



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA



TRABAJO FINAL DE GRADO

**Universitat Politècnica de València – Escola Tècnica Superior
d'Enginyeria Agronòmica i del Medi Natural
Grado en Biotecnología
Curso académico: 2019/2020**

Caracterización de las diferencias de sexo en la enfermedad de Parkinson mediante el metaanálisis de estudios transcriptómicos

Alumna: Violeta Lillo Collado
Tutora: Dña. Esther Giraldo Reboloso
Cotutor: D. Francisco García García

Valencia, Junio 2020

CHARACTERIZATION OF SEX DIFFERENCES IN PARKINSON DISEASE BY META-ANALYSIS OF TRANSCRIPTOMIC STUDIES

Abstract:

Parkinson's disease (PD) is one of the neurological disorders most associated with aging. The incidence of this disease is higher in men than in women and the symptoms vary between both sexes. The presence of motor symptoms is strongly associated with male sex, while dyskinesia and depression occur more frequently in women.

The effect of estrogen on dopaminergic neurons may be a possible source of differences between men and women in the clinical and cognitive characteristics of PD.

Elucidating the molecular mechanisms responsible of these differences could help to find new biomarkers that may enhance the accuracy and effectiveness of clinical evaluation and treatment of the disease.

With this aim, we have carried-out a meta-analysis with specific data from selected genetic expression studies to determine the possible causes of the differences between sexes.

Keywords: Parkinson disease, sex, transcriptomic, meta-analysis, personalized medicine.

CARACTERIZACIÓN DE LAS DIFERENCIAS DE SEXO EN LA ENFERMEDAD DE PARKINSON MEDIANTE EL METAANÁLISIS DE ESTUDIOS TRANSCRIPTÓMICOS

Resumen:

La enfermedad de Parkinson (EP) es uno de los trastornos neurológicos más asociados al envejecimiento. La incidencia de esta enfermedad es mayor en hombres que en mujeres y los síntomas varían entre ambos sexos. La presencia de síntomas motores está fuertemente asociada en hombres, mientras que la discinesia y la depresión se presenta con mayor frecuencia en mujeres.

El efecto del estrógeno sobre las neuronas dopaminérgicas puede ser una posible fuente de diferencias entre hombres y mujeres en cuanto a las características clínicas y cognitivas de la EP.

Elucidar los mecanismos moleculares responsables de estas diferencias podría ayudar a encontrar nuevos biomarcadores que permitan mejorar la precisión y la eficacia en la evaluación clínica y el tratamiento de la enfermedad.

Con este objetivo, hemos realizado un metaanálisis con datos procedentes de estudios seleccionados de expresión génica, para determinar las posibles causas de las diferencias entre ambos sexos.

Palabras clave: Enfermedad de Parkinson, sexo, transcriptómica, metaanálisis, medicina personalizada.

Autora del TFG: Violeta Lillo Collado

Tutora Académica: Dña. Esther Giraldo Reboloso

Cotutor: D. Francisco García García

Valencia, Junio de 2020

AGRADECIMIENTOS

En primer lugar, me gustaría agradecer al Dr. Francisco García-García, jefe de la Unidad de Bioinformática y Bioestadística del CIPF, por haberme hecho un hueco en su grupo, por haber confiado en mí para llevar a cabo este trabajo y por el tiempo dedicado a pesar del obstáculo que ha supuesto la pandemia del SARS-CoV-2.

También me gustaría agradecer a la Dra. Deborah Burks directora del CIPF y jefa del grupo de investigación de Neuroendocrinología Molecular, porque a pesar de no haber podido trabajar en su grupo por problemas técnicos, fue quien me aceptó y confió en mí en primera instancia. Gracias también a Esperanza Irles, investigadora postdoctoral en el grupo de Deborah, por guiarme en lo que inicialmente iba a ser mi TFG.

Gracias a Esther Giraldo, investigadora en el grupo de Regeneración Tisular y Neuronal en el CIPF, por aceptar ser mi tutora UPV e interesarse en mi trabajo.

Gracias a Belén, María, Ángela, Laura y Alba, por acompañarme a lo largo de estos años y por haberse convertido en mucho más que compañeras de grado.

Por último, gracias a mi familia, en especial a mis padres, por apoyarme y ayudarme siempre en las decisiones tomadas a lo largo de mi vida.

ÍNDICE DE CONTENIDOS

Índice de figuras	II
Índice de tablas	III
Índice de abreviaturas	III
1. Introducción	1
1.1. Enfermedad de Parkinson	1
1.1.1. Epidemiología y diferencias entre sexos	1
1.1.2. Bases moleculares de la enfermedad de Parkinson	2
1.1.3. Alternativas terapéuticas de la enfermedad de Parkinson	2
1.2. Datos ómicos y tecnologías de alto rendimiento	3
1.2.1. Transcriptómica	4
1.2.2. Análisis transcriptómicos	5
1.2.3. Caracterización funcional	5
1.2.4. Metaanálisis funcional	6
2. Objetivos	7
3. Material y métodos	7
3.1. Revisión sistemática y selección de estudios	7
3.2. <i>MetaFun</i> : herramienta web para la caracterización molecular de las diferencias por sexo	9
3.3. Análisis individual de los estudios	9
3.3.1. Procesamiento de los datos	9
3.3.2. Análisis de expresión diferencial	10
3.3.3. Análisis de enriquecimiento funcional	10
3.4. Metaanálisis	10
4. Resultados y discusión	11
4.1. Revisión sistemática y selección de estudios	11
4.2. Análisis individual de los estudios	12
4.2.1. Procesamiento de los datos	12
4.2.2. Análisis de expresión diferencial	21
4.2.3. Análisis de enriquecimiento funcional	22
4.3. Metaanálisis	24
4.4. Discusión	28
5. Conclusiones	30
6. Referencias	31

ÍNDICE DE FIGURAS

Figura 1: Dogma central de la biología molecular	4
Figura 2: Flujo de información del proceso de revisión sistemática	8
Figura 3: Diagrama de flujo de la revisión sistemática realizada en este trabajo	11
Figura 4: Número de individuos en cada estudio. Distribuidos según sexo y estado	12
Figura 5: Diagrama de cajas en el que están representados los valores normalizados las muestras empleadas en el estudio GSE7621	15
Figura 6: Análisis de componentes principales (PCA) de las muestras del estudio GSE7621	16
Figura 7: Clustering de las muestras del estudio GSE7621. Cada grupo experimental aparece representado de un color	16
Figura 8: Diagrama de cajas en el que están representados los valores normalizados las muestras empleadas en el estudio GSE20153	17
Figura 9: Análisis de componentes principales (PCA) de las muestras del estudio GSE20153	17
Figura 10: Clustering de las muestras del estudio GSE20153. Cada grupo experimental aparece representado de un color	18
Figura 11: Diagrama de cajas en el que están representados los valores normalizados las muestras empleadas en el estudio GSE20292	18
Figura 12: Análisis de componentes principales (PCA) de las muestras del estudio GSE20292	19
Figura 13: Clustering de las muestras del estudio GSE20292. Cada grupo experimental aparece representado de un color	19
Figura 14: Diagrama de cajas en el que están representados los valores normalizados las muestras empleadas en el estudio GSE128177	20
Figura 15: Análisis de componentes principales (PCA) de las muestras del estudio GSE128177	20
Figura 16: Clustering de las muestras del estudio GSE128177	21
Figura 17: Gráfico de bosque para el término GO:0022904	26
Figura 18: Gráfico de embudo para el término GO:0022904	26
Figura 19: Gráfico de bosque para el término GO:0002250	27
Figura 20: Gráfico de embudo para el término GO:0002250	27

ÍNDICE DE TABLAS

Tabla 1: Estudios finalmente seleccionados tras la revisión sistemática	12
Tabla 2: Individuos incluidos en el estudio GSE7621	13
Tabla 3: Individuos incluidos en el estudio GSE20153	13
Tabla 4: Individuos incluidos en el estudio GSE20292	14
Tabla 5: Individuos incluidos en el estudio GSE128177	14
Tabla 6: Número de genes cuya diferencia de expresión es significativa	21
Tabla 7: Resultados de expresión diferencial del estudio GSE7621	22
Tabla 8: Número de funciones cuya diferencia de expresión es significativa	23
Tabla 9: Resultados del análisis de enriquecimiento funcional del estudio GSE7621 para la categoría “Procesos biológicos” de GO	23
Tabla 10: Resultados del metaanálisis	24

ÍNDICE DE ABREVIATURAS

EP: Enfermedad de Parkinson
GEO: Gene Expression Omnibus
GO: Gene Ontology
GSEA: Gene Set Enrichment Analysis
NGS: Next generation sequencing
PCA: Análisis de componentes principales
RNA-seq: Secuenciación de ARN
RS: Revisión sistemática

1. INTRODUCCIÓN

1.1 ENFERMEDAD DE PARKINSON

La enfermedad de Parkinson (EP) es uno de los trastornos neurológicos más asociados al envejecimiento. Se caracteriza por temblores, bradicinesia y rigidez muscular. Además, la mitad de los pacientes presentan déficits de atención, perturbación verbal y reducción de la velocidad de procesamiento verbal.

Es una enfermedad asociada a los cuerpos de Lewis y a la pérdida neuronal dopaminérgica en la sustancia nigra pars compacta (SNpc). Esta pérdida y el agotamiento de los niveles de dopamina (DA) en el cuerpo estriado representan la patología distintiva de la EP (Maiti et al. 2017).

Su sintomatología es muy heterogénea, lo que aumenta su complejidad involucrando numerosas regiones del sistema nervioso y varios neurotransmisores y proteínas.

1.1.1. Epidemiología y diferencias entre sexos

En el primer mundo, la prevalencia e incidencia de las enfermedades neurológicas ha aumentado considerablemente durante los últimos años. Uno de los motivos fundamentales es el aumento de la esperanza de vida acompañado de mejoras en el tratamiento de enfermedades y en la educación sanitaria.

En un estudio epidemiológico, la prevalencia y la incidencia son los indicadores más precisos para informar sobre el riesgo y las posibles relaciones que existen entre factores a los que se exponen las personas y las enfermedades que presentan.

La incidencia es una medida del número de casos nuevos de una enfermedad durante un periodo de tiempo determinado, mientras que la prevalencia es la proporción de individuos de una población que presentan una característica durante un periodo de tiempo determinado. La prevalencia depende de la incidencia y de la supervivencia. Esta última varía según las condiciones sanitarias y económicas de la población en estudio. Por ello, en el caso concreto de la EP, es difícil comparar la prevalencia en poblaciones donde existen notables diferencias económicas y sociales.

La EP tiene una prevalencia de entre el 0,3% y el 1% en mayores de 60 años, que asciende al 3% en mayores de 80 años. La incidencia varía entre 0,08 y 0,18 por 1000 personas/año (Benito, 2018).

La causa de la EP sigue siendo desconocida, pero parece ser resultado de una interacción compleja de factores genéticos y ambientales que afectan a gran parte de procesos celulares fundamentales. Además, el diagnóstico en estadios tempranos de la enfermedad es muy complejo, así como el tratamiento de etapas más avanzadas. A día de hoy tampoco existe un tratamiento efectivo que frene el daño neurodegenerativo (Kalia y Lang, 2015).

