i>up</i>) o negativo (<i>down</i>) para aquellos genes estadísticamente significativos con un valor p ajustado por Benjamini y Hochberg (BH) menor que 0,05. La comparación completa es (<i>Mujeres patológicas – Mujeres control</i>) – (<i>Hombres patológicos – Hombres control</i>). Los genes <i>up</i> estarán más expresados en mujeres que en hombres, y los genes <i>down</i> al revés. La comparación de las mujeres es <i>Mujeres patológicas – Mujeres control</i> , mientras que la de los hombres es <i>Hombres patológicos – Hombres control</i> . Los genes <i>up</i> estarán más expresados en pacientes patológicos, mientras que los <i>down</i> , lo estarán más en los controles.....	19
Tabla 4.4. Resumen del metaanálisis de genes, indicando el identificador (EntrezID) y el símbolo de cada gen, así como su <i>log fold change</i> (logFC) y el valor p ajustado. Aquellos con un logFC negativo están sobreexpresados en hombres, y los genes con logFC positivo lo están en mujeres.....	19
Tabla 4.5. Selección de los 25 genes más sobrerrepresentados en mujeres en el metaanálisis. Se indica su identificador (EntrezID), el símbolo del gen, el <i>log fold change</i> (logFC) y el valor p ajustado correspondiente. Aquellos genes con un logFC negativo están sobreexpresados en los grupos control, y los genes con logFC positivo lo están en el grupo patológico.....	20
Tabla 4.6. Selección de los 25 genes más sobrerrepresentados en hombres en el metaanálisis. Se indica su identificador (EntrezID), el símbolo del gen, el <i>log fold change</i> (logFC) y el valor p ajustado correspondiente. Aquellos genes con un logFC negativo están sobreexpresados en los grupos control, y los genes con logFC positivo lo están en el grupo patológico.....	21

Abreviaturas

ACV	Accidente cerebrovascular
Affymetrix	Affymetrix Human Genome U133 Plus 2.0 Array
ARN	Ácido ribonucleico
ARNm	Ácido ribonucleico mensajero
AVAD	Años de Vida Ajustados por Discapacidad
BH	Benjamini y Hochberg
DL	DerSimonian y Laird
EAC	Enfermedad de las arterias coronarias
FAIR	Findability, Accessibility, Interoperability, and Reusability
FDR	False Discovery Rate (tasa de descubrimientos falso)
FEC	Fundación Española del Corazón
GBD	Global Burden of Diseases, Injuries, and Risk Factors
GEO	Gene Expression Omnibus
HTS	High-Throughput Sequencing (secuenciación de alto rendimiento)
Illumina	Illumina HumanRef-8 v3.0 expression beadchip
INE	Instituto Nacional de Estadística
logFC	Log fold change (estimación del cambio de \log_2)
MAV	Malformaciones arteriovenosas
MeSH	Medical Subject Headings
NA	Not available (datos no disponibles)
NCBI	National Centre for Biotechnology Information
NHGRI	National Human Genome Research Institute
OMS	Organización Mundial de la Salud
PANTHER	Protein ANalysis THrough Evolutionary Relationships
PCA	Principal component analysis (análisis de componentes principales)
PRISMA	Preferred Reporting Items for Systematic reviews and Meta-Analyses
QUOROM	QUality Of Reporting Of Meta-analysis
RMA	Robust Multichip Average
STRING	Search Tool for the Retrieval of Interacting Genes/Proteins

1. Introducción

1.1. Accidentes cerebrovasculares

El accidente cerebrovascular (ACV) se define como la patología caracterizada por la interrupción o disminución del flujo sanguíneo cerebral. Este cese provoca una falta de nutrientes y oxígeno en el tejido que, si se prolonga, puede desencadenar la muerte celular del tejido (AMERICAN STROKE ASSOCIATION, 2014). Por tanto, esta patología puede ser considerada tanto una enfermedad cardiovascular al alterar el flujo sanguíneo, como una enfermedad neurológica por su impacto último en el cerebro.

Las enfermedades cardiovasculares se definen como aquellas que suponen un riesgo para la circulación sanguínea, provocando un bloqueo o disminución de su caudal (OMS, 2015). El término engloba diferentes enfermedades dependiendo de donde ocurra este fenómeno, como por ejemplo la enfermedad arterial oclusiva periférica o la embolia pulmonar. Sin embargo, las patologías cardiovasculares más importantes a nivel mundial son la enfermedad de las arterias coronarias (EAC) responsable del infarto de miocardio, y el ACV desencadenante del ictus. Al contrario del ACV que suele deberse a un desangramiento o coágulo de sangre, las EACs son debidas principalmente a la aterosclerosis, es decir, a la acumulación de depósitos de grasa formadores de placas, los cuales reducen el flujo sanguíneo al corazón (NATIONAL CANCER INSTITUTE, 2018).

Según la Organización Mundial de la Salud (OMS), las enfermedades cardiovasculares se consideran la principal causa de mortalidad en el mundo, siendo las responsables de 17,5 millones de defunciones en 2012, representando un 30% de las muertes globales. De ellas, 6,7 millones fueron debidas a ictus (OMS, 2015), tendencia que fue evidenciada por Joseph *et al.* (2017). La *Figura 1.1* representa los datos proporcionados por el estudio, indicando el porcentaje de mortalidad en función del tipo de enfermedad cardiovascular.

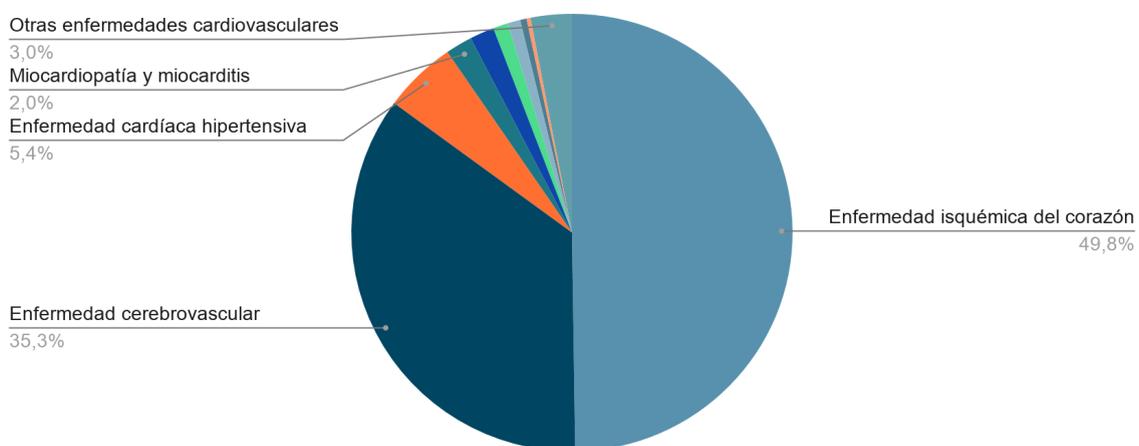


Figura 1.1. Porcentaje de mortalidad dependiendo del tipo de enfermedad cardiovascular (datos adaptados de Joseph *et al.*, 2017).

Por otro lado, las enfermedades neurológicas son aquellas que afectan al sistema nervioso central y periférico, englobando patologías del cerebro, de la médula espinal y de los músculos, entre otros (OMS, 2016). Estos trastornos incluyen demencias, ACV, tumores cerebrales, etc., siendo las más importantes las migrañas y la epilepsia (GBD 2015 Mortality and Causes of Death Collaborators, 2016). Cabe destacar que aunque el ictus está clasificado como enfermedad neurológica por sí sola, también suele ser un factor de riesgo en el desarrollo de otras enfermedades neurológicas como la epilepsia (Berg *et al.*, 2010), además de ser una de las principales causas de muerte prematura en estos pacientes (Thakur *et al.*, 2016).

El estudio *Global Burden of Diseases, Injuries, and Risk Factors (GBD)* partió de un esfuerzo común, entre diferentes colaboradores de todo el mundo, para caracterizar y cuantificar las muertes mundiales y sus causas (GBD 2015 Mortality and Causes of Death Collaborators, 2016). Este estudio ha permitido describir el impacto de una multitud de enfermedades, entre ellas el impacto del ictus, en comparación con el resto de enfermedades neurológicas (*Figura 1.2*). Observamos que, respecto al resto de patologías, el ACV es la principal causa tanto de reducción de años de vida ajustados por discapacidad (AVAD), como de muertes a nivel global.

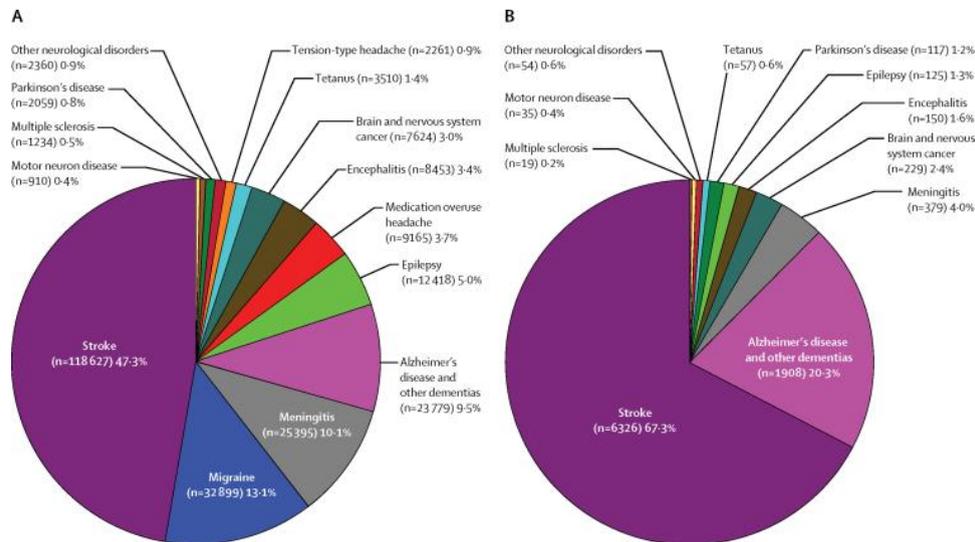


Figura 1.2. Impacto de las enfermedades neurológicas en 2015, en porcentajes, en relación a **a)** los años de vida ajustados por discapacidad (AVAD) y **b)** las muertes globales (GBD 2015 Neurological Disorders Collaborator Group, 2017).

Teniendo en cuenta todos los datos expuestos anteriormente, el ACV es una de las patologías más prevalentes desde el punto de vista cardiovascular y neurológico. Su impacto es doble, tanto en las muertes que causa, como en la posible discapacidad resultante de sobrevivir al evento.

En España, las cifras no se alejan de la tendencia global. La Fundación Española del Corazón (FEC) clasifica a los ACV como la segunda causa de defunción en nuestro país, y la primera en mujeres (FEC, 2012). El Instituto Nacional de Estadística (INE) notificó que en 2018 fallecieron de ictus 26.420 individuos, de los cuales 14.985 eran mujeres (INE, 2018) (*Figura 1.3*).

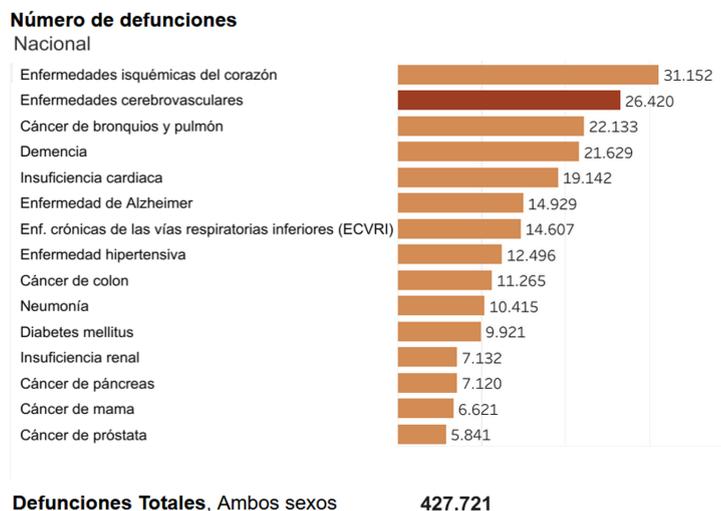


Figura 1.3. Número de defunciones totales por enfermedad en 2018, a nivel nacional para ambos sexos y ordenados de mayor a menor impacto (INE, 2018).

1.1.1. Tipos de accidente cerebrovascular

Aunque la clasificación de los ACV varía en función de la fuente consultada, existe un consenso en los principales subtipos. Esta clasificación depende de la modificación del flujo sanguíneo: isquémicos y hemorrágicos. Otros tipos de ictus menos prevalentes incluyen el ataque isquémico transitorio, el ACV del tronco encefálico y el ictus criptogénico, este último con una etiología indeterminada (AMERICAN STROKE ASSOCIATION, 2014).

Los ictus isquémicos son debidos a la interrupción total del flujo sanguíneo. Este se trata del tipo de ACV más prominente, representando alrededor del 87% de los casos (Sohrabji *et al.*, 2016). Sus causas pueden ser diversas, denotando, por ello, 5 subtipos de ictus isquémico: aterosclerosis de arterias grandes, cardioembolia, oclusión de vasos pequeños, ACV de otra etiología determinada y ACV de etiología indeterminada (Adams *et al.*, 1993).

Sin embargo, la principal causa de los ictus isquémicos es la aterosclerosis. Esta patología se caracteriza por la deposición de grasa, lo cual genera una obstrucción de venas y arterias. Esta oclusión puede ocurrir tanto en los propios vasos sanguíneos cerebrales como en otras partes del sistema circulatorio, habitualmente cerca del corazón. El desprendimiento de esta última podría llegar a desencadenar una embolia en el cerebro (AMERICAN STROKE ASSOCIATION, 2019b).

Por otra parte, el flujo sanguíneo cerebral también puede estar alterado debido a un sangrado interno, denominado ictus hemorrágico. Son menos frecuentes que los isquémicos, implicando un 13% de los casos (Sohrabji *et al.*, 2016). No obstante, son el tipo de ACV más mortal, con mayor número de muertes en jóvenes adultos (Feigin *et al.*, 2017).

Las causas de la hemorragia se originan en vasos sanguíneos debilitados que se rompen y provocan una acumulación de sangre en el tejido cerebral. Existen dos patologías asociadas a esta debilitación: los aneurismas y las malformaciones arteriovenosas (MAV).