A todas las complicaciones que tiene el diagnóstico y tratamiento de esta enfermedad hay que sumarle que existen varias diferencias de sexo. La presencia de síntomas motores está fuertemente asociada con el sexo masculino. Según datos anteriores y un metaanálisis de 17 estudios relevantes, la incidencia en hombres es un 46% mayor que en mujeres, y la prevalencia es hasta 3 veces mayor en hombres que en mujeres entre

los 50 y 59 años (Picillo et al., 2017). Los síntomas motores, la edad de aparición de la enfermedad, temblores y pérdida de receptores de dopamina parecen ser ligeramente más leves en mujeres, posiblemente debido a la protección de los estrógenos. Sin embargo, una vez comienza la enfermedad, en las mujeres hay síntomas que se desarrollan más rápido, como es el caso de las fluctuaciones motoras y movimientos involuntarios (discinesia), que en estadios avanzados suelen ser peores que en hombres. Los síntomas no motores siguen un patrón similar: se presentan más tarde en mujeres, pero empeoran con el progreso de la enfermedad. En cualquier caso, se desarrollan de forma diferente y cada sexo presenta un cuadro distinto: entre mujeres hay más casos de ansiedad, mientras que en hombres prevalece una disminución de los sentidos del olfato y el gusto (Nicoletti et al., 2017).

1.1.2. Bases moleculares de la enfermedad de Parkinson

La EP es una enfermedad multifactorial cuya detección temprana es muy complicada, principalmente porque no se han establecido biomarcadores que puedan estar relacionados con su desarrollo. Factores tanto genéticos como no genéticos están involucrados en el progreso de la EP (Maiti et al. 2017).

Sin embargo, cabe destacar el urato como posible biomarcador. Es un metabolito y un potente antioxidante que ha sido el mejor descrito y estudiado. Si su concentración disminuye, se relaciona con un mayor riesgo de desarrollar EP en hombres, pero en mujeres no se ha detectado asociación (Schwarzschild et al., 2008), (Ascherio et al, 2009), (Jesús et al., 2012).

La detección de patrones genéticos también puede ayudar a la detección de la enfermedad. Las mutaciones en el gen LRRK2 son una causa común de la enfermedad de Parkinson, sin existir diferencias entre hombres y mujeres. Como se ha indicado en modelos animales, estas mutaciones afectan al tráfico vesicular, a la autofagia, a la síntesis de proteínas y a la función del citoesqueleto (Martin et al., 2016). Sí que se ha observado que, en individuos portadores de mutaciones en este gen en los que no se manifiesta la enfermedad, los niveles de urato eran mayores que en aquellos que sí desarrollaban EP. Estos hallazgos corroboran el urato como un biomarcador de resistencia a la EP entre los individuos portadores de la mutación en el gen LRRK2 (Bakshi et al., 2019).

1.1.3. Alternativas terapéuticas de la enfermedad del Parkinson

A pesar de las diferencias que existen en cómo afecta y evoluciona la enfermedad entre hombres y mujeres, hoy en día a ambos sexos se les aplica el mismo tratamiento.

El más común es la administración de levodopa, la cual a largo plazo parece llevar a otras complicaciones motoras (Fabbrini et al., 2007). Este medicamento es un precursor de la dopamina que, a diferencia de ella, es capaz de atravesar la barrera hematoencefálica y ser utilizada por las células nerviosas para fabricar dopamina. Generalmente, la levodopa se administra con carbidopa, un inhibidor periférico de la dopadecarboxilasa (DDC), que retrasa la conversión de levodopa a dopamina hasta que llega al cerebro. De esta manera, la dosis de levodopa administrada puede ser menor, reduciendo así los efectos secundarios.

La levodopa retrasa el inicio de los síntomas de la EP, pero no todos ellos responden igual al medicamento. La bradicinesia y la rigidez suelen mejorar mientras que, el

temblor y el equilibrio pueden no aliviarse en absoluto. Además, la administración de dosis elevadas de levodopa durante largos periodos de tiempo tiene como consecuencia un aumento de los movimientos involuntarios o discinesia. Por ello es importante buscar un equilibrio entre los beneficios y los efectos secundarios propios de este tratamiento (Carretero Colomer, 2006).

Se sugiere que las diferencias epidemiológicas de sexo en la EP pueden deberse en parte a la participación de los estrógenos. Las mujeres tienen una biodisponibilidad significativamente mayor de levodopa ya que la metabolizan más lentamente (Kumagai et al., 2014).

Cuando no hay un alivio significativo después de la medicación, se recurre a tratamientos quirúrgicos para reducir los síntomas, especialmente en etapas avanzadas de la EP, como es la estimulación cerebral profunda. Ésta se ha mostrado efectiva tanto en hombres como en mujeres. Sin embargo, en mujeres este procedimiento es menos frecuente y sólo en etapas muy avanzadas (Hariz et al., 2003).

Es necesario elucidar las causas de estas diferencias, lo cual ayudará a encontrar nuevos biomarcadores que permitan mejorar la detección y el posterior tratamiento, llevándonos más cerca de la medicina personalizada. Además de metaanálisis, la realización de estudios de neuroimagen y de microbiota intestinal, la cual muestra alteraciones en pacientes (Miraglia y Colla, 2019), pueden permitirnos establecer mejor las diferencias entre sexos para la EP.

1.2. DATOS ÓMICOS Y TECNOLOGÍAS DE ALTO RENDIMIENTO

La información de la materia viva está almacenada en moléculas de ADN que consisten en dos cadenas antiparalelas, cada una de ellas formada por uniones covalentes de un azúcar (desoxirribosa) y una molécula de fosfato. Cada azúcar está unido a una base nitrogenada, adenina (A), timina (T), guanina (G) o citosina (C).

Este ADN no es el molde directo para la síntesis de proteínas, primero es necesario pasar por el proceso de transcripción para generar moléculas de ARN y estas sí, actuarán como molde. Esta molécula difiere en tres aspectos fundamentales de la molécula de ADN. El azúcar del ARN es una ribosa en lugar de desoxirribosa, una de las cuatro bases de la ribosa es el uracilo (U) en lugar de la timina (T) y por último la molécula de ARN es de una sola cadena.

Existen muchos tipos de ARN, pero es el ARN mensajero (ARNm) el que actúa como molde para la síntesis de proteínas, que una vez transcrito sale del núcleo celular y se dirige a los ribosomas donde tiene lugar el proceso de traducción. Durante este proceso se forma una cadena de aminoácidos a partir de la cual se sintetiza una proteína que será la encargada de realizar la función que corresponda.

Todo este proceso junto con la replicación y la transcripción reversa queda resumido en el conocido como dogma central de la biología molecular (Figura 1).

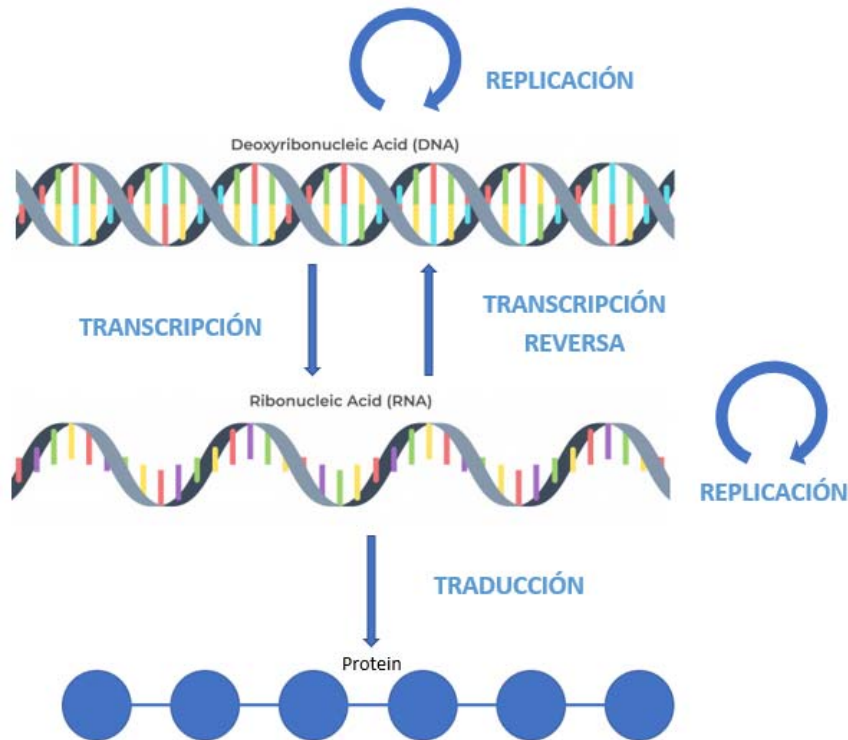


Figura 1: Dogma central de la biología molecular. Elaboración propia.

En los últimos años ha habido una mejora considerable en los avances tecnológicos. Esto nos ha permitido posicionar la biología como una ciencia de la información gracias al estudio intensivo de la naturaleza. Destacan las disciplinas ómicas, que estudian conjuntos biomoleculares analizados mediante procedimientos de alto rendimiento (Barrera et al., 2018).

Entre las principales tecnologías ómicas incluimos la genómica, transcriptómica, proteómica y metabolómica. Con el desarrollo de nuevas tecnologías, los estudios ómicos son cada vez más frecuentes, pero para tener éxito en ellos, es crucial un diseño correcto de la parte experimental, así como tener en cuenta factores como el tamaño de las muestras entre otros (Bedia, 2018).

Debido a la mejora de estas tecnologías de alto rendimiento, se ha generado una gran cantidad de datos biológicos. A raíz de ello, se ha incrementado la necesidad de tener a disposición bases de datos donde poder almacenarlos, clasificarlos y ordenarlos.

1.2.1. Transcriptómica

La transcriptómica es el área que estudia el transcriptoma, es decir, el conjunto de transcritos de ARN que existe en la célula, tanto los codificantes como los no codificantes. El estudio de estos transcritos nos permite obtener información sobre las características de expresión génica específicas de células y tejidos para poder comprender mejor la dinámica del metabolismo celular y tisular. Esta información es fundamental para relacionar cambios en los perfiles del transcriptoma con problemas de salud. Gracias a los avances que se han conseguido en las tecnologías de alto rendimiento, como son los microarrays y RNA-seq, la transcriptómica se ha podido desarrollar en numerosas áreas de estudio (Manzoni et al., 2016).

Los ensayos de hibridación en microarrays consisten en un sustrato al que se une ADN en posiciones conocidas. El ARN que obtenemos de la muestra a estudiar, se convertirá

en ADN complementario (ADNc) y se marcará con una molécula fluorescente o radioactiva para a continuación ponerlo en contacto con el array, hibridando en las posiciones en que se complementen con las sondas. Una vez se ha producido la hibridación, el microarray se lee con un escáner láser y se obtienen unas imágenes que se procesan mediante un software para elaborar una base de datos (Cigudosa, 2004), (Moreno y Solé, 2004).

La secuenciación de ARN es una técnica que permite examinar la cantidad y las secuencias de ARN en una muestra mediante NGS. Consiste en la conversión de ARN a ADNc mediante transcripción inversa. Esta técnica permite estudiar el contenido celular total de ARN, incluidos ARNm, ARNr y ARNt de manera cuantitativa y con alto rendimiento. Esto aporta información vital sobre la expresión de los genes evaluando los cambios asociados a una enfermedad (Wang et al., 2009), (Mackenzie, 2018).

1.2.2. Análisis transcriptómicos

Debido a los rápidos avances en las tecnologías de NGS, el análisis de los datos ómicos es un campo de continua evolución en el que constantemente se están desarrollando nuevos métodos. Por tanto, la estrategia de análisis se tiene que ir adaptando al tipo de datos que se van a analizar y al objetivo final del estudio (Henry et al., 2014).