Los aneurismas intracraneales son lesiones o ensanchamientos producidos en vasos sanguíneos debilitados, generalmente en zonas intracraneales. Entre un 80 y un 85% de los aneurismas se encuentran en la circulación anterior del cerebro (Schievink, 1997). Las formas más habituales se tratan de aneurismas saculares o de bayas (Toth y Cerejo, 2018). Estos pueden estar asociados a ciertos trastornos de vasos sanguíneos, como pueden ser la displasia fibromuscular o la arteritis cerebral, aunque no es la manifestación más habitual. La causa principal de la formación de aneurismas radica en la presión constante del flujo sanguíneo en los puntos ramificados de las arterias, lo cual crea una debilidad en ellas y el progresivo aumento de tamaño (AMERICAN STROKE ASSOCIATION, 2018b).

Por otra parte, las MAV son redes de vasos o nidos que, aun siendo alimentados por arterias y drenados por venas, no tienen capilares intermedios. Esto provoca un paso anómalo del flujo sanguíneo entre los sistemas arterial y venoso, caracterizado por un flujo alto y una baja resistencia en la red (Lawton *et al.*, 2015). Estas suelen ser congénitas pero no hereditarias, y son más prevalentes en hombres que en mujeres (AMERICAN STROKE ASSOCIATION, 2018a). Solo un 1% de las personas con estas malformaciones presentan un riesgo importante de sufrir una rotura, situación capaz de desatar una hemorragia. Sin embargo, una vez esta se produce, la reincidencia se quintuplica (Lawton *et al.*, 2015).

La clasificación del ictus hemorrágico depende de su localización, pudiendo ser de tipo intracerebral o subaracnoide (AMERICAN STROKE ASSOCIATION, 2019a).

El ACV hemorrágico intracerebral es el más prominente de los dos, siendo un 11% de los ictus totales (Feigin *et al.*, 2009). Se define como un sangrado en el parénquima cerebral, no causado a consecuencia de un trauma o cirugía. Se estima que de un 5 a 10% de pacientes con hemorragia intracerebral, manifestarán convulsiones clínicas. Además, un

30% presentarán actividad convulsiva eléctrica tras análisis con electroencefalograma, aunque no presenten síntomas clínicamente evidentes (Rymer, 2011). Por otro lado, el ictus hemorrágico subaracnoide se trata de la rotura de un vaso sanguíneo entre la membrana aracnoidea y la piamadre (Toth y Cerejo, 2018). En el 85% de los casos, este subtipo de ictus es debido a la rotura de un aneurisma intracraneal (Etminan *et al.*, 2019).

El impacto de los diferentes tipos de ACV depende de su localización geográfica. Por ejemplo, la incidencia del ictus hemorrágico intracerebral se ha reducido más que la del ACV isquémico en los países desarrollados. En cambio, la hemorragia subaracnoidea se ha mantenido uniforme en las últimas décadas. Sin embargo, la presencia del ictus hemorrágico es notablemente mayor en los países subdesarrollados (Feigin *et al.*, 2009).

1.1.2. Diferencias debidas al sexo

Debido a la gran prevalencia de esta enfermedad, se han estudiado diversos factores que pudieran afectar directa o indirectamente a los ACV, promoviendo o desfavoreciendo su aparición. Uno de los factores más estudiados es la posible relación entre el sexo y el impacto de la enfermedad.

El análisis llevado a cabo por Feigin *et al.* (2017) a partir de los datos proporcionados por el estudio GBD 2013, muestra una incidencia mayor en hombres frente a mujeres, tanto en el subtipo isquémico como el hemorrágico. No obstante, las proporciones varían a medida que la edad aumenta.

En los primeros años de vida y previo a la edad adulta, los hombres son más propensos a sufrir un ictus. Según Koellhoffer y McCullough (2013), esta diferencia puede ser debida a factores dependientes del sexo pero independientes de hormonas. Esta tendencia continúa hasta las edades más avanzadas. Los hombres son significativamente más jóvenes que las mujeres cuando sufren el primer ACV, siendo 71,1 la edad media para hombres y 75,1 para mujeres. Sin embargo, la incidencia por encima de los 85 años es superior en el caso de las mujeres (Petrea *et al.*, 2009).

Respecto a la incidencia a nivel nacional, el ictus es considerado la segunda causa de muerte, tal y como se ha mencionado anteriormente. Sin embargo, es la primera para mujeres (Figura 1.4). Estas diferencias pueden ser debidas a la mayor esperanza de vida que estas presentan. Además, cabe destacar que las mujeres sufren con mayor frecuencia hipertensión o fibrilación auricular, los cuales se tratan de dos factores de riesgo muy importantes en el desarrollo del ictus (Cordonnier *et al.*, 2017).

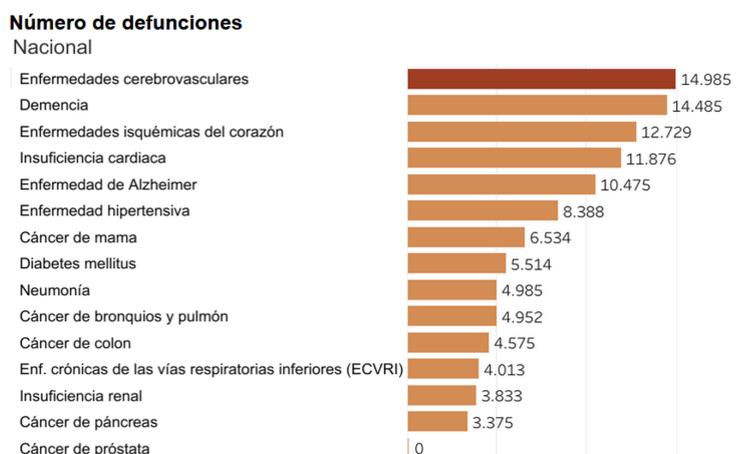


Figura 1.4. Número de defunciones totales por enfermedad en 2018, a nivel nacional para mujeres y ordenados de mayor a menor impacto (INE, 2018).

Defunciones Totales, Mujeres 211.279

Respecto a los subtipos de ACV, se han evidenciado algunas diferencias debidas al sexo. Algunos estudios informan que los ictus hemorrágicos subaracnoides y los isquémicos cardioembólicos son más comunes en mujeres (Appelros *et al.*, 2009), aunque otros indican que este sexo padece más ictus isquémicos en general (Sohrabji *et al.*, 2016).

Aunque existen indicadores clínicos y epidemiológicos que muestran una diferencia a nivel de sexo, la causa exacta sigue sin estar determinada. Cordonnier *et al.* (2017) indica que los factores que más contribuyen a estas diferencias radican en la biología vascular, la inmunidad, la coagulación, los perfiles hormonales y el estilo de vida.

Tanto Cordonnier *et al.* (2017) como Saeed *et al.* (2017) señalan que ciertos factores de riesgo del ictus son únicos para las mujeres, como por ejemplo la menopausia o riesgos del embarazo y posparto. La menopausia está asociada con la pérdida de estrógenos, lo que provoca una menor neuroprotección y un mayor riesgo de ACV (Koellhoffer y McCullough, 2013). Por otra parte, el embarazo también puede conllevar ciertas alteraciones como la enfermedad hipertensiva del embarazo, un factor de riesgo importante. Además, el ictus hemorrágico intracerebral es la causa principal de muerte materna (Cordonnier *et al.*, 2017).

Sin embargo, el principal motivo de la falta de información sobre estas diferencias de sexo es la infrarrepresentación de las mujeres en ensayos clínicos (Cordonnier *et al.*, 2017).

1.2. El papel de la bioinformática en el metaanálisis

Luscombe *et al.* (2001) define la bioinformática como “la conceptualización de la biología en términos de macromoléculas, aplicando técnicas informáticas con el fin de entender y organizar la información asociada a estas moléculas, a gran escala”. Estas técnicas no se limitan únicamente a lenguajes de programación como puede ser R, sino que además, existen diferentes herramientas digitales más ergonómicas y accesibles de gran utilidad para cualquier investigador que las necesite. Un ejemplo de ellas es *Protein ANalysis THrough Evolutionary Relationships* (PANTHER) (Thomas *et al.*, 2003) o *Search Tool for the Retrieval of Interacting Genes/Proteins* (STRING) (Jensen *et al.*, 2009).

Según Bayat (2002), la bioinformática es una herramienta esencial para el procesamiento y el análisis de grandes volúmenes de datos moleculares en estudios biomédicos. Este incremento de datos se ha producido como consecuencia del desarrollo y abaratamiento de las tecnologías de alto rendimiento. Gracias a los repositorios públicos de datos, esta información está disponible para su uso compartido por cualquier miembro de la comunidad científica.

1.2.1. Transcriptómica

El término ciencias ómicas engloba a todo aquel estudio centrado en estudiar la muestra biológica mediante enfoques diversos que cubren desde el genotipo al fenotipo. Como consecuencia, han surgido subclasificaciones en función del escalón biológico analizado: genómica, cuando se estudian los genes; transcriptómica, al hablar de la transcripción de los mismos siendo el producto final el ácido ribonucleico (ARN) transcrito; proteómica, al analizar el resultado de la traducción del ácido ribonucleico mensajero (ARNm); y metabolómica, centrado en la combinación de los diferentes metabolitos presentes en un organismo para un determinado momento (Bedia, 2018) (*Figura 1.5*).

Cuando hablamos de transcriptómica, nos referimos al estudio de todos los transcritos, tanto los ARNm codificantes de proteínas, como los ARN no codificantes que realizan otras funciones variadas (Lowe *et al.*, 2017). Este análisis permite determinar los niveles de expresión de los diferentes genes de un ser vivo en concreto.

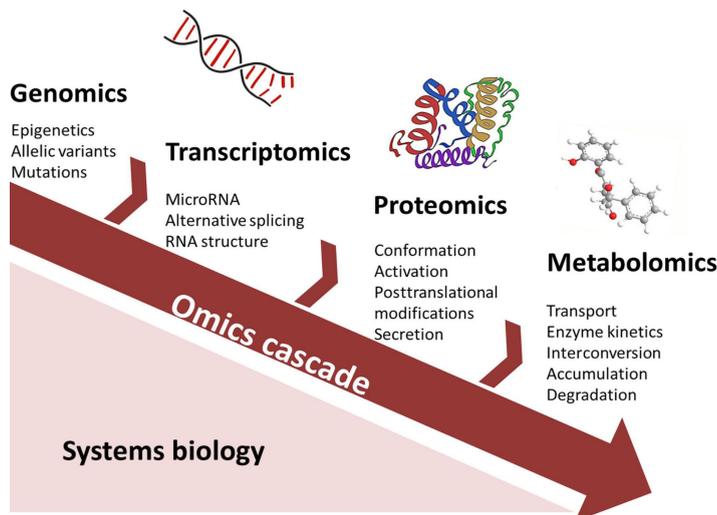


Figura 1.5. Esquematación de la cascada ómica y los diferentes elementos estudiados en cada subtipo (Bedia, 2018).

La generación de datos transcriptómicos puede llevarse a cabo mediante diferentes tecnologías de alto rendimiento, siendo las más importantes el empleo de microarrays y la secuenciación de ARN (RNA-seq).

Según el *National Human Genome Research Institute* (NHGRI) (2015), los microarrays son versiones en miniatura de técnicas realizadas en los laboratorios de genética molecular durante años. Se basan en la hibridación de moléculas de ácido desoxirribonucleico (ADN) o ARN, mediante el apareamiento de bases complementarias. Esta unión se puede cuantificar teniendo en cuenta la etiqueta asociada a las secuencias del microarray o la cantidad de luz producida en dicho apareamiento. Los avances respecto a la detección fluorescente han permitido un incremento en la sensibilidad de la técnica. Las tecnologías más conocidas actualmente que utilizan esta metodología se tratan de los microarrays de dos canales y los GeneChips de Affymetrix (Lowe *et al.*, 2017).

Por otro lado, la tecnología RNA-seq emplea *high-throughput sequencing* (HTS) o secuenciación de alto rendimiento. Esta se basa en la generación de una librería de ADN complementario (ADNc) con adaptadores en sus extremos, pudiendo realizar secuenciación por un solo extremo o por ambos, denominándose *single-end* o *pair-end sequencing* respectivamente (Wang *et al.*, 2009). Este método de secuenciación permite generar gran cantidad de fragmentos, incluyendo la cuantificación precisa de transcriptomas humanos completos. La principal tecnología que utiliza esta técnica es Illumina (Lowe *et al.*, 2017).

La elección de la tecnología depende tanto del diseño del experimento como de la muestra. Si se conoce la secuencia del gen expresado a analizar, puede resultar más útil la realización de un análisis transcriptómico mediante microarrays. Sin embargo, si se dispone de poca muestra y se desea reducir la mayor cantidad posible de ruido de fondo, es recomendable utilizar Illumina por su mayor eficiencia (Wang *et al.*, 2009).

El constante refinamiento y mejora de estas técnicas ha posibilitado el abaratamiento de costes. Ello ha provocado una expansión en la cantidad de datos transcriptómicos que se encuentran disponibles. No solo eso, sino que además, permite analizar más cantidad de muestras y comparar el nivel de expresión entre ellas. Esta mayor accesibilidad proporciona obtener variaciones de expresión debidas al tejido, sexo o estado fisiológico del individuo.

No obstante, para llevar a cabo esta extrapolación y el análisis posterior a gran escala, se requiere recopilar un volumen importante de datos transcriptómicos, así como emplear las herramientas necesarias. Por todo ello, resulta fundamental recurrir a bases de datos y a herramientas bioinformáticas complementarias para realizar un metaanálisis de grupos de datos de diversos estudios, que permitirán conocer la señal biológica común en todos ellos, proporcionando un perfil de expresión consensuado.

1.2.2. La revisión sistemática y el metaanálisis

Según Ahn y Khan (2018), el uso de la revisión sistemática y el metaanálisis permiten obtener información a gran escala, superando con ello las dificultades que dicho estudio controlado y aleatorio podría conllevar. Esto ha propiciado el aumento de este tipo de análisis en los últimos años.

Si clasificamos los diferentes tipos de estudios que se pueden realizar en función del nivel de evidencia científica, la revisión sistemática y el metaanálisis se situarían en su escalón más alto (*Figura 1.6*). La razón de ello se debe a la gran cantidad de datos empleados en este tipo de análisis.



Figura 1.6. Calidad de evidencia, ordenado de mayor a menor según el tipo de estudio realizado (adaptado de Ahn y Khan, 2018).