En los análisis transcriptómicos se requiere de un pre-procesamiento de datos específico según la tecnología que se vaya a utilizar. Como resultado, obtenemos una matriz de expresión con la cuantificación de las unidades biológicas de interés de cada una de las muestras (García, 2016). Algunas de las estrategias para las cuales vamos a hacer uso de esta matriz como punto de partida son:

- Análisis de expresión diferencial: se emplea para encontrar genes que se expresan a diferentes niveles en varias condiciones experimentales.
- Descubrimiento de clases o análisis clúster: se emplea cuando se cree que puede haber subtipos interesantes de genes que no se conocen, pero que pueden ser revelados por firmas de expresión. Asimismo, también permite la detección de subgrupos de individuos de interés con un perfil transcriptómico común.
- Predicción de clases: conjunto de perfiles de entrenamiento con una clasificación conocida y cuyo objetivo es generar un predictor (también conocido como clasificador) que se utilizará para clasificar nuevos perfiles (Grant et al., 2007).

1.2.3. Caracterización funcional

A partir del conjunto de genes resultante de los análisis realizados de los experimentos, nos interesa saber si este conjunto muestra o no enriquecimiento de una propiedad biológica predefinida. Dado que a partir del análisis de expresión diferencial podemos obtener una lista de elementos biológicos de interés, estos los podemos combinar con información biológica de bases de datos para facilitar su interpretación. Algunas de las bases de datos que contienen esta información son:

- The Gene Ontology (GO): proporciona un vocabulario dinámico y controlado para tres categorías: procesos biológicos, componentes celulares y funciones moleculares (Grant et al., 2007), (Ashburner et al., 2000).

- Kyoto Encyclopedia of Genes and Genomes (KEGG): descripción de rutas de señalización asociadas a genes y metabolitos (Kanehisa, 2000).

Para realizar la combinación de los datos obtenidos y los procedentes de las bases de datos utilizadas, llevamos a cabo un análisis funcional de los genes que consiste en identificar, mediante análisis estadísticos, anotaciones funcionales (procesos biológicos, rutas de señalización...) con las que los genes analizados están significativamente relacionados (SERVICIO DE BIOINFORMÁTICA DE NUCLEUS, 2020).

- **Over-representation analysis (ORA):** se trata de un enfoque empleado para determinar si una serie de funciones biológicas están sobrerrepresentando o enriqueciendo funcionalmente a un grupo de genes específicos (García, 2016).
- **Gene set enrichment analysis (GSEA):** este procedimiento considera todos los genes que participan en un experimento (y no sólo los significativos como en ORA), ordenados por alguna variable de interés, como puede ser el nivel de expresión diferencial, el número de mutaciones de cada gen, etc. A continuación, para cada función, los GSEA buscan grupos de genes con un patrón común de expresión (u otra variable biológica), que además participen en una misma función biológica. De esta forma, caracterizaremos funcionalmente la lista de genes de partida con la información biológica de la base de datos que hayamos seleccionado (Korotkevich et al., 2016), (Hung et al., 2011).
- **Integrative and modular enrichment analysis (MEA):** es la única clase que emplea relaciones que puede haber entre diferentes anotaciones durante el enriquecimiento. Puede reducir la redundancia y evitar ignorar conceptos biológicos importantes (Tipney y Hunter, 2010).

1.2.4. Metaanálisis funcional

El metaanálisis es un conjunto de técnicas cuyo objetivo consiste en combinar evidencias de diferentes estudios de una misma temática en uno solo (Hoffman, 2015).

Los objetivos específicos del metaanálisis son:

- Aumentar el poder estadístico.
- Tratar la controversia cuando los estudios individuales no están de acuerdo.
- Mejorar las estimaciones del tamaño del efecto.
- Responder nuevas preguntas no planteadas previamente en estudios de componentes.

Los resultados del metaanálisis pueden ser cualitativos y cuantitativos, permitiendo revelar fortalezas y debilidades de otros estudios. Además, estos resultados proporcionan mejores recomendaciones de tratamiento y orientan en el diseño de posibles ensayos clínicos (Russo, 2007).

Aunque la mayoría de los métodos de metaanálisis están centrados a nivel genómico, dirigir esta metodología a nivel de función nos permite realizar una interpretación de los resultados de estudios genómicos más acertada, ya que los resultados de estudios individuales pueden presentar incongruencias debido a problemáticas como un reducido tamaño muestral o la heterogeneidad en el diseño de los estudios (García, 2016), (Russo, 2007).

2. OBJETIVOS

El objetivo general de este trabajo es la detección y comprensión de los mecanismos moleculares que afectan de forma diferencial a hombres y mujeres en la EP. Indicadores clínicos y epidemiológicos han demostrado, en ocasiones, que no se conoce bien a que se deben estas diferencias. Caracterizando mejor estos mecanismos, podríamos ser más precisos en el diagnóstico de la enfermedad y en la selección de un tratamiento según el grupo de pacientes.

Para alcanzar este objetivo general, requerimos del cumplimiento de los siguientes objetivos específicos:

1. Revisión sistemática y selección de estudios transcriptómicos de la enfermedad de Parkinson con datos de microarrays o RNA-seq en los repositorios Gene Expression Omnibus (GEO) y ArrayExpress.
2. Análisis individual de cada estudio seleccionado, incluyendo un análisis de expresión diferencial y un análisis de enriquecimiento funcional.
3. Metaanálisis de genes y de los resultados funcionales que permitirá la detección de biomarcadores específicos por sexo.

3. MATERIALES Y MÉTODOS

3.1. REVISIÓN SISTEMÁTICA Y SELECCIÓN DE ESTUDIOS

Las revisiones sistemáticas (RS) son investigaciones en las cuales se realiza un análisis de estudios originales primarios. Las RS sintetizan resultados obtenidos de distintas investigaciones de manera que se limita el error debido a la aleatoriedad. Para realizar este proceso de selección, se llevan a cabo una serie de estrategias ejecutadas de manera rigurosa y objetiva con el fin de obtener un diseño de investigación eficiente que permita incrementar el poder, precisión, consistencia y generalización de los resultados y de la estimación. Para que esta metodología sea útil, se ha de ser muy cauteloso en la elección de los estudios y en la interpretación de los resultados debido, entre otros motivos, a la heterogeneidad de los estudios (Ferreira et al., 2011), (Manterola et al., 2013). En este trabajo seguimos el diagrama de flujo del proceso de revisión sistemática que se muestra en la Figura 2.

Criterios de inclusión y exclusión

En primer lugar, se llevó a cabo una búsqueda exhaustiva en las bases de datos públicas GEO y ArrayExpress, durante marzo y abril de 2020. Para ello comenzamos buscando con la palabra clave *Parkinson* y a partir de los estudios obtenidos se realizó una selección rigurosa de los mismos.

Dado que buscamos datos de expresión génica, otro de los criterios establecidos para seleccionar el tipo de tecnología, fue la búsqueda de estudios a partir de las palabras clave *expression profiling by array* y *expression profiling by high throughput sequencing*.

De los estudios de expresión génica realizados mediante arrays y secuenciación de alto rendimiento, se excluyeron aquellos que no estudian específicamente la enfermedad de Parkinson. También fueron excluidos aquellos que no disponían de información sobre el sexo y la edad de los pacientes de los que se obtenían las muestras. Otro criterio de

exclusión se basó en el tamaño muestral: los estudios con un número de muestras menor de 12, fueron excluidos de la selección.

Los estudios que finalmente se incluyeron en el análisis proceden de la base de datos GEO.

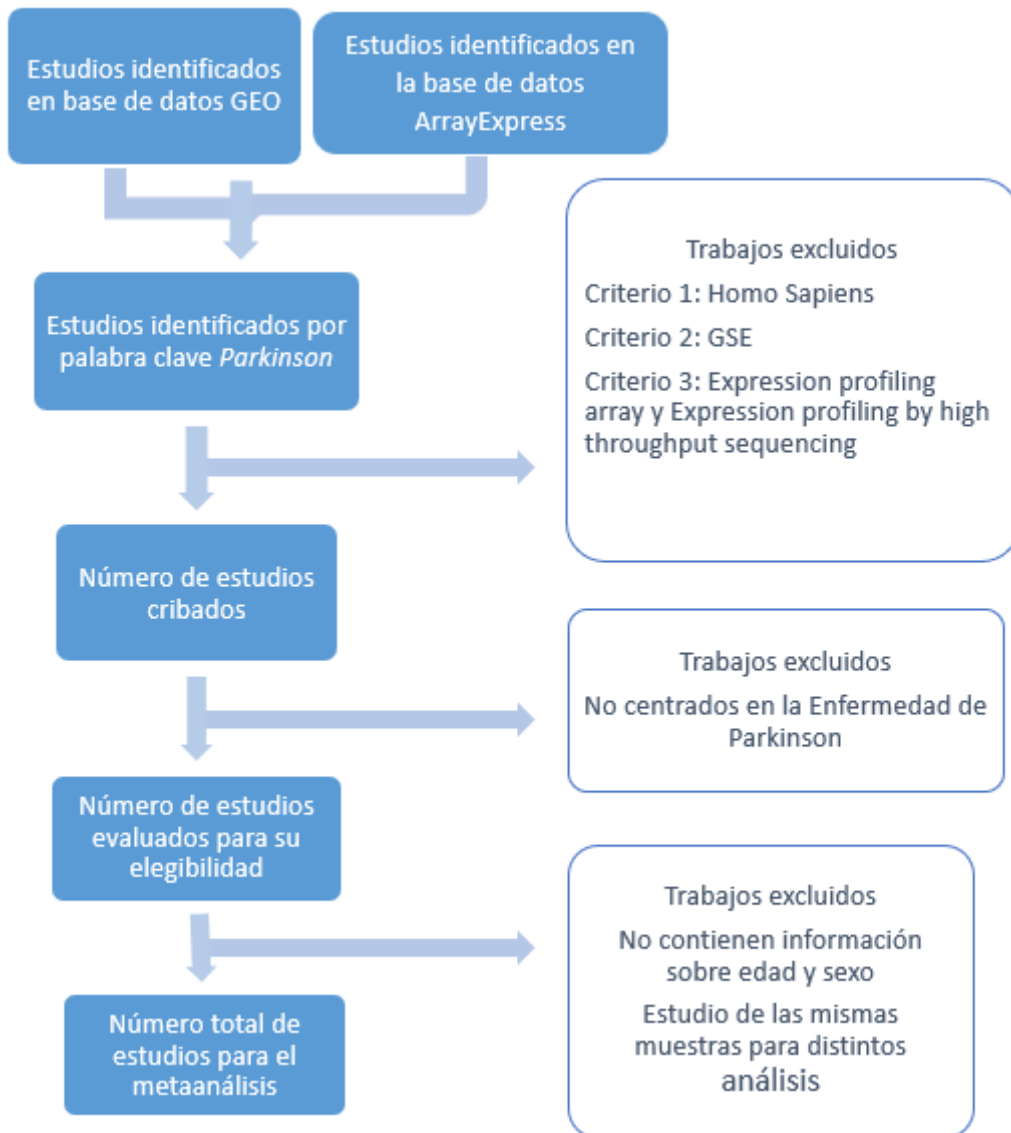


Figura 2: Flujo de información del proceso de revisión sistemática. Adaptado de (Rodrigo Ruíz, Pérez González & Cejudo, 2017).

3.2. METAFUN: HERRAMIENTA WEB PARA LA CARACTERIZACIÓN MOLECULAR DE LAS DIFERENCIAS POR SEXO

Para realizar la caracterización de las diferencias de sexo en la EP, hemos hecho uso de la herramienta web *MetaFun* “cuyo objetivo es proporcionar un recurso bioinformático a la comunidad científica que permita una mejor comprensión de los resultados obtenidos en la integración de diversos estudios, con técnicas de metaanálisis” (Malmierca, 2019).

Esta herramienta realiza un metaanálisis, tomando como unidad de análisis la función que lleva a cabo el gen, en lugar del propio gen. Gracias a su sencillez en la utilización, esta herramienta está al alcance de cualquier investigador sin experiencia en el campo de la bioinformática.

En los apartados 3.3 y 3.4 se explican los análisis desarrollados por esta herramienta entre los que se encuentran: análisis exploratorio, análisis de expresión diferencial, análisis de enriquecimiento funcional y metaanálisis.

3.3. ANÁLISIS INDIVIDUAL DE LOS ESTUDIOS

3.3.1. Procesamiento de los datos

ESTANDARIZACIÓN NOMENCLATURA

Para evitar problemas en el metaanálisis debido a las diferencias en los diseños experimentales de los diferentes estudios, el primer paso fue estandarizar la nomenclatura de los datos.

Para ello se tuvieron en cuenta dos variables:

- Variable “sexo”: Clasifica a los individuos como “Hombre” o “Mujer”.
- Variable “estado”: Clasifica a los individuos como “Enfermo”, para aquellos que sufren EP en cualquier estadio y como “Sano” para los individuos que no sufren EP y por tanto actúan como controles experimentales.

Así pues, generamos 4 grupos: HS (Hombre sano), HE (Hombre enfermo), MS (Mujer sana), ME (Mujer enferma).

ESTANDARIZACIÓN ANOTACIÓN DE LAS SONDAS

Dado que los estudios seleccionados para el análisis fueron realizados en diferentes plataformas tecnológicas, se estandarizó la anotación de las sondas de los microarrays a un código común, el código ENTREZID que sirve como identificador de los genes y secuencias en la base de datos del National Center for Biotechnology Information (NCBI).

ANÁLISIS EXPLORATORIO

MetaFun emplea los diagramas de cajas (boxplot), los análisis de componentes principales (PCA) y los análisis clúster de las muestras para cada uno de los estudios a analizar.

En *MetaFun*, los análisis clúster de las muestras se utilizan complementariamente con los de PCA. De esta manera, en caso de tener que eliminar alguna muestra considerada anómala, es posible comparar ambos análisis para tener una mayor seguridad de que dicha anomalía no es consecuencia del tipo de análisis utilizado.

En resumen, con este tipo de análisis se pretende comprobar la variabilidad y agrupación de las muestras, pudiendo detectar patrones anómalos.

3.3.2. Análisis de expresión diferencial

Con el análisis de expresión diferencial buscamos encontrar diferencias significativas en la expresión de los genes de interés entre ambos sexos.

Con los resultados obtenidos, conoceremos qué genes están sobreexpresados o infraexpresados en los hombres con EP y en mujeres con EP, en comparación con hombres y mujeres sanos respectivamente [(ME-MS) vs (HE-HS)].

3.3.3. Análisis de enriquecimiento funcional

Una vez realizado el análisis de expresión diferencial, se lleva a cabo un análisis de enriquecimiento funcional con GSEA. El objetivo de este tipo de análisis es indicar cual es la función biológica que hay detrás de un conjunto de genes que se han clasificado según su expresión diferencial.

Para ello se asocia cada gen de la matriz de expresión con su término GO de manera que puedan ser ordenados según: procesos biológicos, componentes celulares y funciones moleculares.

3.4. METAANÁLISIS

Una vez completado el análisis bioinformático individual para cada estudio (expresión diferencial y caracterización funcional), realizamos el metaanálisis a nivel de función. Éste nos permite ver diferencias significativas que, a partir de un solo conjunto de datos no sería posible.

La herramienta *MetaFun* nos devuelve los resultados del metaanálisis con dos tipos de representaciones gráficas: gráfico de bosque o *forest plot* y gráfico de embudo o *funnel plot*.

En el primero; se representa en el eje vertical los estudios que han formado parte del metaanálisis y en el eje horizontal, el valor combinado de la variable de interés, *Log Odds Ratio* del término representado obtenido a partir de la caracterización funcional, con su intervalo de confianza. El tamaño del cuadrado negro en el centro del gráfico es proporcional a la precisión de las estimaciones.

En el segundo; se enfrenta el valor cuantitativo de la variable de interés con su error estándar, de manera que, una vez obtenido el gráfico con los estudios representados, aquellos que aparezcan alejados del 0 en el eje horizontal serán estudios poco precisos. En cuanto al eje vertical, aquellos que estén cercanos al 0, tendrán un error estándar menor. La variable de interés representada en el eje horizontal es la misma que la representada en el *forest plot*. Para los estudios que se encuentran fuera de la región de confianza delimitada, debería ser revisada su inclusión en el metaanálisis.

4. RESULTADOS

Para la obtención de los resultados del análisis individual de los estudios y del metaanálisis, hemos hecho uso de la herramienta web *MetaFun* (Malmierca, 2020).

4.1 REVISIÓN SISTEMÁTICA Y SELECCIÓN DE ESTUDIOS

Tras aplicar los criterios de selección (palabra clave "Parkinson" y tecnologías: microarrays o secuenciación de alto rendimiento) en GEO, nos quedamos con un total de 31 estudios candidatos para incluir en el metaanálisis.

Tras estudiar más a fondo estos 31 estudios, excluimos 17 por no centrarse realmente en la EP y 10 por no contener información sobre la edad y el sexo de las muestras o por emplear las mismas muestras para realizar diversos estudios, tal y como se muestra en la Figura 3.

Así pues, finalmente 4 estudios fueron seleccionados para el metaanálisis (Tabla 1).

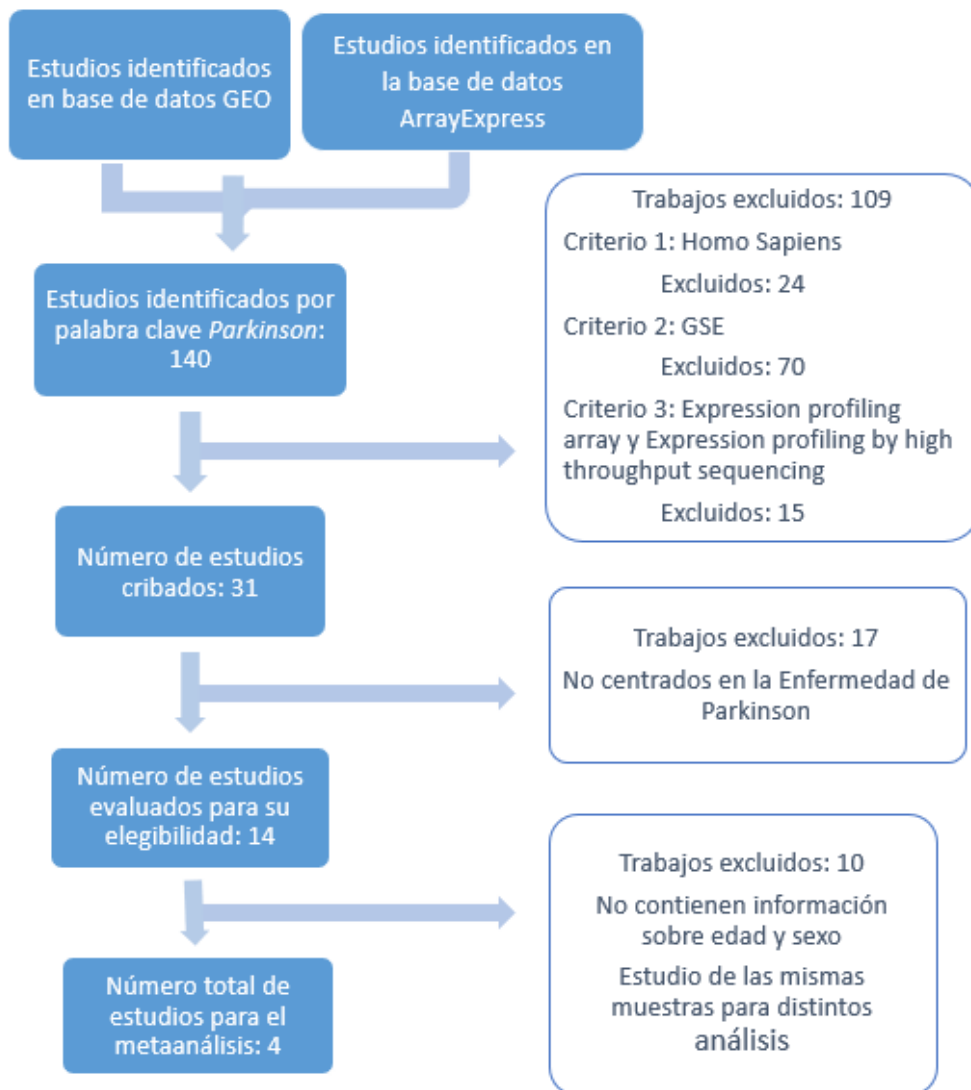


Figura 3: Diagrama de flujo de la revisión sistemática realizada en este trabajo.

Estudio	Plataforma	Tejido
GSE128177	GPL24676 Illumina NovaSeq 6000 (Homo Sapiens)	Musculo esquelético. Musculo vasto lateral
GSE20292	GPL96 Affymetrix Human Genome U133A Array	Región cerebral: Sustancia negra completa del cerebro postmortem
GSE20153	GPL570 Affymetrix Human Genome U133 Plus 2.0 Array	Linfocitos B de sangre periférica transformados con EBV
GSE7621	GPL570 Affymetrix Human Genome U133 Plus 2.0 Array	Sustancia negra humana postmortem

Tabla 1: Estudios finalmente seleccionados tras la revisión sistemática. Descripción de la plataforma tecnológica y el tejido empleado en cada estudio.

4.2 ANÁLISIS INDIVIDUAL DE LOS ESTUDIOS

4.2.1. Procesamiento de los datos

Con los datos de cada estudio descargados y estandarizada la nomenclatura, se describió el número de pacientes clasificados según el sexo en cada estudio, como se puede observar en la Figura 4. La evaluación de la normalización de los valores de expresión se realizó mediante diagramas de cajas (Figuras 5,8,11 y 14).

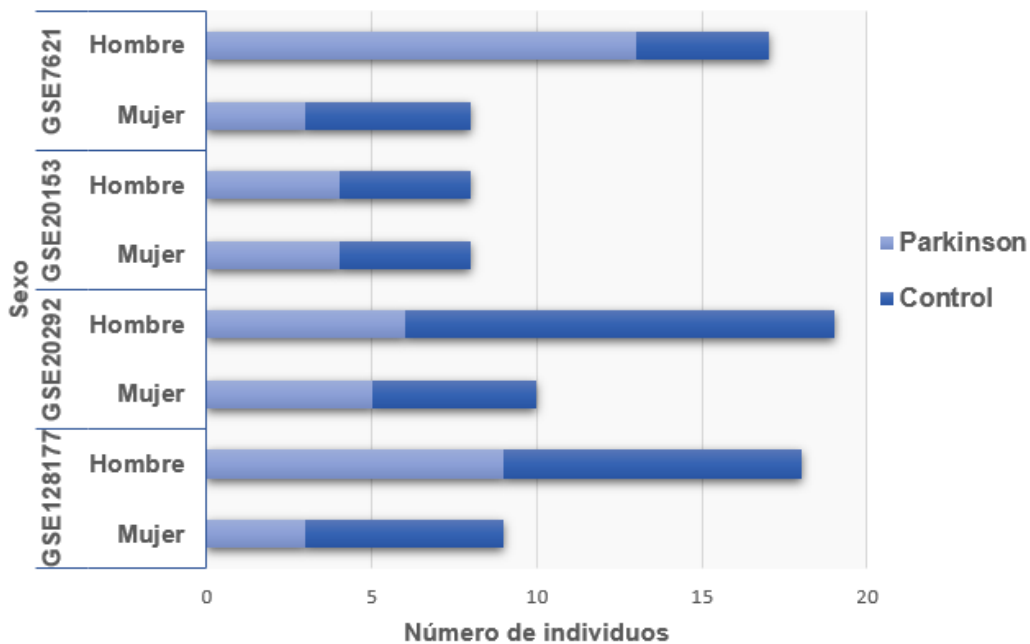


Figura 4: Número de individuos en cada estudio. Distribuidos según sexo y estado.