Sin embargo, este gran volumen de datos requiere su almacenamiento y accesibilidad para usos posteriores, necesitándose para ello bases de datos que permitan su gestión. Una de las bases de datos más importantes a día de hoy se trata de *Gene Expression Omnibus* (GEO) (Edgar *et al.*, 2002), con casi 5 millones de muestras. Esta recopila todo tipo de información transcriptómica procedente de diferentes plataformas, desde arrays hasta HTS. Además, permite buscar por palabras clave, así como filtrar por organismo, tipo de plataforma, número de muestras, etc. Su sistema de búsqueda se basa en los términos *Medical Subject Headings* (MeSH), un diccionario propio del *National Centre for Biotechnology Information* (NCBI). De esta manera, se puede complementar y ampliar la búsqueda mediante términos similares de la enfermedad a estudiar.

La revisión sistemática posibilita la recopilación de estudios obtenidos a partir de las bases de datos y otros tipos de fuentes de información, fijando como objetivo la extrapolación de conclusiones biológicas (Ahn y Kang, 2018). Se basa en la búsqueda primaria y complementaria, excluyendo todo estudio que no tenga las características deseadas. Por ello, se debe evaluar tanto la calidad del estudio como su pertinencia al posterior metaanálisis que se va a realizar (Ahn y Kang, 2018).

Sin embargo, se debe seguir un protocolo estándar con el fin de asegurar que esta búsqueda ha sido eficiente y suficiente, además de asegurar que se han realizado todos los pasos necesarios para incluir o excluir un estudio del análisis. Por este motivo, Moher *et al.* (1999) redactaron el *Quality Of Reporting Of Meta-analysis* (QUOROM), el cual describe los pasos a seguir en estos procesos junto con su razonamiento.

Desde entonces, se han descrito otros métodos más afines a la cantidad de información disponible actualmente. La mayoría de revisiones sistemáticas se desarrollan siguiendo el estándar *Preferred Reporting Items for Systematic reviews and Meta-Analyses* (PRISMA), establecido por Liberati *et al.* (2009).

Por otro lado, el metaanálisis permite realizar un resumen de la información expuesta en los diferentes estudios encontrados mediante revisión sistemática (Normand, 1999). Esto permite detectar el efecto general y el beneficio de un tratamiento, evaluar la variabilidad entre estudios o identificar diferencias dependiendo de una patología.

La combinación de estas dos técnicas complementarias proporciona un mayor número de resultados fiables que a partir de un solo estudio, al eliminar posibles factores de confusión (Ahn y Kang, 2018).

2. Objetivos

El objetivo principal de este trabajo se basa en estudiar las diferencias debidas al sexo que pueden propiciar el desarrollo del ictus, identificando un perfil consenso transcriptómico. Se aplicará una estrategia bioinformática con el fin de alcanzar los siguientes objetivos específicos:

- 1) Llevar a cabo una revisión sistemática y selección de estudios relacionados con la patología e información del sexo, mediante repositorios públicos de datos ómicos.
- 2) Realizar un análisis bioinformático capaz de caracterizar las diferencias de sexo, para cada uno de los estudios seleccionados.
- 3) Integrar los resultados del conjunto de estudios mediante técnicas de metaanálisis.

El cumplimiento de estos objetivos permitirá identificar un perfil consenso transcriptómico de las diferencias de sexo en los ACV.

3. Materiales y métodos

El metaanálisis realizado, y todos sus pasos previos, se ha realizado en el lenguaje de programación R (R Core Team, 2020). Todo el código de programación generado en este trabajo está disponible en el repositorio gitlab (<https://gitlab.com/ubb-cipf/tfgs-stroke/tfg-irene-canovas>).

3.1. Revisión sistemática

La revisión sistemática se llevó a cabo siguiendo el estándar de PRISMA (Liberati *et al.*, 2009). La *Figura 3.1* detalla los pasos seguidos, afines al estándar. Esta metodología permite llevar a cabo la búsqueda de datos de forma eficiente y escalonada.

Además, este estándar aconseja la realización de la revisión sistemática por dos o más personas diferentes, con el fin de cotejar los resultados. Por ello, esta revisión se realizó por varios integrantes del laboratorio.

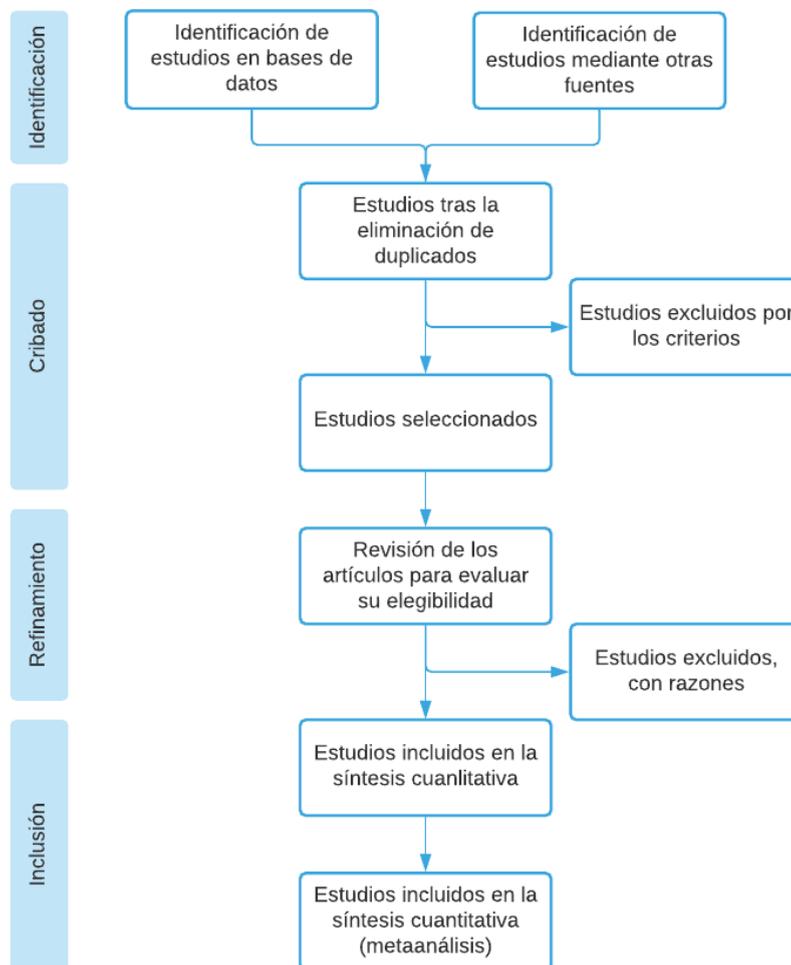


Figura 3.1. Revisión sistemática siguiendo el estándar *Preferred Reporting Items for Systematic reviews and Meta-Analyses* (PRISMA).

En primer lugar se llevó a cabo una búsqueda en la base de datos GEO (Edgar *et al.*, 2002): "stroke"[MeSH Terms] OR stroke[All Fields] AND ("gse"[Filter] AND "Homo sapiens"[Organism] AND ("Expression profiling by array"[Filter] OR "Expression profiling by high throughput sequencing"[Filter]) AND ("6"[n_samples] : "1000000"[n_samples])).

Se realizaron dos búsquedas más, utilizando términos MeSH de stroke. En la primera se añadió *strokes* a la búsqueda: "stroke"[MeSH Terms] OR stroke[All Fields] OR Strokes[All Fields] AND ("gse"[Filter] AND "Homo sapiens"[Organism] AND ("Expression profiling by array"[Filter] OR "Expression profiling by high throughput sequencing"[Filter]) AND ("6"[n_samples] : "1000000"[n_samples])). En la siguiente búsqueda se añadió el término "cerebrovascular accident": "stroke"[MeSH Terms] OR stroke[All Fields] OR Cerebrovascular Accident[All Fields] AND ("gse"[Filter] AND "Homo sapiens"[Organism] AND ("Expression profiling by array"[Filter] OR "Expression profiling by high throughput sequencing"[Filter]) AND ("6"[n_samples] : "1000000"[n_samples])).

Siguiendo el estándar PRISMA, se realizó también una búsqueda complementaria en otra base de datos distinta, *ArrayExpress* (Athar et al., 2019), con el fin de abarcar la máxima información disponible. En esta base de datos se filtró por organismo (*Homo sapiens*), por tipo de experimento (*RNA Assay* → *Sequencing Assay / Array Assay*) y por array (*All arrays*). Por último se especificó que se mostraran los datos únicos a la base de datos *ArrayExpress*, con el fin de reducir los duplicados.

Tras esta búsqueda, se excluyeron los duplicados, eliminando principalmente las *SuperSeries* relacionadas de la búsqueda en GEO, que suelen englobar varios estudios.

Se verificó que los estudios cumplieran con los requisitos de inclusión, eliminando aquellos que no cumplieran alguno de estos criterios:

- Muestras relacionadas con pacientes que han sufrido un ACV
- Datos transcriptómicos
- Muestras de *Homo sapiens*

Seguidamente, se revisaron los artículos relacionados con los datos transcriptómicos encontrados. Los estudios fueron excluidos si no cumplían con los requisitos de inclusión, si no había suficientes muestras, o si no se aportaba la información de sexo.

Para el grupo de estudios seleccionados, se descargaron sus datos y a continuación se llevó a cabo un análisis bioinformático en cada uno de ellos.

3.2. Análisis individual de los estudios

Previo al análisis conjunto de todos los estudios, estos se deben estudiar individualmente para confirmar la inclusión del estudio mediante el análisis exploratorio y generar los datos de partida para la realización del propio metaanálisis.

3.2.1. Imputación del sexo

Aunque en nuestra revisión sistemática encontramos 5 estudios con información de sexo que cumplieran con los requisitos de inclusión, no todos ellos proporcionaban la información en la base de datos de origen.

Para aquellos estudios, en los que se indicaban la cantidad de individuos de cada sexo sin identificar cada una de ellas, se realizó una imputación de sexo utilizando la función *imputeSex* del paquete *MetalIntegrator* (Haynes et al., 2016). Esta función analiza la expresión genética de cada muestra segregando aquellos genes característicos de cada sexo. Si no se aportan genes específicos, *imputeSex* emplea los genes provistos en su función. Para identificar mujeres, utiliza genes asociados al cromosoma X que escapan a su inactivación, como *XIST*, *RPS4X* o *CD40LG*. Por otro lado, utiliza genes ligados al cromosoma Y para identificar las muestras de hombre, como *KDM5D*, *RPS4Y1* o *EIF1AY*.

3.2.2. Análisis exploratorio

Cuando se utilizan datos públicos que provienen de microarrays es posible que existan valores perdidos o *not available* (NA). Por tanto, es necesario realizar una imputación de estos valores para poder llevar a cabo un correcto análisis exploratorio. Para ello existen varias técnicas, desde sustituir el valor perdido por la mediana de la expresión del gen en las demás muestras, hasta imputar el valor teniendo en cuenta los valores vecinos. Se utilizó esta segunda técnica mediante la función *knn.impute* del paquete *bnstruct* (Franzin *et al.*, 2017).

Seguidamente, se homogeneizó la escala de los datos mediante transformación logarítmica en base 2. Este paso es esencial cuando se analizan diferentes estudios ya que los datos pueden estar presentados en diferentes escalas dependiendo del estudio, o provienen de distintas plataformas, por lo que su análisis directo no proporciona resultados fiables.

A continuación, se clasificaron las muestras según:

- La condición de la muestra: si corresponde a un paciente control o a un paciente que ha sufrido un ictus. Esta información se obtuvo del *dataset* proporcionado en GEO.
- El sexo de la muestra: si es mujer o hombre. Esta información se obtuvo bien del *dataset* proporcionado en GEO, y en su ausencia, por la imputación de sexo, teniendo en cuenta los datos proporcionados por el estudio.

Por último, se realizaron diversos diagramas para obtener una visión generalizada de las regularidades existentes en los datos de cada estudio. El primero fue un *boxplot* o diagrama de caja y bigotes, para verificar que la mediana de expresión génica era similar en todas las muestras. Esto asegura que los datos han sido normalizados y puedan ser comparados entre estudios. Para ello se empleó la función *boxplot* propia de R.

Además, y como primer indicador de agrupaciones entre muestras, se realizó un *principal component analysis* (PCA) o análisis de componentes principales utilizando las funciones *pcaGenes* y *plotPCA* de *metaFunR* (<https://gitlab.com/ubb-cipf/metafunr>). También se hizo un *clustering* o análisis de grupos con la función *plotTreeClust* del mismo repositorio. Las representaciones gráficas de los resultados de este análisis exploratorio ayudan en la detección de muestras anómalas.

3.2.3. Expresión diferencial

En primer lugar se realizó una anotación de las sondas, traduciéndolas a un identificador único, *Entrez Gene* (EntrezID) (Maglott *et al.*, 2011). Esta conversión se realizó utilizando el paquete de *AnnotationDbi* (Pagès *et al.*, 2021), junto con el paquete *hgu133plus2.db* (Carlson, 2016), correspondiente a las sondas de “Affymetrix Human Genome U133 Plus 2.0 Array” (Affymetrix), y el paquete *illuminaHumanv4.db* (Dunning *et al.*, 2015), correspondiente a las anotaciones de las sondas de “Illumina HumanHT12v4” (Illumina).

Para realizar la expresión diferencial se empleó la expresión *diffExp*, que forma parte de *metaFunR*. Esta función necesita los datos transcriptómicos, el grupo de cada muestra, y el contraste o comparación deseado. El resultado es un objeto *MArrayLM* del paquete *limma* (Ritchie *et al.*, 2015).

Usando la función *topTable* del mismo paquete, se obtiene una tabla con:

- *genelist*: anotación de la sonda
- *logFC*: *log fold change* o estimación del cambio de \log_2 , respecto al contraste

- CI.L: límite izquierdo del intervalo de confianza para logFC
- CI.R: límite derecho del intervalo de confianza para logFC
- p.value: *p-value* o valor p sin ajustar
- adj.p.value: *adjusted p-value* o valor p ajustado. Este ajuste se realizó mediante el método de Benjamini y Hochberg (1995), también conocido como BH

La comparación de interés realizada fue la siguiente:

(Mujer Enferma – Mujer Sana) – (Hombre Enfermo – Hombre Sano)

Teniendo en cuenta este contraste, los genes con un logFC positivo, estarían expresados diferencialmente en las mujeres, con un nivel de expresión mayor en el grupo enfermo respecto al control; mientras aquellos con un logFC negativo, lo son en hombres. Sin embargo, solo los genes con un valor p ajustado por BH menor que 0,05 son estadísticamente expresados diferencialmente.