GSE7621

En este estudio se utilizó tejido de sustancia negra del cerebro postmortem de individuos que sufrían EP y de individuos sanos. El objetivo era conocer el programa de expresión génica que predispone a la EP (Lesnick et al., 2007).

En la Tabla 2 se muestran los individuos que han participado en el estudio. Están clasificados según el sexo y si padecen o no la enfermedad de Parkinson.

	Mujer	Hombre	Total
Parkinson	3	13	16
Control	5	4	9
Total	8	17	25

Tabla 2: Individuos incluidos en el estudio GSE7621. Clasificados por sexo y estado.

Los análisis exploratorios de PCA y *Clustering*, no mostraron agrupamiento de las muestras (Figuras 6 y 7).

GSE20153

En este estudio se realizó un análisis del transcriptoma de todo el genoma de los cambios de expresión en líneas celulares transformadas por el virus de Epstein-Barr (EBV) del depósito de células Coriell (Coriell Institute for Medical Research) en sujetos Parkinson y Control (Zheng et al., 2010).

En la Tabla 3 se muestran los individuos que han participado en el estudio. Están clasificados según el sexo y si padecen o no la enfermedad de Parkinson.

	Mujer	Hombre	Total
Parkinson	4	4	8
Control	4	4	8
Total	8	8	16

Tabla 3: Individuos incluidos en el estudio GSE20153. Clasificados por sexo y estado.

Los análisis exploratorios de PCA y *Clustering*, no mostraron agrupamiento de las muestras (Figuras 9 y 10).

GSE20292

En el trabajo llevado a cabo por Zhang et al. en 2005 se realizaron perfiles de expresión génica y análisis funcionales de grupos de genes de tres regiones cerebrales distintas (sustancia negra, putamen y área 9) en tejido postmortem de individuos que sufrían EP y de individuos control. En este trabajo hemos incluido únicamente los datos obtenidos de la sustancia negra para evitar así la sobrepresencia de los individuos de este estudio en el metaanálisis.

En la Tabla 4 se muestran los individuos que han participado en el estudio. Están clasificados según el sexo y si padecen o no la enfermedad de Parkinson.

	Mujer	Hombre	Total
Parkinson	5	6	11
Control	5	13	18
Total	10	19	29

Tabla 4: Individuos incluidos en el estudio GSE20292. Clasificados por sexo y estado.

Los análisis exploratorios de PCA y *Clustering*, no mostraron agrupamiento de las muestras (Figuras 12 y 13).

GSE128177

En estudios anteriores, se observó que un deterioro en la integridad de la unión neuromuscular podría estar asociada con una activación excesiva de la unidad motora. Así pues, en este estudio, el objetivo fue probar la hipótesis de que esta activación excesiva, se asociaría a distintas redes de expresión génica. Para ello se generaron datos de RNA-seq del transcriptoma de músculo esquelético de pacientes sanos y con EP (Lavin et al., 2020).

En la Tabla 5 se muestran los individuos que han participado en el estudio. Están clasificados según el sexo y si padecen o no la enfermedad de Parkinson. En este estudio hemos prescindido de las muestras de individuos sanos jóvenes ya que este grupo no se tiene en cuenta en ninguno de los otros estudios que hemos empleado para el metaanálisis.

	Mujer	Hombre	Total
Parkinson	3	9	12
Control	6	9	15
Total	9	18	27

Tabla 5: Individuos incluidos en el estudio GSE128177. Clasificados por sexo y estado.

El análisis exploratorio de PCA mostró agrupamiento de las muestras en los grupos experimentales (Figura 15), pero el análisis clúster no (Figura 16).

La gran heterogeneidad que observamos en los gráficos de PCA (Figuras 6,9,12 y 15) y *Clustering* (Figuras 7,10,13 y 16) son de esperar ya que las muestras se han tomado en estadios distintos de la enfermedad. Además, como está indicado en la Tabla 1, el estudio GSE128177 recoge muestras de dos tejidos distintos, lo que incrementa aún más esta heterogeneidad.

Observando la variabilidad de las muestras en los diagramas de cajas o *boxplot* (Figuras 5,8,11 y 14) vemos que el número de hombres y mujeres está desbalanceado en todos los estudios, a excepción del GSE20153.

En general, la dispersión de los datos de expresión es similar cuando comparamos muestras de un mismo estudio. Lo mismo podemos decir acerca de la mediana que está representada por la línea que cruza la caja y de los valores máximos y mínimos que están representados por los bigotes.

GSE7621

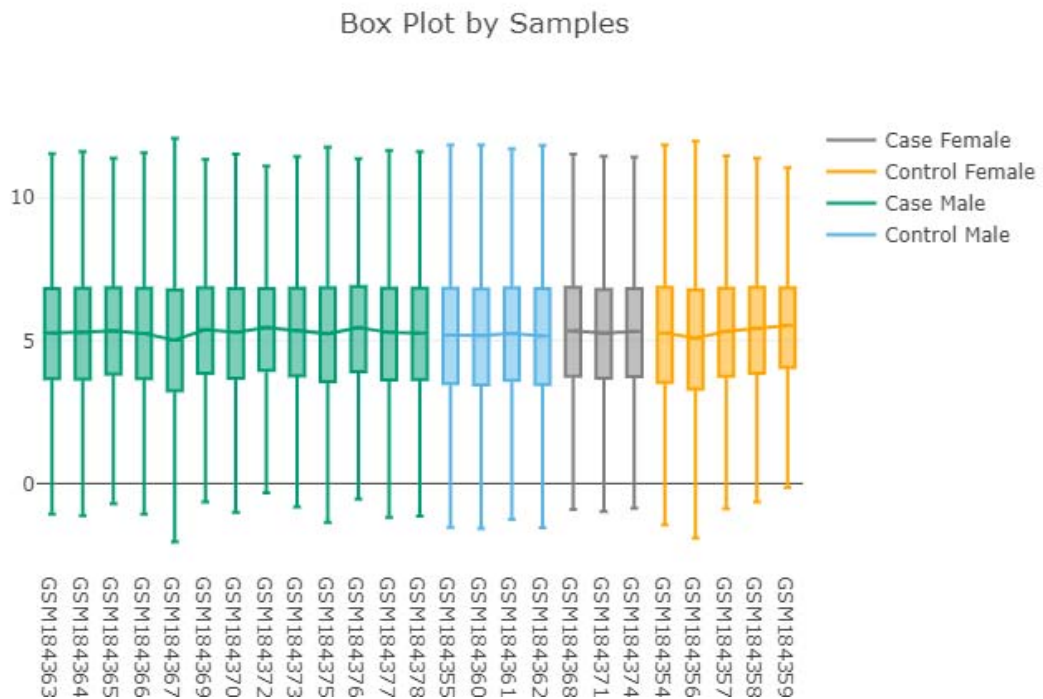


Figura 5: Diagrama de cajas en el que están representados los valores normalizados de las muestras empleadas en el estudio GSE7621, coloreadas según sexo y estado.

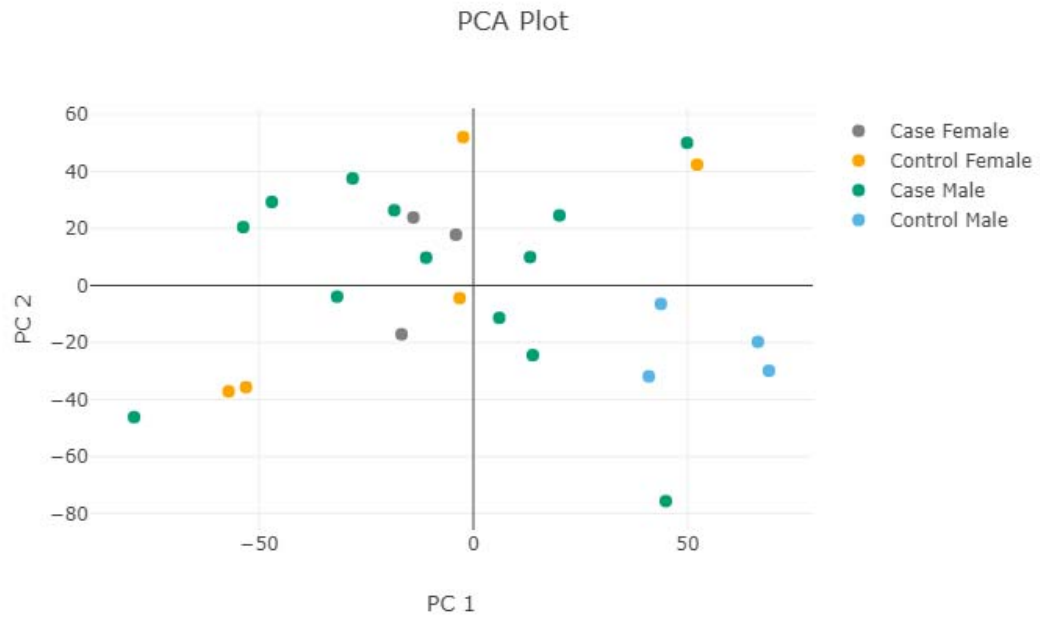


Figura 6: Análisis de componentes principales (PCA) de las muestras del estudio GSE7621.

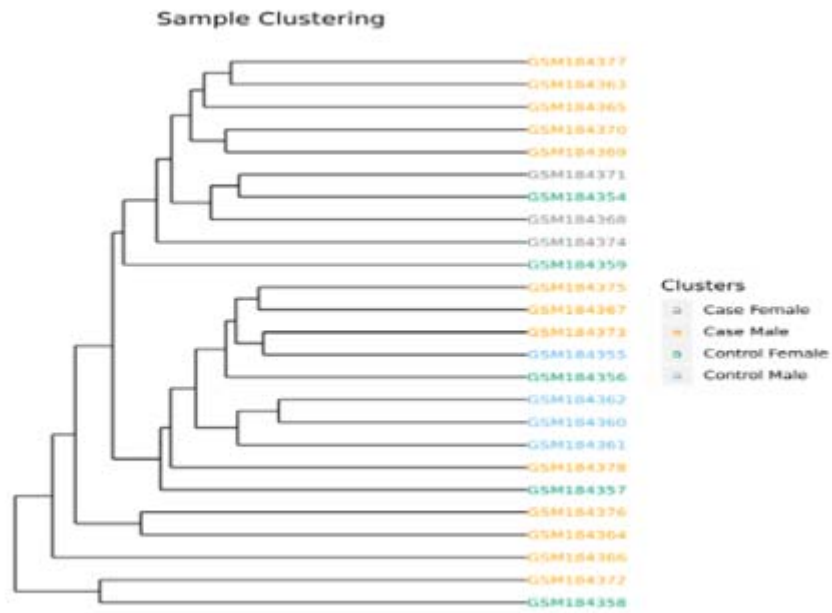


Figura 7: Clustering de las muestras del estudio GSE7621. Cada grupo experimental aparece representado de un color.

GSE20153

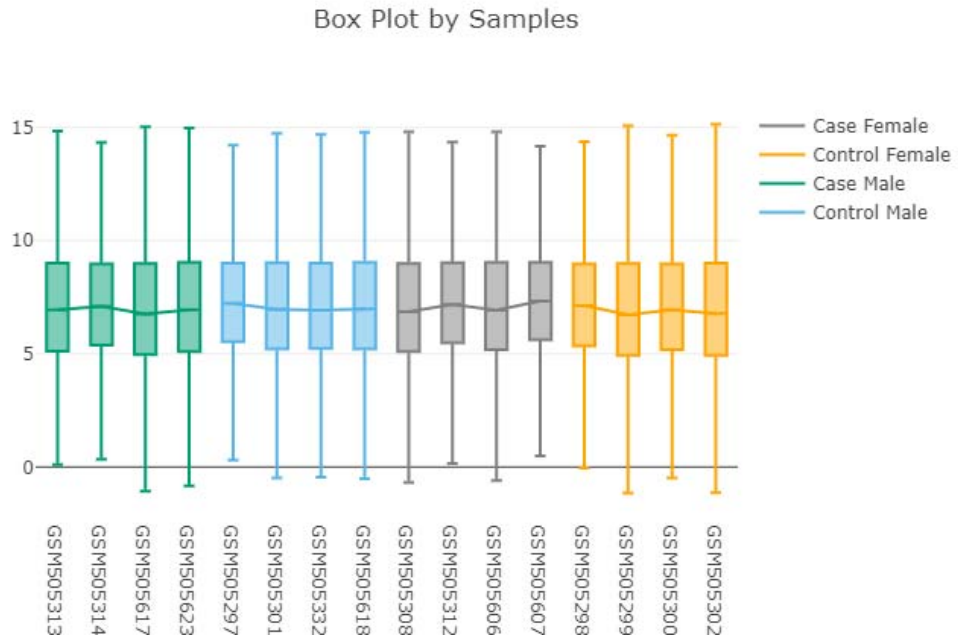


Figura 8: Diagrama de cajas en el que están representados los valores normalizados de las muestras empleadas en el estudio GSE20153, coloreadas según sexo y estado.

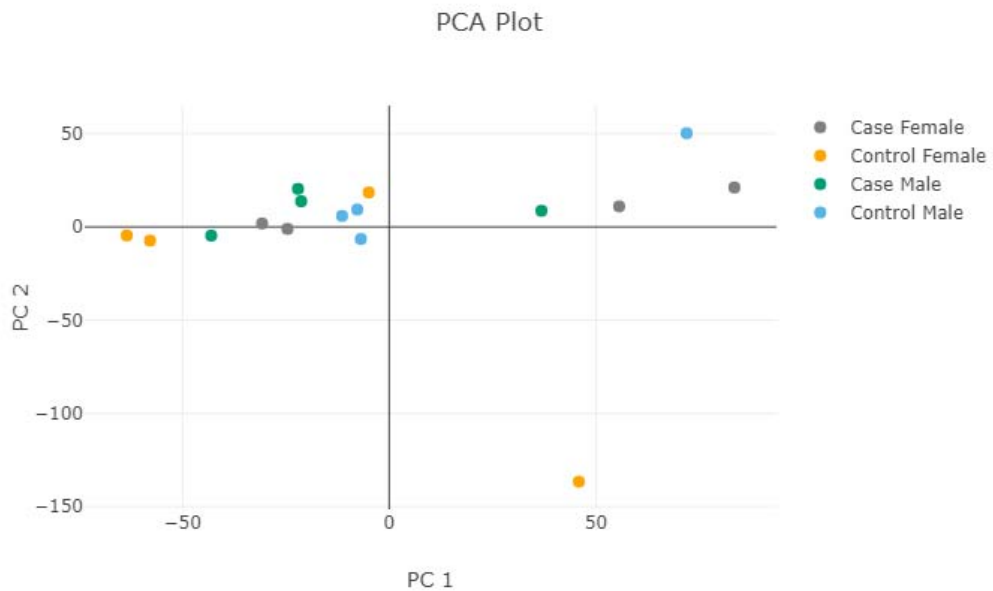


Figura 9: Análisis de componentes principales (PCA) de las muestras del estudio GSE20153.

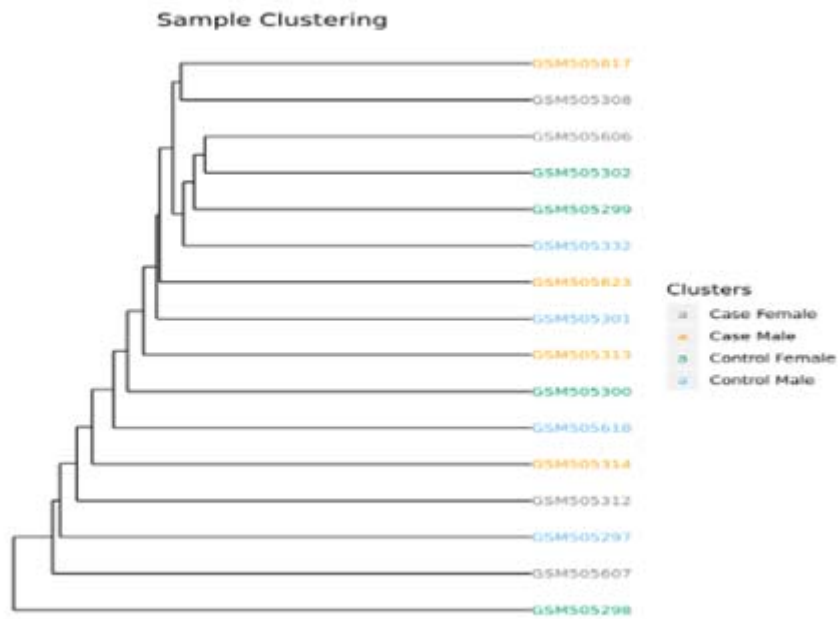


Figura 10: Clustering de las muestras del estudio GSE20153. Cada grupo experimental aparece representado de un color.

GSE20292

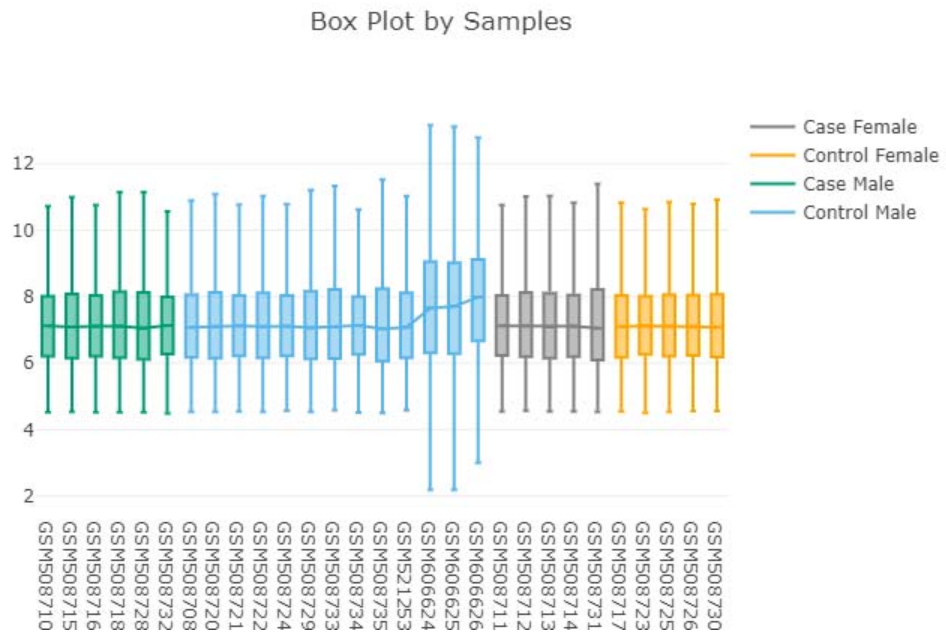


Figura 11: Diagrama de cajas en el que están representados los valores normalizados de las muestras empleadas en el estudio GSE20292, coloreadas según sexo y estado.

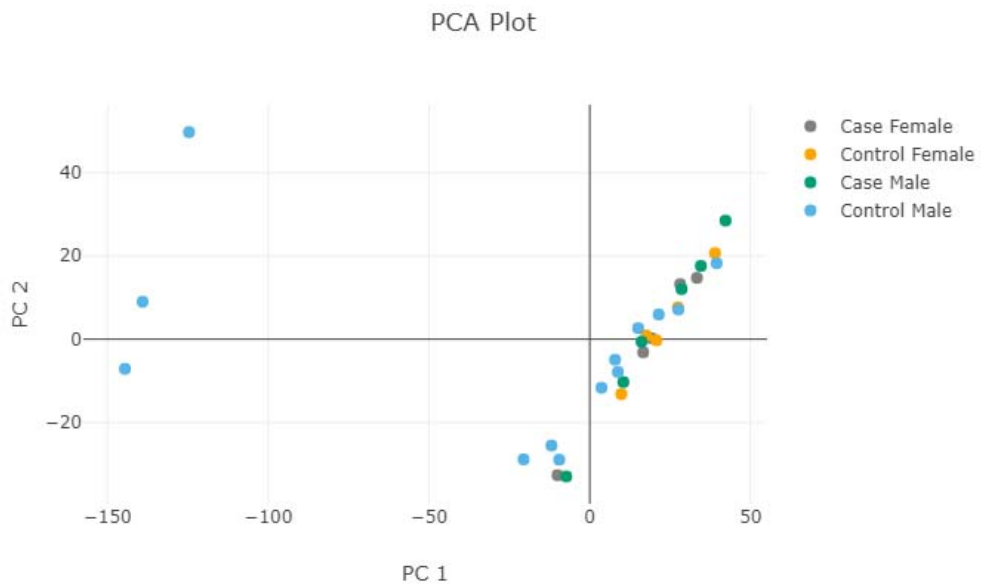


Figura 12: Análisis de componentes principales (PCA) de las muestras del estudio GSE20292.

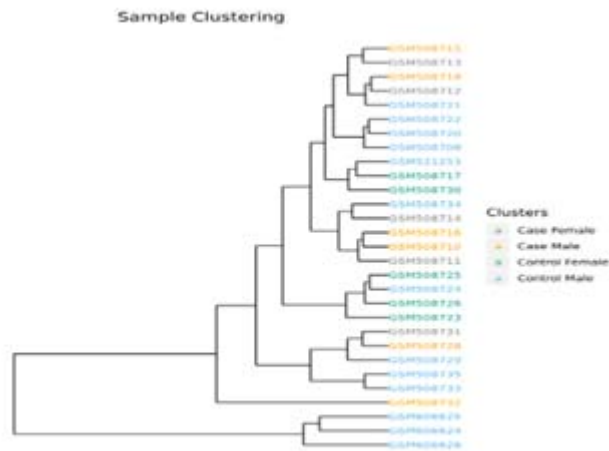


Figura 13: Clustering de las muestras del estudio GSE20292. Cada grupo experimental aparece representado de un color.