Además se realizó dos comparaciones complementarias para las mujeres y los hombres, respectivamente:

Mujer Enferma – Mujer Sana

Hombre Enfermo – Hombre Sano

En este caso, los genes con un logFC positivo corresponden a los genes diferencialmente expresados en los individuos patológicos, mientras que aquellos con un logFC negativo corresponden a individuos controles. De igual manera, solo se tienen en cuenta aquellos genes con un valor p ajustado por BH menor que 0,05.

3.3. Metaanálisis de genes

En primer lugar se combinaron los diferentes logFC del análisis de la expresión diferencial, respecto a cada gen y estudio. Se conservaron solo aquellos genes que presentaban un logFC en al menos 4 de los estudios. Utilizando la función *rma* del paquete *metafor* (Viechtbauer, 2010), se calculó la medida de expresión *Robust Multichip Average* (RMA) o promedio robusto de múltiples trazos. Se optó por utilizar el método de DerSimonian y Laird (1986), también conocido como DL, para estimar la heterogeneidad. Por último, el valor p ajustado se obtuvo gracias a la función *p.adjust* del paquete *stats* (R Core Team, 2020), por el método *False Discovery Rate* (FDR) o tasa de descubrimientos falso, siendo sinónimo del ajuste realizado por BH.

El resultado final es una tabla con los siguientes valores:

- logFC: *log fold change* o estimación del cambio de \log_2 , respecto al contraste
- p.value: *p-value* o valor p sin ajustar
- SE: *standard error* o error estándar
- *p.adjust.fdr. adjusted p-value by FDR* o valor p ajustado por FDR

De esta tabla, se seleccionaron aquellos genes con un valor p ajustado por FDR menor 0,05, así como los 25 genes con mayor logFC y 25 genes con menor logFC.

Para caracterizar funcionalmente los resultados del metaanálisis se llevó a cabo: 1) un análisis de interacción proteína-proteína que permite conocer el nivel de conexión entre los genes con mayor expresión diferencial, utilizando STRING (Jensen *et al.*, 2009) y 2) un análisis de sobrerrepresentación de los procesos biológicos, para identificar las funciones en las que participan los genes con un perfil específico de expresión por sexos, utilizando PANTHER (Thomas *et al.*, 2003). Ambas estrategias nos ayudarán a comprender el papel de los genes que son diferentes entre hombres y mujeres, en la patología estudiada.

4. Resultados

4.1. Revisión sistemática

Como podemos observar en la *Figura 4.1*, tras la revisión sistemática se encontraron 86 estudios en la plataforma GEO, mientras que en la plataforma *ArrayExpress* solo se encontraron 9. Eliminando los duplicados, se cribaron 91 estudios en total.

De estos estudios, 69 tuvieron que ser descartados por tres razones principales: 1) los estudios no estaban relacionados con ACV, 2) los datos proporcionados no correspondían a la expresión génica, 3) o bien las muestras no eran de humano. Seguidamente, se procedió a revisar los textos, excluyendo con ello 17 estudios más. Las causas de esta exclusión son debidas a las características de las muestras: por falta de ellas, tanto de pacientes control como de ACV; por combinación de muestras de diferentes individuos; y por falta de información de sexo en el texto.

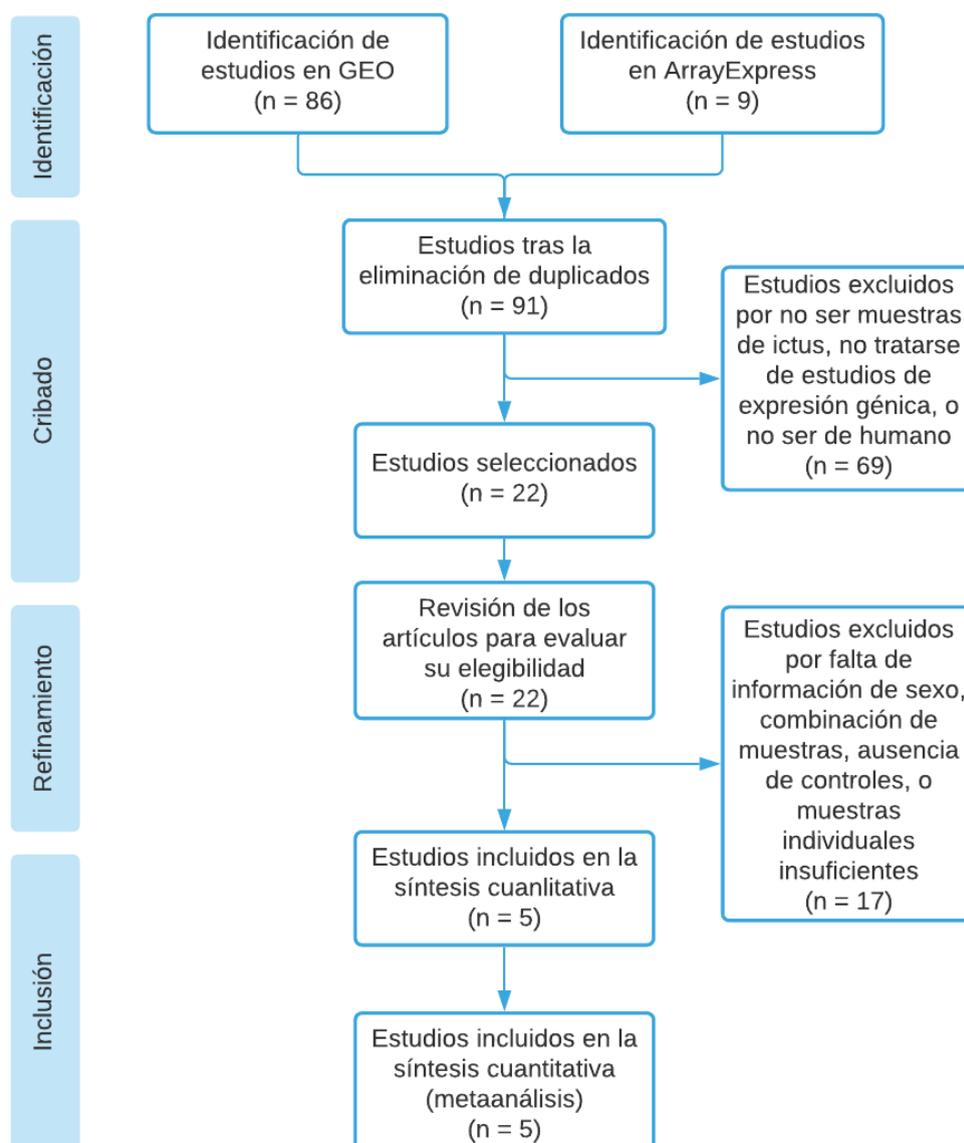


Figura 4.1. Resultado de la revisión sistemática siguiendo el estándar *Preferred Reporting Items for Systematic reviews and Meta-Analyses* (PRISMA).

Por tanto, se incluyeron 5 estudios en el análisis exploratorio así como en el metaanálisis posterior. Estos estudios se detallan en la *Tabla 4.1*, indicando la plataforma utilizada y la referencia.

Tabla 4.1. Estudios incluidos en el análisis tras la revisión sistemática en *Gene Expression Omnibus* (GEO), indicando la plataforma utilizada para la obtención de los datos y la referencia.

Estudio	Plataforma utilizada	Referencia
GSE13353	Affymetrix Human Genome U133 Plus 2.0 Array	Kurki <i>et al.</i> (2011)
GSE16561	Illumina HumanRef-8 v3.0 expression beadchip	Barr <i>et al.</i> (2010)
GSE22255	Affymetrix Human Genome U133 Plus 2.0 Array	Krug <i>et al.</i> (2012)
GSE58294	Affymetrix Human Genome U133 Plus 2.0 Array	Stamova <i>et al.</i> (2014)
GSE66724	Affymetrix Human Genome U133 Plus 2.0 Array	Allende <i>et al.</i> (2016)

4.2. Análisis individual de los estudios

4.2.1. Imputación del sexo

De los 5 estudios seleccionados, solo dos de ellos (GSE16561 y GSE22255) contenían la información del sexo en el *dataset* de GEO. Sin embargo, los estudios sin dicha información sí que detallan el número de muestras por grupo y sexo, por medio de tablas en los textos. Por ello, se optó por realizar una imputación del sexo de estas muestras.

La *Tabla 4.2* muestra el resultado de esta imputación, comparando el número de muestras por sexo y grupo de cada estudio. Podemos observar que la imputación ha resultado exitosa a excepción del estudio GSE22255 (Krug *et al.*, 2012), con un error de clasificación en las muestras control. Al no haber ningún error adicional, y teniendo la información correcta de este estudio proporcionada por la base de datos GEO, suponemos que los resultados de esta imputación son correctos.

Tabla 4.2. Resultados de la imputación de sexo, resaltando las diferencias en rojo. La información del sexo se obtuvo por *Gene Expression Omnibus* (GEO), estudios destacados en negrita, o mediante los datos propocionados por el estudio.

Estudio	Información proporcionada				Resultado de la imputación			
	Control		Ictus		Control		Ictus	
	Mujer	Hombre	Mujer	Hombre	Mujer	Hombre	Mujer	Hombre
GSE13353	5	3	8	3	5	3	8	3
GSE16561	14	10	22	17	14	10	22	17
GSE22255	10	10	10	10	11	9	10	10
GSE58294	12	11	12 (x3)	11 (x3)	12	11	12 (x3)	11 (x3)
GSE66724	3	5	3	5	3	5	3	5

4.2.2. Análisis exploratorio

GSE13353

El estudio realizado por Kurki *et al.* (2011) se centra en analizar el aneurisma intracraneal sacular, cuya rotura puede provocar un ictus hemorrágico. Para ello, obtienen el perfil de expresión del genoma completo, mediante la plataforma Affymetrix, con el fin de determinar mecanismos celulares y moleculares. Se compararon 11 muestras de aneurismas rotos, tras 15 horas de la rotura, con 8 muestras no rotas.

Los datos tuvieron que ser ajustados a escala logarítmica en base 2. La *Figura 4.2* muestra este ajuste, representando los datos en un diagrama de caja y bigotes. Tanto el PCA como el análisis de grupos se encuentran en el Anexo (*Figura S1* y *Figura S2*).

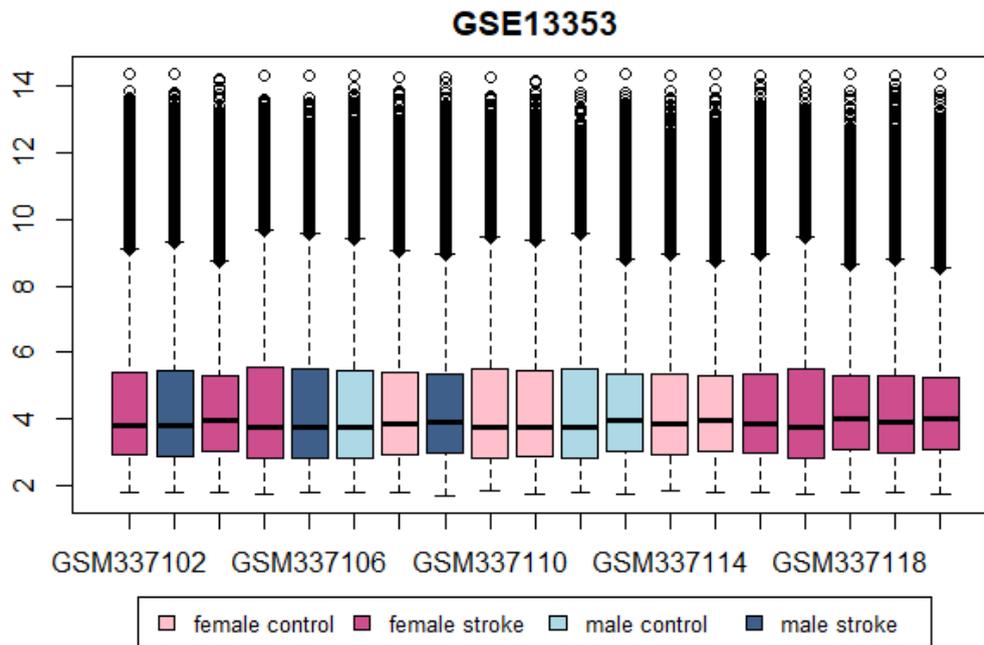


Figura 4.2. Diagrama de cajas y bigotes correspondiente al estudio GSE13353, por muestra. Las muestras rosa oscuro corresponden a mujeres patológicas; rosa claro, a mujeres control; azul oscuro, a hombres patológicos; y azul claro, a hombres control.

GSE16561

Con el fin de identificar genes para el diagnóstico preventivo de ictus isquémico, Barr *et al.* (2010) estudia la expresión génica en la sangre de pacientes tras el evento. Se extrajo el ARN total de sangre de 39 pacientes que sufrieron un ACV isquémico, analizando mediante Illumina, y comparando con 24 individuos sanos neurológicamente.

Los datos tuvieron que ser ajustados a escala logarítmica en base 2, realizando el subsecuente diagrama de caja y bigotes, PCA y análisis de grupos. El PCA se muestra en la *Figura 4.3*, mientras que el resto de diagramas están en el Anexo (*Figura S3* y *Figura S4*).

GSE22255

Krug *et al.* (2012) estudiaron las diferencias de expresión resultantes del ictus isquémico, en muestras de células mononucleares de sangre periférica. Se analizaron 20 pacientes patológicos y 20 pacientes controles, mediante la tecnología de microarrays Affymetrix.

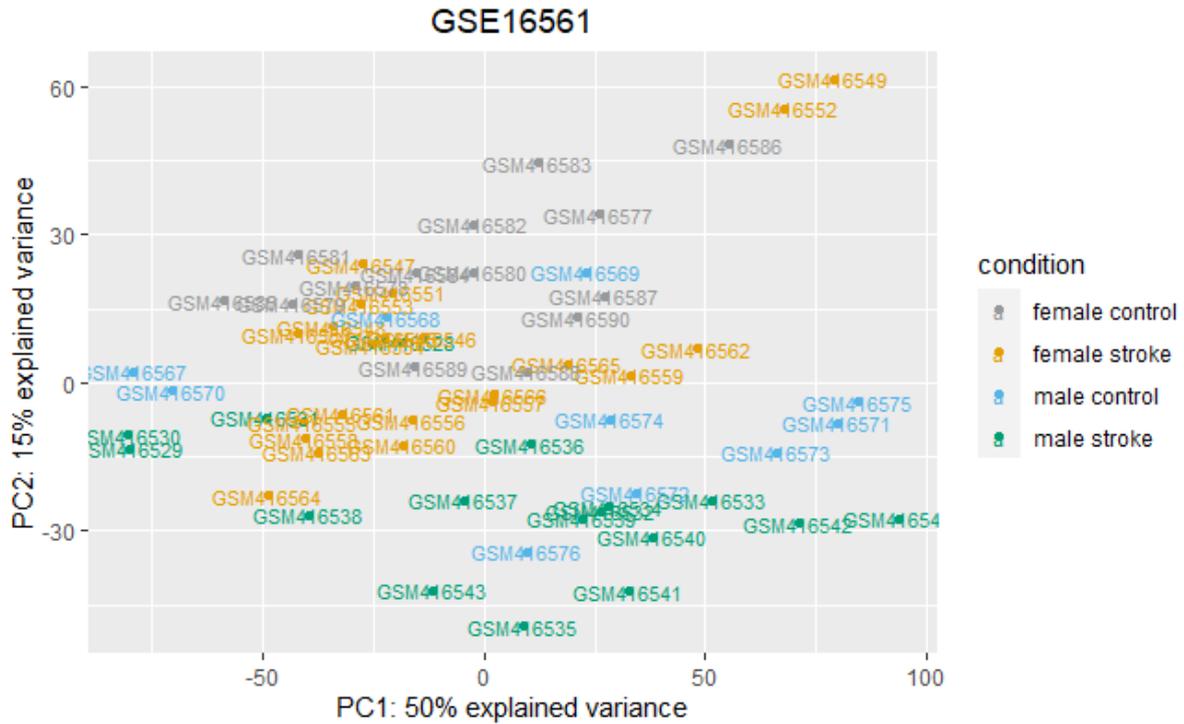


Figura 4.3. PCA correspondiente al estudio GSE16561. Las muestras naranjas corresponden a mujeres patológicas; gris, a mujeres control; verde, a hombres patológicos; y azul, a hombres control.

Los datos estaban normalizados pero presentaban muchas NA, por lo que se procedió a la imputación por valores próximos. Seguidamente, se realizó el diagrama de caja y bigotes, PCA y análisis de grupos. Este último se muestra en la *Figura 4.4*, estando el resto diagramas en el Anexo (*Figura S5* y *Figura S6*).

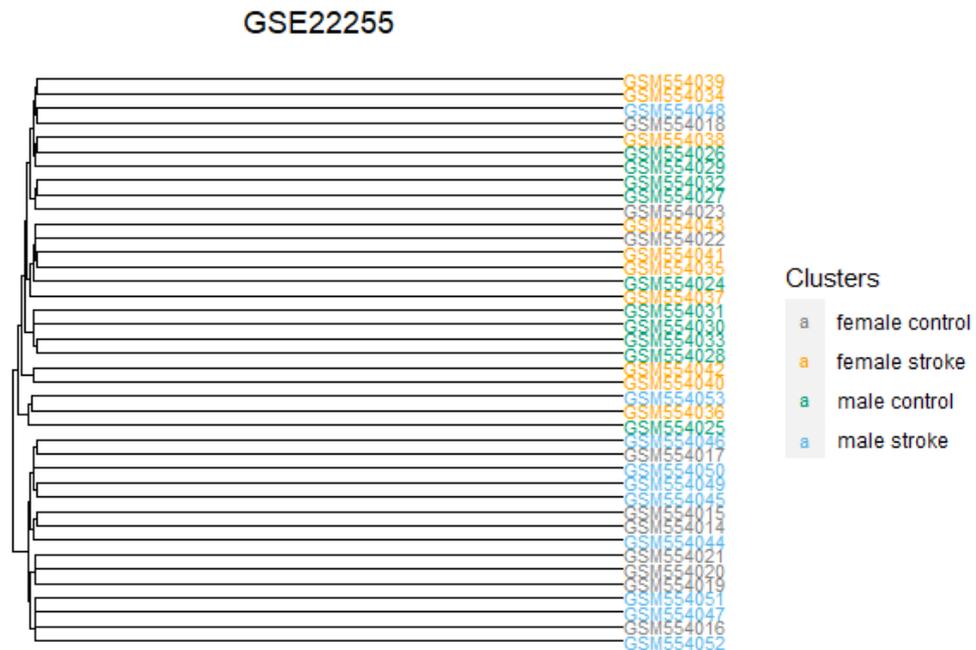


Figura 4.4. Análisis de grupos correspondiente al estudio GSE22255. Las muestras naranjas corresponden a mujeres patológicas; gris, a mujeres control; azul, a hombres patológicos; y verde, a hombres control.

GSE58294

El estudio realizado por Stamova *et al.* (2014) se centra en analizar las diferencias en expresión génica tras un ictus isquémico cardioembólico. Para ello, cuenta con 23 pacientes patológicos, a los que ha extraído muestras 3, 5 y 24 horas después del evento. Se realizó el análisis en Affymetrix, comparando con 23 muestras control.

Con el fin de tener un conjunto de muestras homogéneo, se descartaron las repeticiones, realizando los análisis únicamente con aquellos datos transcriptómicos correspondientes a las muestras obtenidas tras 24 horas del evento.

Se realizaron los correspondientes gráficos que se encuentran en el Anexo, siendo la *Figura S7* el diagrama de cajas y bigotes, la *Figura S8* el PCA y la *Figura S9* el análisis de grupos.

GSE66724

El estudio de Allende *et al.* (2016) analiza las diferencias en el perfil de expresión génica en pacientes con fibrilación auricular, ya que esta es un factor de riesgo para el ictus isquémico cardioembólico. Por ello, realizaron un análisis con microarray, comparando pacientes que habían sufrido un ictus con aquellos que, aunque presentaban la patología, no habían sufrido un evento.

Las muestras totales analizadas fueron 19, siendo 8 de pacientes isquémicos y 8 pacientes control. Tanto el diagrama de cajas y bigotes, el PCA y el análisis de grupos, se encuentran en el Anexo (*Figura S10*, *Figura S11* y *Figura S12*).

4.2.3. Expresión diferencial

Tras realizar la anotación correspondiente, se realizó el análisis de la expresión diferencial. La *Tabla 4.3* muestra el número de genes estadísticamente significativos, con un valor *p* ajustado por BH menor a 0,05. La comparación completa es:

(Mujeres patológicas – Mujeres control) – (Hombres patológicos – Hombres control)

Siendo los genes expresados más en mujeres aquellos con un logFC positivo (*up*), observamos que el único estudio con expresión diferencial es el GSE22255. Los genes más expresados en hombre, con un logFC negativo (*down*), siguen la misma tendencia, solo presentando resultados significativos en el mismo estudio.

Por otro lado, la comparación de las mujeres y hombres son, respectivamente:

Mujeres patológicas – Mujeres control y Hombres patológicos – Hombres control

Aquí observamos más genes diferencialmente expresados, aunque ni el estudio GSE13353 ni el GSE66724 presentan genes expresados diferencialmente de forma estadísticamente significativa. En este caso, los genes *up* serían aquellos sobrerrepresentados en los pacientes patológicos, y los *down*, sobrerrepresentados en pacientes control.

Sin embargo, hay que tener en cuenta las limitaciones de las comparaciones realizadas. Por ejemplo, si un gen está subrepresentado en un grupo, aparecerá como sobrerrepresentado en el otro, aunque la expresión sea basal.

Tabla 4.3. Resumen de la expresión diferencial por estudio. Se indica el número de genes totales, la comparación realizada, y si el *log fold change* es positivo (*up*) o negativo (*down*) para aquellos genes estadísticamente significativos con un valor p ajustado por Benjamini y Hochberg (BH) menor que 0,05. La comparación completa es (*Mujeres patológicas – Mujeres control*) – (*Hombres patológicos – Hombres control*). Los genes *up* estarán más expresados en mujeres que en hombres, y los genes *down* al revés. La comparación de las mujeres es *Mujeres patológicas – Mujeres control*, mientras que la de los hombres es *Hombres patológicos – Hombres control*. Los genes *up* estarán más expresados en pacientes patológicos, mientras que los *down*, lo estarán más en los controles.

Estudio	Genes totales	Completa		Mujer		Hombre	
		<i>Up</i>	<i>Down</i>	<i>Up</i>	<i>Down</i>	<i>Up</i>	<i>Down</i>
GSE13353	20962	0	0	0	0	0	0
GSE16561	17794	0	0	819	717	416	345
GSE22255	20962	2016	1968	133	271	1631	1847
GSE58294	20962	0	0	4546	923	2242	1261
GSE66724	20962	0	0	0	0	0	0

4.3. Metaanálisis de genes

El metaanálisis de genes identificó 7 genes, con un valor p ajustado por FDR menor que 0,05 (*Tabla 4.4*). De ellos, dos presentaban un logFC negativo, por lo que estaban sobreexpresados en hombres. El resto de genes con logFC positivo, están sobreexpresados en mujeres.

Tabla 4.4. Resumen del metaanálisis de genes, indicando el identificador (EntrezID) y el símbolo de cada gen, así como su *log fold change* (logFC) y el valor p ajustado. Aquellos con un logFC negativo están sobreexpresados en hombres, y los genes con logFC positivo lo están en mujeres.

EntrezID	Gen	logFC	Valor p ajustado
100129029	PKP4-AS1	-0,386	0,012
93624	TADA2B	-0,338	0,040
29097	CNIH4	0,578	0,040
388327	C17orf100	0,638	0,038
26157	GIMAP2	0,652	0,001
388588	SMIM1	0,671	0,046
80862	ZNRD1ASP	0,687	0,020

Con el fin de obtener una representación de las interacciones entre los diferentes genes, se procedió a seleccionar los 25 genes más sobrerrepresentados en mujeres (mayor logFC). Los genes se detallan en la *Tabla 4.5*, indicando su identificador, símbolo, logFC y valor p ajustado. Se realizó la misma selección para los hombres, seleccionando los 25 genes más sobrerrepresentados (menor logFC). Los genes se detallan en la *Tabla 4.6*, indicando su identificador, símbolo, logFC y valor p ajustado.

Tabla 4.5. Selección de los 25 genes más sobrerrepresentados en mujeres en el metaanálisis. Se indica su identificador (EntrezID), el símbolo del gen, el *log fold change* (logFC) y el valor p ajustado correspondiente. Aquellos genes con un logFC negativo están sobreexpresados en los grupos control, y los genes con logFC positivo lo están en el grupo patológico.

EntrezID	Gen	logFC	Valor p ajustado
8781	PSPHP1	1,957	0,906
1669	DEFA4	1,125	0,734
3429	IFI27	1,124	0,536
54674	LRRN3	1,11	0,85
7504	XK	0,935	0,862
10561	IFI44	0,921	0,446
1178	CLC	0,914	0,867
10964	IFI44L	0,891	0,457
91543	RSAD2	0,85	0,536
51327	AHSP	0,84	0,85
3434	IFIT1	0,822	0,808
566	AZU1	0,806	0,137
56670	SUCNR1	0,801	0,054
3045	HBD	0,794	0,858
55278	QRSL1	0,785	0,817
152518	NFXL1	0,784	0,785
1088	CEACAM8	0,782	0,902
101928370	LOC101928370	0,777	0,137
101928429	LOC101928429	0,773	0,826
729230	CCR2	0,773	0,851
8743	TNFSF10	0,773	0,815
400713	ZNF880	0,770	0,617
671	BPI	0,769	0,812
166929	SGMS2	0,766	0,855
57182	ANKRD50	0,766	0,870

Tabla 4.6. Selección de los 25 genes más sobrerrepresentados en hombres en el metaanálisis. Se indica su identificador (EntrezID), el símbolo del gen, el *log fold change* (logFC) y el valor p ajustado correspondiente. Aquellos genes con un logFC negativo están sobrerrepresentados en los grupos control, y los genes con logFC positivo lo están en el grupo patológico.

EntrezID	Gen	logFC	Valor p ajustado
114614	MIR155HG	-2,314	0,906
6351	CCL4	-1,775	0,734
3553	IL1B	-1,761	0,536
2921	CXCL3	-1,682	0,85
6364	CCL20	-1,568	0,862
22822	PHLDA1	-1,511	0,446
84419	C15orf48	-1,468	0,867
5743	PTGS2	-1,433	0,457
2152	F3	-1,385	0,536
4929	NR4A2	-1,327	0,85
8507	ENC1	-1,311	0,808
23645	PPP1R15A	-1,258	0,137
100302650	BABAM2-AS1	-1,233	0,054
5806	PTX3	-1,225	0,858
3569	IL6	-1,200	0,817
9308	CD83	-1,187	0,785
2069	EREG	-1,160	0,902
5209	PFKFB3	-1,132	0,137
54855	TENT5C	-1,115	0,826
8843	HCAR3	-1,107	0,851
2919	CXCL1	-1,071	0,815
7130	TNFAIP6	-1,066	0,617
58476	TP53INP2	-1,064	0,812
5142	PDE4B	-1,056	0,855
23764	MAFF	-1,044	0,870

Se realizó un análisis mediante STRING, que muestra en un gráfico de vértices y nodos las interacciones de proteínas presentes en una lista de genes proporcionados. Además, los genes fueron evaluados mediante PANTHER, herramienta que indica las posibles funciones que desempeñan los genes proporcionados.