GSE128177

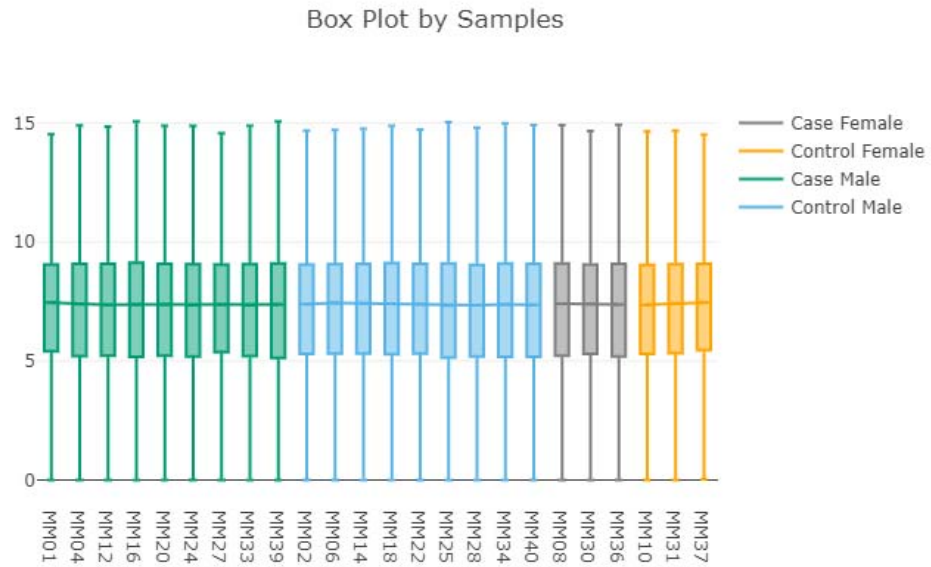


Figura 14: Diagrama de cajas en el que están representados los valores normalizados de las muestras empleadas en el estudio GSE128177, coloreadas según sexo y estado.

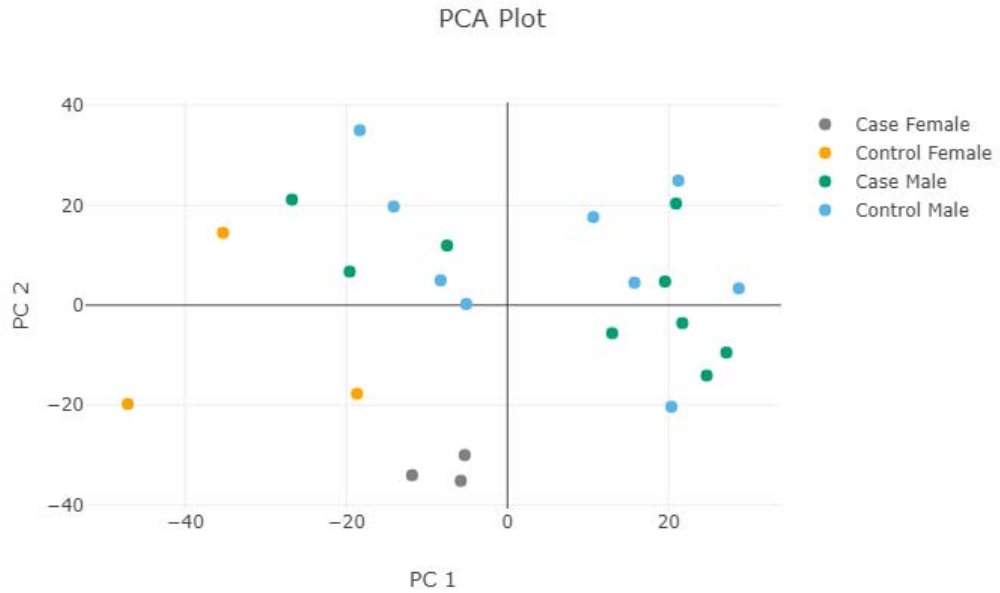


Figura 15: Análisis de componentes principales (PCA) de las muestras del estudio GSE128177.

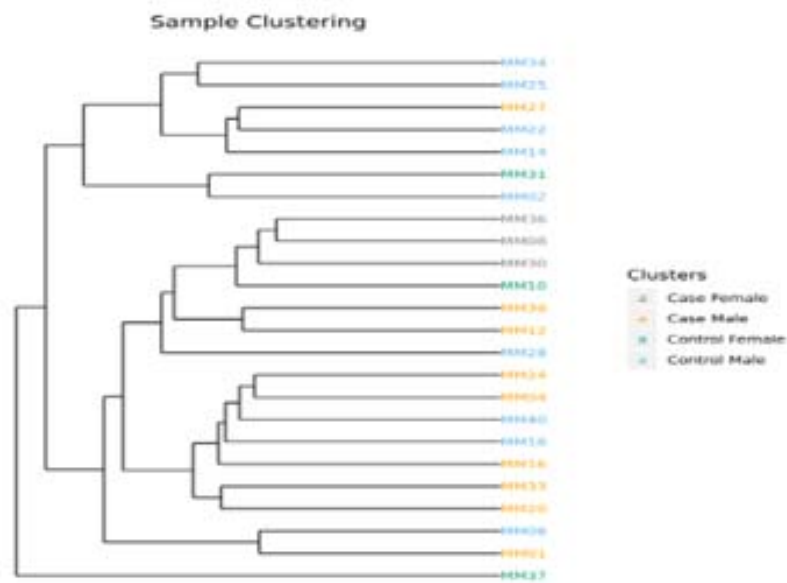


Figura 16: Clustering de las muestras del estudio GSE128177. Cada grupo experimental aparece representado de un color.

4.2.2. Análisis de expresión diferencial

MetaFun nos proporciona una tabla resumen con el número total de genes diferencialmente expresados en cada uno de los estudios, distinguiendo entre genes sobreexpresados en mujeres (sig.UP) y en hombres (sig.DOWN). En los estudios evaluados no se detectan genes cuya diferencia de expresión sea estadísticamente significativa entre hombres y mujeres cuando comparamos los grupos de enfermos frente a los sanos: (ME-MS) vs. (HE-HS) (Tabla 6).

Differential Expression Summary				
Study Name	Total.genes	sig.Total	sig.UP	sig.DOWN
gse7621	20930	0	0	0
gse20153	20974	0	0	0
gse20292	12993	0	0	0
gse128177	14369	0	0	0

Tabla 6: Número de genes cuya diferencia de expresión es significativa. Están clasificados en sig.Up y sig.Down según si se encuentran sobrerrepresentados en mujeres o en hombres respectivamente. Tabla proporcionada por MetaFun.

Además de esta tabla resumen, *MetaFun* proporciona para cada uno de los estudios una tabla con el identificador y nombre de los genes de los 50 mejores resultados obtenidos. En ella se incluyen parámetros que nos permiten saber cuáles de estos genes están sobreexpresados en cada uno de los sexos y si dicha diferencia de expresión es significativa (Tabla 7).

Entrez ID	Gene Name	logFC	P value	Adjusted P value
7004	TEA domain transcription factor 4	-2.662	2.0974e-5	2.4771e-1
3958	galectin 3	-2.744	3.4574e-5	2.4771e-1
101929645	uncharacterized LOC101929645	-5.980	4.0704e-5	2.4771e-1
59351	prostate and breast cancer overexpressed 1	-2.711	4.9662e-5	2.4771e-1
51029	desumoylating isopeptidase 2	-1.958	6.2823e-5	2.4771e-1
152189	CKLF like MARVEL transmembrane domain containing 8	-4.068	7.6664e-5	2.4771e-1
84688	chromosome 9 open reading frame 24	1.500	9.5394e-5	2.4771e-1
4620	myosin heavy chain 2	-3.689	9.7427e-5	2.4771e-1
25816	TNF alpha induced protein 8	-2.141	1.0652e-4	2.4771e-1
100101266	hepatitis A virus cellular receptor 1 pseudogene 1	-3.909	1.3364e-4	2.5888e-1

Tabla 7: Resultados de expresión diferencial del estudio GSE7621. Representación parcial de la tabla proporcionada por *MetaFun*. *logFC* (logaritmo del Fold Change), nos indica cuanto cambia la expresión cuando comparamos (ME-MS) vs. (HE-HS), *P value* y *Adjusted P value* (respectivamente, valor de *p* y valor de *p* ajustado que nos informan del nivel de significación de cada gen).

4.2.3. Análisis de enriquecimiento

De la misma manera que en el análisis de expresión diferencial, *MetaFun* nos proporciona una tabla resumen del número total de funciones con mayor sobrerrepresentación en cada uno de los estudios. En esta tabla también se indica el número total de funciones significativas, distinguiendo entre las que están más activas diferencialmente en mujeres (sig.UP) y las que están más activas en hombres (sig.DOWN) (Tabla 8).

Functional Profiling Summary				
Study Name	Total.functions	sig.UP	sig.DOWN	sig.Total
gse7621	8294	600	735	5340
gse20153	8298	28	366	4399
gse20292	7779	698	916	4678
gse128177	7578	113	454	3910

Tabla 8: Número de funciones cuya diferencia de expresión es significativa. Están clasificadas en sig.Up y sig.Down según si se encuentran sobrerrepresentados en mujeres o en hombres respectivamente. Tabla proporcionada por MetaFun.

Además de esta tabla resumen, *MetaFun* proporciona para cada uno de los estudios una tabla con el identificador y nombre de las funciones de los 50 mejores resultados obtenidos. En ella se incluyen parámetros que nos permiten saber cuáles de estas funciones están sobreexpresadas en cada uno de los sexos y si dicha diferencia de expresión es significativa. Además, la columna correspondiente a la variable N nos indica el número de genes que participan en cada función (Tabla 9).

GO ID	GO Term	LOR	P value	Adjusted P value	N
GO:0006119	oxidative phosphorylation	0.722	3.6343e-17	2.3034e-13	133
GO:0009123	nucleoside monophosphate metabolic process	0.427	2.2792e-15	7.2226e-12	356
GO:0009161	ribonucleoside monophosphate metabolic process	0.433	5.7225e-15	1.2090e-11	337
GO:0045333	cellular respiration	0.574	1.2448e-14	1.9724e-11	183
GO:0009126	purine nucleoside monophosphate metabolic process	0.426	4.1105e-14	3.2973e-11	325
GO:0009167	purine ribonucleoside monophosphate metabolic process	0.427	4.0388e-14	3.2973e-11	324
GO:0022900	electron transport chain	0.574	4.1619e-14	3.2973e-11	175
GO:0022904	respiratory electron transport chain	0.715	3.8994e-14	3.2973e-11	108
GO:0009141	nucleoside triphosphate metabolic process	0.410	7.6065e-14	5.3567e-11	344
GO:0009144	purine nucleoside triphosphate metabolic process	0.418	1.4376e-13	9.1117e-11	325

Tabla 9: Resultados del análisis de enriquecimiento funcional del estudio GSE7621 para la categoría "Procesos biológicos" de GO. GO ID (identificador del término de la Gene Ontology), LOR (logaritmo del Odds Ratio), P value y Adjusted P value (respectivamente, valor de p y valor de p ajustado que nos informan del nivel de significación de cada función), N: número de genes que participan en cada función. Representación parcial de la tabla proporcionada por MetaFun.

4.3 METAANÁLISIS

En este último apartado, *MetaFun* proporciona una tabla con las funciones obtenidas como resultado final a nivel de función tras realizar el metaanálisis (Tabla 10). En ella para cada una de las funciones se aportan los gráficos de bosque y embudo (Figuras 17 a 20) que nos permiten saber la relevancia de cada estudio en el metaanálisis.

En el gráfico de bosque para el término GO:0022904 (Figura 17), dado que el área del cuadrado negro es proporcional a la contribución del estudio en el resultado global, podemos ver que el estudio GSE7621 participa más que el GSE20292. Pero los dos estudios que aparecen representados en este diagrama son estadísticamente significativos para este término GO, ya que sus intervalos de confianza no cruzan el valor nulo de la variable de resultados.