El diagrama de los genes más expresados en mujeres, muestra 23 conexiones, creando dos redes principales (Figura 4.5). Estas interacciones proteína-proteína son significativas, ya que para una red con este mismo número de nodos, el número de conexiones al azar sería 1, siendo el valor p del enriquecimiento menor a 10^{-16} . Se señalan en rojo aquellos genes que tienen funciones relacionadas con el sistema inmunológico, mientras que en azul se indican aquellas que están relacionadas con la respuesta inmune, en especial, vírica. Destacar que el análisis de PANTHER también denotaba funciones con regulación de la producción de citocinas de células T, aunque esto no se mostraba en STRING. [A]

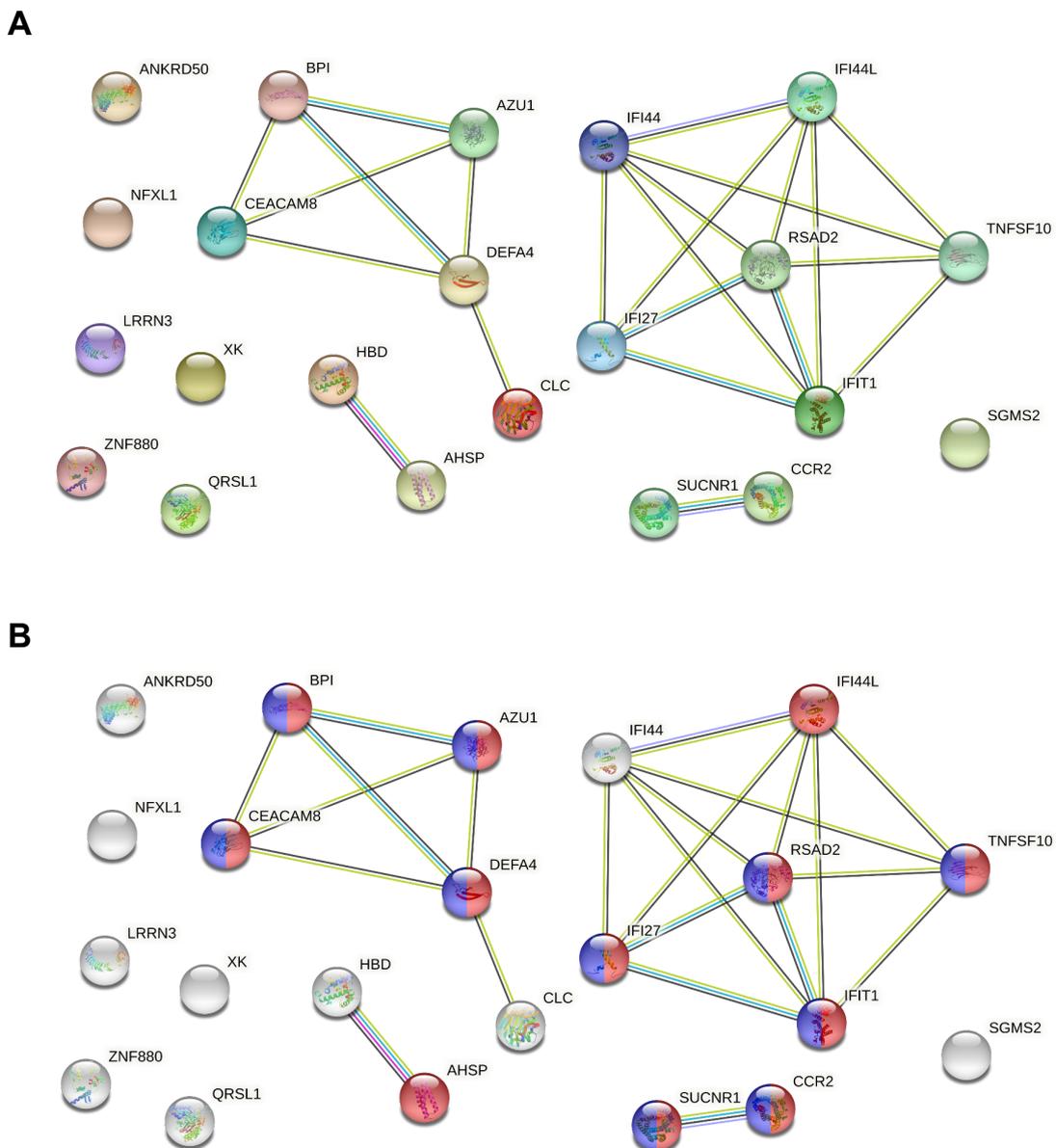
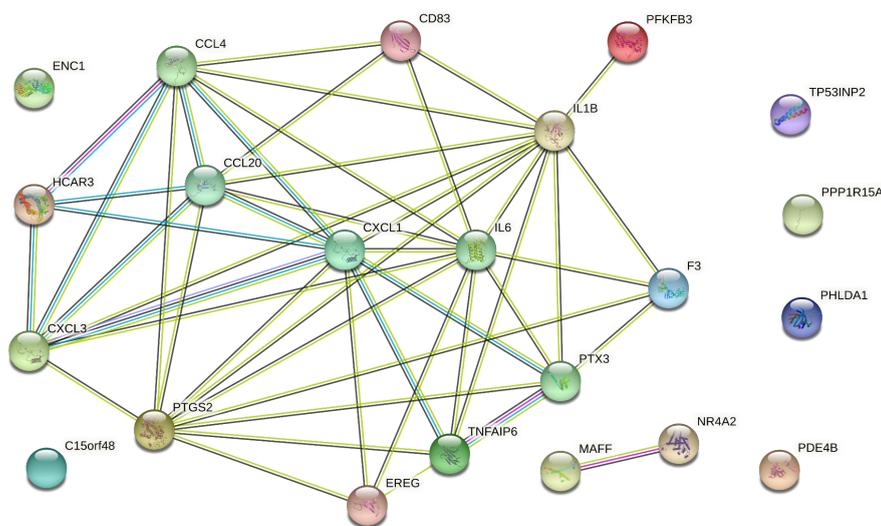


Figura 4.5. Representación en *Search Tool for the Retrieval of Interacting Genes/Proteins* (STRING) **a)** de los 25 genes más sobrerrepresentados en mujeres, **b)** marcando en rojo aquellos genes con funciones relacionadas con el sistema inmunológico y en azul aquellos con funciones relacionadas con la respuesta inmune.

Este doble análisis también se realizó con los genes más expresados en hombre (*Figura 4.6*). El diagrama creado por STRING muestra una gran red principal, con muchas interacciones entre sí. El número de conexiones es 48, siendo 7 el número de conexiones al azar que la red tendría y el valor p del enriquecimiento menor a 10^{-16} . Por lo tanto, existe un nivel significativo de conectividad entre los genes con mayor expresión diferencial, lo que apunta a que probablemente estén compartiendo funciones biológicas además de presentar un perfil común de expresión.

La evaluación mediante PANTHER mostró muchos genes con funciones relacionadas con la respuesta inflamatoria, en rojo; relacionadas con quimiocina, en azul; y relacionadas con citoquina en verde. También halló funciones de regulación positiva de la respuesta neuroinflamatoria y de señalización de interleucina-10, aunque estas no se mostraron en STRING.

A



B

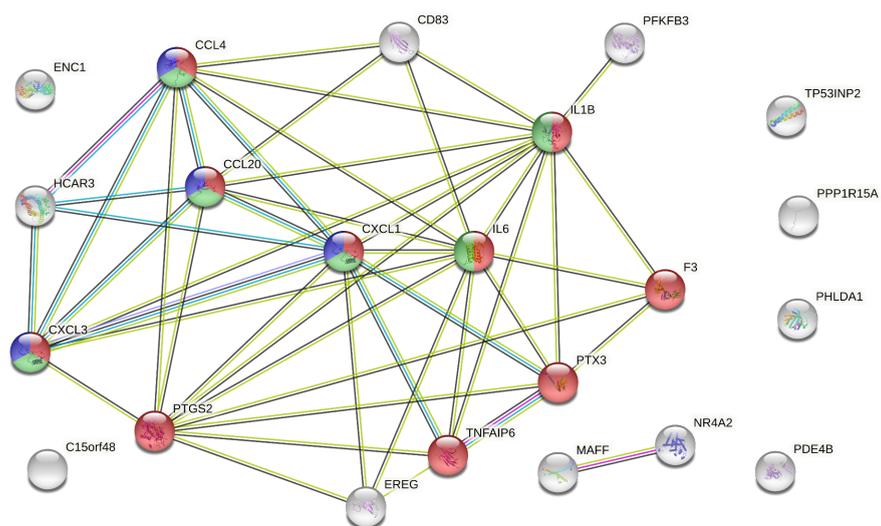


Figura 4.6. Representación en *Search Tool for the Retrieval of Interacting Genes/Proteins* (STRING) **a)** de los 25 genes más sobrerrepresentados en hombres, **b)** marcando en rojo aquellos genes con funciones relacionadas con la respuesta inflamatoria; en azul aquellos relacionados con quimiocina; y en verde aquellos relacionados con citoquina.

5. Discusión

El ACV es una patología con gran impacto en nuestra sociedad, no solo a nivel de defunciones, sino también debido a la posible discapacidad asociada (GBD 2015 Mortality and Causes of Death Collaborators, 2016). Por ello, es necesario determinar los posibles mecanismos moleculares desencadenantes del accidente. Su análisis permitirá una mejor comprensión de las causas de esta patología y su temprana detección.

No obstante, es evidente que el impacto de la enfermedad sobre los sexos es desigual. Las mujeres están más afectadas, presentando más muertes a nivel nacional (INE, 2018). Sin embargo, existen pocos estudios donde el sexo se haya incorporado en el diseño o en el análisis de sus datos. Además, las poblaciones de estudio suelen estar exclusivamente compuestas por hombres, estando significativamente subrepresentadas las mujeres (Cordonnier *et al.*, 2017). Tras la revisión sistemática se ha evidenciado que incluso aquellos estudios donde su artículo recalcan el haber utilizado una población heterogénea respecto al sexo, no indican el sexo de la muestra en los datos transcriptómicos proporcionados a la base de datos.

Son pocos los estudios donde se han abordado en mayor o menor medida las diferencias de sexo en ACV. Uno de ellos fue el trabajo presentado por Li *et al.* (2016). El estudio tenía un objetivo muy similar al presente trabajo: caracterizar las diferencias de sexo en pacientes que habían sufrido un ictus, teniendo en cuenta el logFC comparando entre dos grupos. Para ello, realizaron una revisión sistemática, parecida a la realizada en este trabajo. Únicamente pudieron encontrar dos estudios que cumplían con sus requisitos: GSE16561 y GSE22255. Aunque han pasado 5 años, la revisión sistemática de nuestro trabajo, sin incluir los estudios donde se imputó el sexo, sería coincidente con la selección realizada por Li *et al.* (2016). (*Tabla 4.2*).

Esta ausencia de información se ha intentado solventar mediante la incorporación de un estándar que homogenice la anotación de los datos. Wilkinson *et al.* (2016) describen una metodología basada en cuatro principios: *Findability, Accessibility, Interoperability, and Reusability*. Este estándar, apodado FAIR por su acrónimo en inglés, tiene como objetivo asegurar la transparencia, reproducibilidad y reutilización de los datos.

El primero es la localización de los datos. Estos deben ser registrados e indexados con un identificador único a nivel global. Con ello, también se tiene que asegurar la accesibilidad a los datos. Deben ser recuperables con su identificador, mediante un protocolo estándar. También se debe asegurar la interoperabilidad de la información, es decir, que diferentes sistemas pueden intercambiar y utilizar los datos. Para ello, se debe utilizar un lenguaje estándar y compartido, incluyendo referencias a otros datos si fuera necesario. Por último, se debería tener en cuenta la reutilización de los datos. Es necesario una descripción detallada, con todos los atributos propios de cada uno. El cumplimiento de todos estos principios permite una revisión sistemática con mejores resultados, así como su posterior metaanálisis.

Durante el desarrollo del presente trabajo, se han tenido en cuenta estos estándares FAIR, reutilizando información de otros estudios y homogeneizando los datos. Con ello, se tiene como objetivo integrar conclusiones, encontrando la información consenso entre los diferentes estudios analizados. Además, se ha facilitado la localización, accesibilidad e interoperabilidad de la información, publicando y compartiendo todo el código utilizado en GitLab.

Entre los pocos abordajes realizados sobre las diferencias transcriptómicas del ictus debidas al sexo, se encuentra el estudio llevado a cabo por Li *et al.* (2016), previamente mencionado. Este trabajo se basa en la identificación de genes que están expresados en tres comparaciones: 1) grupo control frente patológicos, 2) hombres frente a mujeres,

separando entre grupo control y patológico y 3) individuos mayores frente a jóvenes separando entre grupo control y patológico. Sin embargo, no realizan una comparación completa entre hombres y mujeres enfermos, descontando el efecto basal del grupo control. Tampoco incluye un metaanálisis de genes y sólo realizan una evaluación conjunta de dos estudios en su análisis.

En nuestro trabajo se ha optado por realizar un metaanálisis, ya que permite obtener resultados robustos, combinando diferentes estudios e incrementando el poder estadístico de los resultados (Normand, 1999). A diferencia del trabajo de Li *et al.* (2016) que realiza una intersección de genes, se ha llevado a cabo un análisis integrativo de los datos encontrados tras la revisión sistemática, que permite encontrar respuestas a preguntas no planteadas por los estudios originales.

El metaanálisis de la comparación entre los grupos enfermo y control, entre sexos, detectó, con un valor p ajustado menor que 0,05, un total de 7 genes diferencialmente expresados: 2 en hombres (*PKP4-AS1* y *TADA2B*) y 5 en mujeres (*CNIH4*, *C17orf100*, *GIMAP2*, *SMIM1* y *ZNRD1ASP*) (Tabla 4.4). Sin embargo, no existen indicaciones claras hasta el momento en la bibliografía de la implicación de estos marcadores en el ACV, presentando un novedoso interés.

El gen *PKP4-AS1*, también conocido como *PKP4 antisense RNA 1*, ha sido relacionado con *BARD1*, un gen importante en el desarrollo del cáncer de mama (Chen *et al.*, 2019). A su vez, el gen *ZNRD1ASP*, *zinc ribbon domain containing 1 antisense pseudogene*, se ha asociado a la misma patología (Dashti *et al.*, 2020). Por otro lado, el gen *CNIH4*, o *cornichon family AMPA receptor auxiliary protein 4*, ha sido asociado al cáncer de colon metastásico (Mishra *et al.*, 2019).

En cambio, *GIMAP2*, o *GTPase IMAP family member 2*, se ha relacionado con el síndrome de Behcet (Asadi, 2020). El síndrome de Behcet se caracteriza por la inflamación de los vasos sanguíneos, que puede llevar a su ensanchamiento. Aunque este gen como tal no ha sido asociado directamente al ACV, el síndrome de Behcet podría ser un factor de riesgo, al propiciar la debilitación de los vasos sanguíneos y su ensanchamiento, que podría resultar en un ictus.

C17orf100, o *chromosome 17 open reading frame 100*, ha sido asociado al síndrome nefrótico idiopático pediátrico (Kang *et al.*, 2017). Por otro lado, el gen *SMIM1*, o *small integral membrane protein 1*, es una proteína transmembrana anclada en la cola y forma fácilmente homodímeros en un sistema libre de células. Su delección provoca que el individuo no tenga la proteína Vel, siendo su grupo sanguíneo Vel - (Nylander *et al.*, 2020).

Por último, *TADA2B*, o transcriptional adaptor 2B, es un *housekeeping gene*, es decir, es un gen cuya expresión no varía y se utiliza para normalizar datos transcriptómicos (Preuss *et al.*, 2020). Este resultado nos puede indicar que *TADA2B* puede que no sea un *housekeeping gene* en sangre, siendo un mal gen de referencia.

Debido a la escasa evidencia que estos genes presentan relacionados al ACV, se optó por analizar los genes más diferencialmente expresados, sin tener en cuenta su valor p ajustado. Se estudiaron los 25 genes más expresados en mujeres (Tabla 4.5), y los 25 genes más expresados en hombres (Tabla 4.6).

Estos genes se buscaron en la base de datos *OpenTarget* (Carvalho-Silva *et al.*, 2019), bajo el término *stroke*. Esta base de datos integra información genética, genómica, transcriptómica, farmacológica, de modelos animales y basadas en la literatura científica para realizar asociaciones entre la enfermedad y posibles dianas.

En las mujeres, se encontraron 8 genes asociados a la patología: *LRRN3*, *IFI44*, *CLC*, *AHSP*, *IFIT1*, *AZU1*, *CCR2* y *TNFSF10*.

Por otro lado, en los hombres, se encontraron 16 genes asociados a la patología: *CCL4*, *IL1B*, *CCL20*, *PTGS2*, *F3*, *NR4A2*, *ENC1*, *PPP1R15A*, *PTX3*, *IL6*, *CD83*, *EREG*, *PFKFB3*, *CXCL1*, *TP53INP2* y *PDE4B*.

En el estudio por Li *et al.* (2016), también encontraron que los genes *IL6* y *IL1B* estaban sobreexpresados en hombres y baja expresión en mujeres. Estos genes forman parte de un grupo llamado *anti-stroke genes*. Por ello, es posible que las mujeres no estén más afectadas que los hombres, sino que los hombres tengan un sistema de defensa más efectivo.

Por otro lado, *PTGS2* también es considerado como un gen *anti-stroke*. Este gen no resultó significativo en el estudio de Li *et al.* (2016) comparando mujeres y hombres, pero sí que lo hizo comparando pacientes mayores respecto a los jóvenes.

El doble análisis realizado mediante STRING y PANTHER apoya esta tendencia. Los hombres presentan funciones relacionadas con la interleucina-10, que forma parte de los genes *anti-stroke*. También mostraron una regulación positiva de la respuesta neuroinflamatoria, la cual podría frenar la posible aparición del ACV. Además, como muestra la *Figura 4.5* y la *Figura 4.6*, muchos de los genes que no han tenido resultados significativos tras el metanálisis, sí que muestran interacciones con genes expresados diferencialmente. Estos genes, aunque no directamente ligados al ictus, podrían formar parte de una cascada que desencadene uno de los genes o podría ser factor de riesgo.

El conjunto de estos resultados, muestra una clara diferencia transcriptómica del ACV entre sexos, identificando un conjunto de genes asociados en hombres y en mujeres. Será necesario el desarrollo de estudios posteriores que amplíen el conocimiento del papel de estos potenciales biomarcadores y su aplicación diagnóstica. Es imprescindible que la información del sexo sea incluida y considerada en los estudios biomédicos, para mejorar y personalizar el diagnóstico y tratamiento del ictus.

6. Conclusiones

1. El metaanálisis es un método robusto para la evaluación e integración de datos provenientes de una revisión sistemática de estudios transcriptómicos, potenciando la reutilización de datos de otros estudios y cumpliendo con los criterios FAIR.
2. La revisión sistemática y selección de estudios en AVC confirma la insuficiente información del sexo en los estudios biomédicos y su necesaria inclusión para garantizar resultados de calidad que puedan ser útiles al conjunto de la población.
3. Hemos identificado 7 biomarcadores de AVC específicos por sexo: 2 en hombres y 5 en mujeres.
4. El mayor impacto de la enfermedad sobre las mujeres puede ser consecuencia de su baja expresión de genes *anti-stroke* en comparación con los hombres.
5. Las funciones más representadas en hombres son relacionadas con la respuesta inflamatoria, quimiocina o citoquina, mientras que las de las mujeres son relacionadas con el sistema inmunológico o la respuesta inmune.
6. La estrategia propuesta en este trabajo permite detectar y caracterizar diferencias debidas al sexo en ACV, incorporando la necesaria perspectiva de la medicina personalizada y de precisión en los estudios biomédicos.

7. Bibliografía

- ADAMS, H.P.; BENDIXEN, B.H.; KAPPELLE, L.J.; BILLER, J.; LOVE, B.B.; GORDON, D.L.; MARSH, E.E. (1993). Classification of subtype of acute ischemic stroke. Definitions for use in a multicenter clinical trial. TOAST. Trial of Org 10172 in Acute Stroke Treatment. *Stroke*, 24(1), 35-41. <https://doi.org/10.1161/01.STR.24.1.35>
- AHN, E.; KANG, H. (2018). Introduction to systematic review and meta-analysis. *Korean Journal of Anesthesiology*, 71(2), 103-112. <https://doi.org/10.4097/kjae.2018.71.2.103>
- ALLENDE, M.; MOLINA, E.; GURUCEAGA, E.; TAMAYO, I.; GONZÁLEZ-PORRAS, J.R.; GONZALEZ-LÓPEZ, T.J.; TOLEDO, E.; RABAL, O.; UGARTE, A.; ROLDÁN, V.; RIVERA, J.; OYARZABAL, J.; MONTES, R.; HERMIDA, J. (2016). Hsp70 protects from stroke in atrial fibrillation patients by preventing thrombosis without increased bleeding risk†. *Cardiovascular Research*, 110(3), 309-318. <https://doi.org/10.1093/cvr/cvw049>
- AMERICAN STROKE ASSOCIATION (2014). *About Stroke*, visto el 17 de Mayo de 2021, <https://www.stroke.org/en/about-stroke/>
- AMERICAN STROKE ASSOCIATION (2018a). *What is an Arteriovenous Malformation*, visto el 17 de Mayo de 2021, <https://www.stroke.org/en/about-stroke/types-of-stroke/hemorrhagic-strokes-bleeds/what-is-an-arteriovenous-malformation>
- AMERICAN STROKE ASSOCIATION (2018b). *What You Should Know About Cerebral Aneurysms*, visto el 17 de Mayo de 2021, <https://www.stroke.org/en/about-stroke/types-of-stroke/hemorrhagic-strokes-bleeds/what-you-should-know-about-cerebral-aneurysms>
- AMERICAN STROKE ASSOCIATION (2019a). *Hemorrhagic Strokes (Bleeds)*, visto el 17 de Mayo de 2021, <https://www.stroke.org/en/about-stroke/types-of-stroke/hemorrhagic-strokes-bleeds>
- AMERICAN STROKE ASSOCIATION (2019b). *Ischemic Strokes (Clots)*, visto el 17 de Mayo de 2021, <https://www.stroke.org/en/about-stroke/types-of-stroke/ischemic-stroke-clots>
- APPELROS, P.; STEGMAYR, B.; TERÉNT, A. (2009). Sex Differences in Stroke Epidemiology. *Stroke*, 40(4), 1082-1090. <https://doi.org/10.1161/STROKEAHA.108.540781>
- ASADI, S. (2020). The Role of Mutations on Genes GIMAP1, GIMAP2, GIMAP4, HLA-B27 in Behcet's Syndrome. *Journal of Genetics and Genomics*, 5, 1-6. <https://doi.org/10.24966/GGS-2485/100017>
- ATHAR, A.; FÜLLGRABE, A.; GEORGE, N.; IQBAL, H.; HUERTA, L.; ALI, A.; SNOW, C.; FONSECA, N.A.; PETRYSZAK, R.; PAPATHEODOROU, I.; SARKANS, U.; BRAZMA, A. (2019). ArrayExpress update—from bulk to single-cell expression data. *Nucleic Acids Research*, 47(D1), D711-D715. <https://doi.org/10.1093/nar/gky964>
- BARR, T.L.; CONLEY, Y.; DING, J.; DILLMAN, A.; WARACH, S.; SINGLETON, A.; MATARIN, M. (2010). Genomic biomarkers and cellular pathways of ischemic stroke by RNA gene expression profiling. *Neurology*, 75(11), 1009-1014. <https://doi.org/10.1212/WNL.0b013e3181f2b37f>
- BAYAT, A. (2002). Bioinformatics. *BMJ*, 324(7344), 1018-1022. <https://doi.org/10.1136/bmj.324.7344.1018>
- BEDIA, C. (2018). Chapter Two—Experimental Approaches in Omic Sciences, en: *Comprehensive Analytical Chemistry (Vol. 82)*, Elsevier. Amsterdam, 13-36 <https://doi.org/10.1016/bs.coac.2018.07.002>
- BENJAMINI, Y.; HOCHBERG, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1), 289-300. <http://www.jstor.org/stable/2346101>

- BERG, A.T.; BERKOVIC, S.F.; BRODIE, M.J.; BUCHHALTER, J.; CROSS, J.H.; VAN EMDE BOAS, W.; ENGEL, J.; FRENCH, J.; GLAUSER, T.A.; MATHERN, G.W.; MOSHÉ, S.L.; NORDLI, D.; PLOUIN, P.; SCHEFFER, I.E. (2010). Revised terminology and concepts for organization of seizures and epilepsies: Report of the ILAE Commission on Classification and Terminology, 2005–2009. *Epilepsia*, 51(4), 676-685. <https://doi.org/10.1111/j.1528-1167.2010.02522.x>
- CARLSON, M. (2016). *hgu133plus2.db: Affymetrix Human Genome U133 Plus 2.0 Array annotation data (chip hgu133plus2)*. R package version 3.2.3, <https://doi.org/10.18129/B9.bioc.hgu133plus2.db>
- CARVALHO-SILVA, D.; PIERLEONI, A.; PIGNATELLI, M.; ONG, C.; FUMIS, L.; KARAMANIS, N.; CARMONA, M.; FAULCONBRIDGE, A.; HERCULES, A.; MCAULEY, E.; MIRANDA, A.; PEAT, G.; SPITZER, M.; BARRETT, J.; HULCOOP, D.G.; PAPA, E.; KOSCIELNY, G.; DUNHAM, I. (2019). Open Targets Platform: New developments and updates two years on. *Nucleic Acids Research*, 47(D1), D1056-D1065. <https://doi.org/10.1093/nar/gky1133>
- CHEN, Y.Z.; ZUO, D.; REN, H.L.; FAN, S.J.; YING, G. (2019). Bioinformatics Analysis of Expression and Alterations of BARD1 in Breast Cancer. *Technology in Cancer Research & Treatment*, 18, 1533033819892260. <https://doi.org/10.1177/1533033819892260>
- CORDONNIER, C.; SPRIGG, N.; SANDSET, E.C.; PAVLOVIC, A.; SUNNERHAGEN, K.S.; CASO, V.; CHRISTENSEN, H. (2017). Stroke in women—From evidence to inequalities. *Nature Reviews Neurology*, 13(9), 521-532. <https://doi.org/10.1038/nrneurol.2017.95>
- DASHTI, S.; TAHERIAN-ESFAHANI, Z.; KHOLGHI-OSKOOEI, V.; NOROOZI, R.; ARSANG-JANG, S.; GHAFOURI-FARD, S.; TAHERI, M. (2020). In silico identification of MAPK14-related lncRNAs and assessment of their expression in breast cancer samples. *Scientific Reports*, 10(1), 8316. <https://doi.org/10.1038/s41598-020-65421-2>
- DERSIMONIAN, R.; LAIRD, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials*, 7(3), 177-188. [https://doi.org/10.1016/0197-2456\(86\)90046-2](https://doi.org/10.1016/0197-2456(86)90046-2)
- DUNNING, M.; LYNCH, A.; ELDRIDGE, M. (2015). *illuminaHumanv4.db: Illumina HumanHT12v4 annotation data (chip illuminaHumanv4)*. R package version 1.26.0, <https://doi.org/10.18129/B9.bioc.illuminaHumanv4.db>
- EDGAR, R.; DOMRACHEV, M.; LASH, A.E. (2002). Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Research*, 30(1), 207-210. <https://doi.org/10.1093/nar/30.1.207>
- ETMINAN, N.; CHANG, H.S.; HACKENBERG, K.; DE ROOIJ, N.K.; VERGOUWEN, M.D.I.; RINKEL, G.J.E.; ALGRA, A. (2019). Worldwide Incidence of Aneurysmal Subarachnoid Hemorrhage According to Region, Time Period, Blood Pressure, and Smoking Prevalence in the Population: A Systematic Review and Meta-analysis. *JAMA Neurology*, 76(5), 588-597. <https://doi.org/10.1001/jamaneurol.2019.0006>
- FEIGIN, V.L.; LAWES, C.M.; BENNETT, D.A.; BARKER-COLLO, S.L.; PARAG, V. (2009). Worldwide stroke incidence and early case fatality reported in 56 population-based studies: a systematic review. *The Lancet Neurology*, 8(4), 355-369. [https://doi.org/10.1016/S1474-4422\(09\)70025-0](https://doi.org/10.1016/S1474-4422(09)70025-0)
- FEIGIN, V.L.; NORRVING, B.; MENSAH, G.A. (2017). Global Burden of Stroke. *Circulation Research*, 120(3), 439-448. <https://doi.org/10.1161/CIRCRESAHA.116.308413>
- FRANZIN, A.; SAMBO, F.; DI CAMILLO, B. (2017). bnstruct: an R package for Bayesian Network structure learning in the presence of missing data. *Bioinformatics*, 33(8), 1250-1252. <https://doi.org/10.1093/bioinformatics/btw807>
- FUNDACIÓN ESPAÑOLA DEL CORAZÓN (FEC) (2012). *El ictus es la segunda causa de muerte en nuestro país y la primera en mujeres*, visto el 25 de Abril de 2021, <https://fundaciondelcorazon.com/prensa/notas-de-prensa/2498-ictus-es-segunda-causa-muerte-en-nuestro-pais-y-primer-a-en-mujeres.html>

- GBD 2015 MORTALITY AND CAUSES OF DEATH COLLABORATORS (2016). Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980-2015: a systematic analysis for the Global Burden of Disease Study 2015. *The Lancet*, 388(10053), 1459-1544. [https://doi.org/10.1016/S0140-6736\(16\)31012-1](https://doi.org/10.1016/S0140-6736(16)31012-1)
- GBD 2015 NEUROLOGICAL DISORDERS COLLABORATOR GROUP (2017). Global, regional, and national burden of neurological disorders during 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015. *The Lancet. Neurology*, 16(11), 877-897. [https://doi.org/10.1016/S1474-4422\(17\)30299-5](https://doi.org/10.1016/S1474-4422(17)30299-5)
- HAYNES, W.A.; VALLANIA, F.; LIU, C.; BONGEN, E.; TOMCZAK, A.; ANDRES-TERRÈ, M.; LOFGREN, S.; TAM, A.; DEISSEROTH, C.A.; LI, M.D.; SWEENEY, T.E.; KHATRI, P. (2016). Empowering Multi-Cohort Gene Expression Analysis to Increase Reproducibility. *Pacific Symposium on Biocomputing*, 22, 144-153. https://doi.org/10.1142/9789813207813_0015
- INSTITUTO NACIONAL DE ESTADÍSTICA (INE) (2018). *Infografía causas de muerte más frecuentes. Años 2002-2019*, visto el 28 de Abril de 2021, https://public.tableau.com/views/CAUSAS_DE_MUERTE/Dashboard1?:showVizHome=no&embed=true
- JENSEN, L.J.; KUHN, M.; STARK, M.; CHAFFRON, S.; CREEVEY, C.; MULLER, J.; DOERKS, T.; JULIEN, P.; ROTH, A.; SIMONOVIC, M.; BORK, P.; VON MERING, C. (2009). STRING 8—A global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Research*, 37(suppl_1), D412-D416. <https://doi.org/10.1093/nar/gkn760>
- JOSEPH, P.; LEONG, D.; MCKEE, M.; ANAND, S.S.; SCHWALM, J.D.; TEO, K.; MENTE, A.; YUSUF, S. (2017). Reducing the Global Burden of Cardiovascular Disease, Part 1: The Epidemiology and Risk Factors. *Circulation Research*, 121(6), 677-694. <https://doi.org/10.1161/CIRCRESAHA.117.308903>
- KANG, H.G.; SEO, H.; LIM, J.H.; KIM, J.I.; HAN, K.H.; PARK, H.W.; KOO, J.W.; KIM, K.H.; KIM, J.H.; CHEONG, H.I.; HA, I.S. (2017). Markers of disease and steroid responsiveness in paediatric idiopathic nephrotic syndrome: Whole-transcriptome sequencing of peripheral blood mononuclear cells. *Journal of International Medical Research*, 45(3), 948-963. <https://doi.org/10.1177/0300060516652762>
- KOELLHOFFER, E.C.; MCCULLOUGH, L.D. (2013). The Effects of Estrogen in Ischemic Stroke. *Translational Stroke Research*, 4(4), 390-401. <https://doi.org/10.1007/s12975-012-0230-5>
- KRUG, T.; GABRIEL, J.P.; TAIPA, R.; FONSECA, B.V.; DOMINGUES-MONTANARI, S.; FERNANDEZ-CADENAS, I.; MANSO, H.; GOUVEIA, L.O.; SOBRAL, J.; ALBERGARIA, I.; GASPAS, G.; JIMÉNEZ-CONDE, J.; RABIONET, R.; FERRO, J.M.; MONTANER, J.; VICENTE, A.M.; SILVA, M.R.; MATOS, I.; LOPES, G.; OLIVEIRA, S.A. (2012). TTC7B Emerges as a Novel Risk Factor for Ischemic Stroke Through the Convergence of Several Genome-Wide Approaches. *Journal of Cerebral Blood Flow & Metabolism*, 32(6), 1061-1072. <https://doi.org/10.1038/jcbfm.2012.24>
- KURKI, M.I.; HÄKKINEN, S.K.; FRÖSEN, J.; TULAMO, R.; VON UND ZU FRAUNBERG, M.; WONG, G.; TROMP, G.; NIEMELÄ, M.; HERNESNIEMI, J.; JÄÄSKELÄINEN, J.E.; YLÄ-HERTTUALA, S. (2011). Upregulated Signaling Pathways in Ruptured Human Saccular Intracranial Aneurysm Wall: An Emerging Regulative Role of Toll-Like Receptor Signaling and Nuclear Factor- κ B, Hypoxia-Inducible Factor-1A, and ETS Transcription Factors. *Neurosurgery*, 68(6), 1667-1676. <https://doi.org/10.1227/NEU.0b013e318210f001>
- LAWTON, M.T.; RUTLEDGE, W.C.; KIM, H.; STAPF, C.; WHITEHEAD, K.J.; LI, D.Y.; KRINGS, T.; TERBRUGGE, K.; KONDZIOLKA, D.; MORGAN, M.K.; MOON, K.; SPETZLER, R.F. (2015). Brain arteriovenous malformations. *Nature Reviews Disease Primers*, 1(1), 15008. <https://doi.org/10.1038/nrdp.2015.8>
- LI, W.X.; DAI, S.X.; WANG, Q.; GUO, Y.C.; HONG, Y.; ZHENG, J.J.; LIU, J.Q.; LIU, D.; LI, G.H.; HUANG, J.F. (2016). Integrated analysis of ischemic stroke datasets revealed sex and age difference in anti-stroke targets. *PeerJ*, 4, e2470. <https://doi.org/10.7717/peerj.2470>

- LIBERATI, A.; ALTMAN, D.G.; TETZLAFF, J.; MULROW, C.; GØTZSCHE, P.C.; IOANNIDIS, J.P.A.; CLARKE, M.; DEVEREAUX, P.J.; KLEIJNEN, J.; MOHER, D. (2009). The PRISMA Statement for Reporting Systematic Reviews and Meta-Analyses of Studies That Evaluate Health Care Interventions: Explanation and Elaboration. *PLOS Medicine*, 6(7), e1000100. <https://doi.org/10.1371/journal.pmed.1000100>
- LOWE, R.; SHIRLEY, N.; BLEACKLEY, M.; DOLAN, S.; SHAFEE, T. (2017). Transcriptomics technologies. *PLOS Computational Biology*, 13(5), e1005457. <https://doi.org/10.1371/journal.pcbi.1005457>
- LUSCOMBE, N.M.; GREENBAUM, D.; GERSTEIN, M. (2001). What is bioinformatics? A proposed definition and overview of the field. *Methods of Information in Medicine*, 40(4), 346-358. <https://doi.org/10.1055/s-0038-1634431>
- MAGLOTT, D.; OSTELL, J.; PRUITT, K.D.; TATUSOVA, T. (2011). Entrez Gene: Gene-centered information at NCBI. *Nucleic Acids Research*, 39(suppl_1), D52-D57. <https://doi.org/10.1093/nar/gkq1237>
- MISHRA, S.; BERNAL, C.; SILVANO, M.; ANAND, S.; RUIZ I ALTABA, A. (2019). The protein secretion modulator TMED9 drives CNIH4/TGF α /GLI signaling opposing TMED3-WNT-TCF to promote colon cancer metastases. *Oncogene*, 38(29), 5817-5837. <https://doi.org/10.1038/s41388-019-0845-z>
- MOHER, D.; COOK, D.J.; EASTWOOD, S.; OLKIN, I.; RENNIE, D.; STROUP, D.F. (1999). Improving the quality of reports of meta-analyses of randomised controlled trials: The QUOROM statement. *The Lancet*, 354(9193), 1896-1900. [https://doi.org/10.1016/S0140-6736\(99\)04149-5](https://doi.org/10.1016/S0140-6736(99)04149-5)
- NATIONAL CANCER INSTITUTE (2018). *NCI Dictionary of Cancer Terms - CAD*, visto el 5 de Mayo de 2021, <https://www.cancer.gov/espanol/publicaciones/diccionarios/diccionario-cancer/def/eac>
- NATIONAL HUMAN GENOME RESEARCH INSTITUTE (NHGRI) (2015). Tecnología de microarrays (chips de ADN o ARN), visto el 27 de Abril de 2021, <https://www.genome.gov/es/genetics-glossary/Tecnologia-de-microarrays>
- NORMAND, S.L.T. (1999). Meta-analysis: formulating, evaluating, combining, and reporting. *Statistics in Medicine*, 18(3), 321-359. [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)
- NYLANDER, A.; LEZNICKI, P.; VIDOVIC, K.; HIGH, S.; OLSSON, M.L. (2020). SMIM1, carrier of the Vel blood group, is a tail-anchored transmembrane protein and readily forms homodimers in a cell-free system. *Bioscience Reports*, 40(5). <https://doi.org/10.1042/BSR20200318>
- ORGANIZACIÓN MUNDIAL DE LA SALUD (OMS) (2015). *¿Qué son las enfermedades cardiovasculares?*, visto el 25 de Abril de 2021, https://www.who.int/cardiovascular_diseases/about_cvd/es/
- ORGANIZACIÓN MUNDIAL DE LA SALUD (OMS) (2016). *Mental health: neurological disorders*, visto el 25 de Abril de 2021, <https://www.who.int/news-room/q-a-detail/mental-health-neurological-disorders>
- PAGÈS, H.; CARLSON, M.; FALCON, S.; LI, N. (2021). *AnnotationDbi: Manipulation of SQLite-based annotations in Bioconductor*. R package version 1.54.1, <https://doi.org/10.18129/B9.bioc.AnnotationDbi>
- PETREA, R.E.; BEISER, A.S.; SESHADRI, S.; KELLY-HAYES, M.; KASE, C.S.; WOLF, P.A. (2009). Gender differences in stroke incidence and poststroke disability in the Framingham heart study. *Stroke*, 40(4), 1032-1037. <https://doi.org/10.1161/STROKEAHA.108.542894>

- PREUSS, C.; PANDEY, R.; PIAZZA, E.; FINE, A.; UYAR, A.; PERUMAL, T.; GARCEAU, D.; KOTREDES, K.P.; WILLIAMS, H.; MANGRAVITE, L.M.; LAMB, B.T.; OBLAK, A.L.; HOWELL, G.R.; SASNER, M.; LOGSDON, B.A.; CARTER, G.W.; THE MODEL-AD CONSORTIUM (2020). A novel systems biology approach to evaluate mouse models of late-onset Alzheimer's disease. *Molecular Neurodegeneration*, 15(1), 67. <https://doi.org/10.1186/s13024-020-00412-5>
- R CORE TEAM (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
- RITCHIE, M.E.; PHIPSON, B.; WU, D.; HU, Y.; LAW, C.W.; SHI, W.; SMYTH, G.K. (2015). Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*, 43(7), e47-e47. <https://doi.org/10.1093/nar/gkv007>
- RYMER, M.M. (2011). Hemorrhagic Stroke: Intracerebral Hemorrhage. *Missouri Medicine*, 108(1), 50-54. PMID: PMC6188453
- SAEED, A.; KAMPANGKAEW, J.; NAMBI, V. (2017). Prevention of Cardiovascular Disease in Women. *Methodist DeBakey cardiovascular journal*, 13(4), 185-192. <https://doi.org/10.14797/mdcj-13-4-185>
- SCHIEVINK, W.I. (1997). Intracranial Aneurysms. Intracranial aneurysms. *The New England Journal of Medicine*, 336(1), 28-40. <https://doi.org/10.1056/NEJM199701023360106>
- SOHRABJI, F.; WELSH, C.J.; REDDY, D.S. (2016). Sex differences in neurological diseases, en: *Sex differences in the central nervous system*, Academic Press. London, 297-323
- STAMOVA, B.; JICKLING, G.C.; ANDER, B.P.; ZHAN, X.; LIU, D.; TURNER, R.; HO, C.; KHOURY, J.C.; BUSHNELL, C.; PANCIOLI, A.; JAUCH, E.C.; BRODERICK, J.P.; SHARP, F.R. (2014). Gene Expression in Peripheral Immune Cells following Cardioembolic Stroke Is Sexually Dimorphic. *PLOS ONE*, 9(7), e102550. <https://doi.org/10.1371/journal.pone.0102550>
- THAKUR, K.T.; ALBANESE, E.; GIANNAKOPOULOS, P.; JETTE, N.; LINDE, M.; PRINCE, M.J.; STEINER, T. J.; DUA, T. (2016). Neurological Disorders, en: *Mental, Neurological, and Substance Use Disorders: Disease Control Priorities, 3rd Edition (Volume 4)* The International Bank for Reconstruction and Development / The World Bank. Washington DC, 87-108. <http://www.ncbi.nlm.nih.gov/books/NBK361950/>
- THOMAS, P.D.; CAMPBELL, M.J.; KEJARIWAL, A.; MI, H.; KARLAK, B.; DAVERMAN, R.; DIEMER, K.; MURUGANUJAN, A.; NARECHANIA, A. (2003). PANTHER: A Library of Protein Families and Subfamilies Indexed by Function. *Genome Research*, 13(9), 2129-2141. <https://doi.org/10.1101/gr.772403>
- TOTH, G.; CEREJO, R. (2018). Intracranial aneurysms: Review of current science and management. *Vascular Medicine*, 23(3), 276-288. <https://doi.org/10.1177/1358863X18754693>
- VIECHTBAUER, W. (2010). Conducting Meta-Analyses in R with the metafor Package. *Journal of Statistical Software*, 36(1), 1-48. <https://doi.org/10.18637/jss.v036.i03>
- WANG, Z.; GERSTEIN, M.; SNYDER, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics*, 10(1), 57-63. <https://doi.org/10.1038/nrg2484>
- WILKINSON, M.; DUMONTIER, M.; AALBERSBERG, I.; APPLETON, G.; AXTON, M.; BAAK, A.; BLOMBERG, N.; BOITEN, J.; DA SILVA SANTOS, L.B.; BOURNE, P.E.; BOUWMAN, J.; BROOKES, A.J.; CLARK, T.; CROSAS, M.; DILLO, I.; DUMON, O.; EDMUNDS, S.; EVELO, C. T.; FINKERS, R.; GONZALEZ-BELTRAN, A.; GRAY, A.; GROTH, P.; GOBLE, C.; GRETHE, J.S.; HERINGA, J.; HOEN, P.; HOOFT, R.; KUHN, T.; KOK, R.; KOK, J.; LUSHER, S.J.; MARTONE, M.E.; MONS, A.; PACKER, A.L.; PERSSON, B.; ROCCA-SERRA, P.; ROOS, M.; VAN SCHAIK, R.; SANSONE, S.A.; SCHULTES, E.; SENGSTAG, T.; SLATER, T.; STRAWN, G.; SWERTZ, M.A.; THOMPSON, M.; VAN DER LEI, J.; VAN MULLIGEN, E.; VELTEROP, J.; WAAGMEESTER, A.; WITTENBURG, P.; WOLSTENCROFT, K.; ZHAO, J.; MONS, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3(1), 160018. <https://doi.org/10.1038/sdata.2016.18>