En el gráfico de bosque para el término GO:0002250 (Figura 19), observamos que el cuadrado negro del estudio GSE7621 es el que tiene mayor área, de manera que al igual que en el caso anterior, es el que más contribuye al resultado global de este término. Dado que en ninguno de los estudios que participan en este término GO los intervalos de confianza cruzan el valor nulo, todos ellos son estadísticamente significativos.

En cuanto a los diagramas de embudo de los términos GO:0022904 (Figura 18) y GO:0002250 (Figura 20), vemos que los estudios que han contribuido a dichos términos no se encuentran muy cercanos al 0. Es por ello por lo que no podemos considerarlos estudios de gran precisión.

Meta-Analysis Results

GO Term	Name	Summary LOR
GO:0001525	angiogenesis	-0.339 (♂)
GO:0002250	adaptive immune response	-0.355 (♂)
GO:0002694	regulation of leukocyte activation	-0.277 (♂)
GO:0006119	oxidative phosphorylation	0.833 (♀)
GO:0006520	cellular amino acid metabolic process	0.389 (♀)
GO:0009123	nucleoside monophosphate metabolic process	0.494 (♀)
GO:0009126	purine nucleoside monophosphate metabolic process	0.504 (♀)

GO:0009141	nucleoside triphosphate metabolic process	0.484 (♀)
GO:0009144	purine nucleoside triphosphate metabolic process	0.485 (♀)
GO:0009161	ribonucleoside monophosphate metabolic process	0.513 (♀)
GO:0009167	purine ribonucleoside monophosphate metabolic process	0.505 (♀)
GO:0009199	ribonucleoside triphosphate metabolic process	0.491 (♀)
GO:0009205	purine ribonucleoside triphosphate metabolic process	0.489 (♀)
GO:0010257	NADH dehydrogenase complex assembly	0.858 (♀)
GO:0022900	electron transport chain	0.604 (♀)
GO:0022904	respiratory electron transport chain	0.827 (♀)
GO:0030099	myeloid cell differentiation	-0.310 (♂)
GO:0030335	positive regulation of cell migration	-0.285 (♂)
GO:0032981	mitochondrial respiratory chain complex I assembly	0.858 (♀)
GO:0033108	mitochondrial respiratory chain complex assembly	0.816 (♀)

Tabla 10: Resultados del metaanálisis. GO Term (identificador del término de la Gene Ontology), Name (función a la que hace referencia el GO Term), LOR (logaritmo del Odds Ratio), el signo positivo y negativo nos indican si la función está sobrerrepresentada en mujeres o en hombres respectivamente. Tabla proporcionada por MetaFun

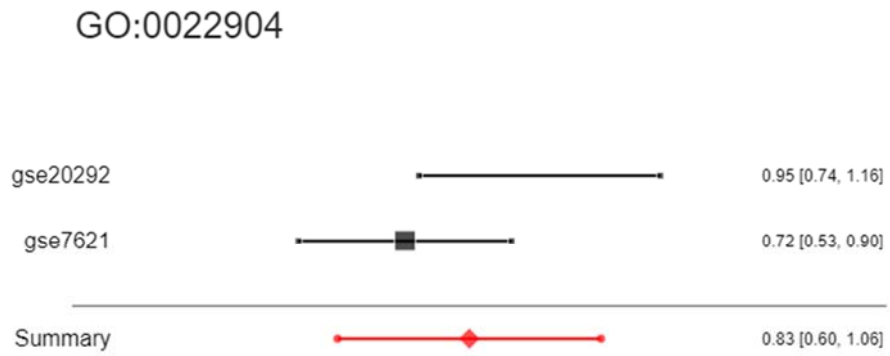


Figura 17: Gráfico de bosque para el término GO:0022904 (Respiratory electron transport chain).

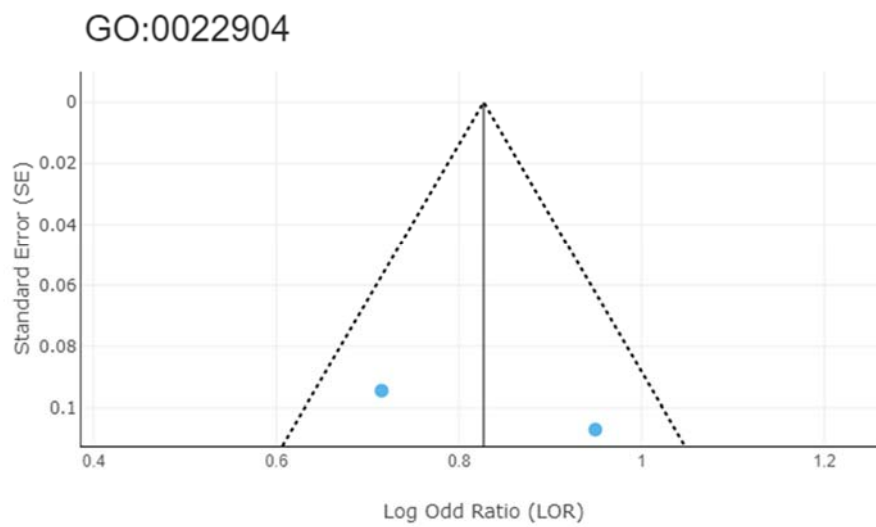


Figura 18: Gráfico de embudo para el término GO:0022904 (Respiratory electron transport chain).

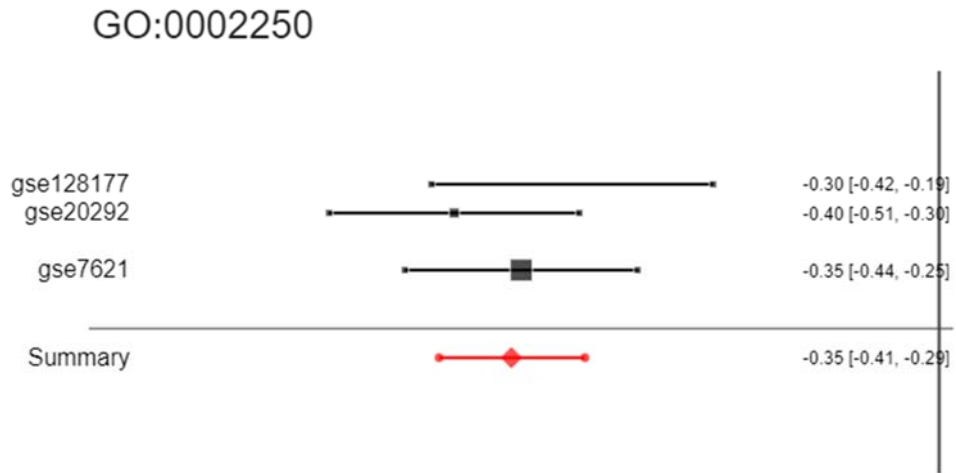


Figura 19: Gráfico de bosque para el término GO:0002250 (*Adaptative immune system*).

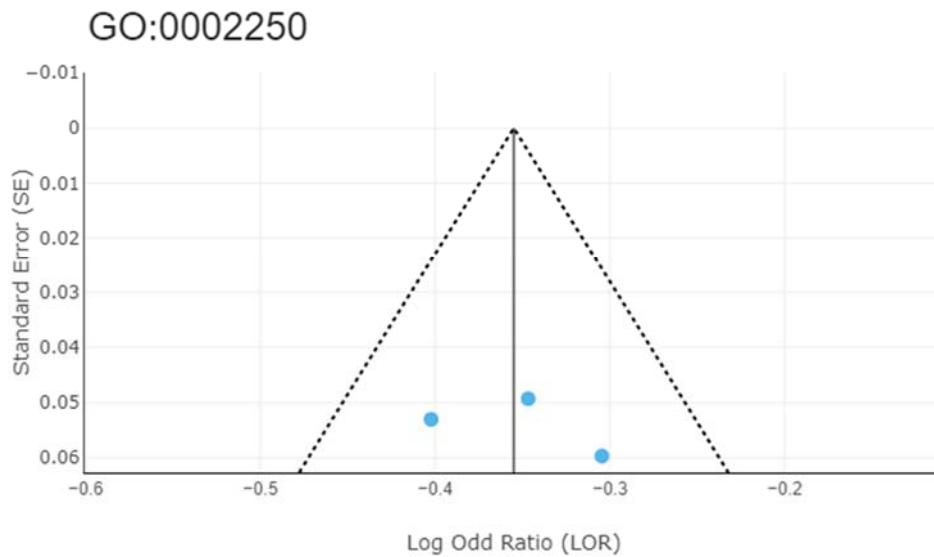


Figura 20: Gráfico de embudo para el término GO:0002250 (*Adaptative immune system*).

4.4 DISCUSIÓN

La enfermedad de Parkinson es un trastorno neurodegenerativo cuya prevalencia e incidencia ha aumentado considerablemente en los últimos años. A pesar de que la causa de la EP sigue siendo desconocida y a día de hoy no existe un tratamiento que frene el daño neurodegenerativo, se sabe que existen diferencias en el desarrollo de la enfermedad en hombres y mujeres. Algunas de estas diferencias se asocian a la presencia de estrógenos en mujeres. Estas hormonas parecen proteger a las mujeres, de manera que los síntomas motores son más leves que en los hombres.

A pesar de que estas diferencias son conocidas en la comunidad científica, se siguen realizando muchos estudios que no incluyen la información del sexo o que simplemente incluyen en él individuos del sexo masculino. Debido a esta falta de información junto con la falta de estandarización de los datos disponibles, la dificultad para reutilizar y reanalizar los datos es mayor.

Así pues, aunque existe una gran cantidad de estudios que podrían haberse utilizado para realizar el metaanálisis, la mayoría de ellos fueron descartados debido a la falta de dicha información.

Pese al número reducido de estudios disponibles y la gran heterogeneidad de estos (incluyendo los diferentes tipos de muestra biológicas utilizadas), se aplicaron las técnicas de metaanálisis para la caracterización las diferencias del sexo en esta enfermedad. Con la potencia de este método es posible detectar diferencias significativas que, a partir de un solo conjunto de datos no habría sido posible.

Como hemos podido observar, en algunos estudios, no obtenemos resultados significativos en el análisis de la expresión diferencial, sin embargo, con la propuesta que presentamos mediante el metaanálisis de los análisis de enriquecimiento funcional de los estudios, es posible conocer qué funciones están caracterizando el conjunto de los estudios.

Analizando los resultados del metaanálisis funcional a nivel de términos GO, hemos sido capaces de detectar, a grandes rasgos, un enriquecimiento de funciones relacionadas con la respiración celular en mujeres y un enriquecimiento de funciones relacionadas con el sistema inmune y la angiogénesis en hombres.

Los términos GO:0006119, GO:0022900, GO:0022904, GO:0032981 y GO:0033108 hacen referencia a la fosforilación oxidativa, cadena de transporte de electrones y complejo de la cadena respiratoria mitocondrial. Estos están sobrerrepresentados en mujeres y se sabe que el metabolismo energético dependiente de las mitocondrias y la producción de ATP está notablemente alterado en los individuos con EP. También se ha encontrado daño oxidativo de ácidos nucleicos en pacientes con EP, así como acumulación de 8-oxoguanina en el ADN mitocondrial en pacientes con EP. Esto sugiere que estas lesiones pueden causar la pérdida de neuronas dopaminérgicas. Es probable que los ácidos nucleicos sean un factor de riesgo para la EP, por lo que comprendiendo bien estos mecanismos se podrían desarrollar nuevas estrategias para proteger el cerebro contra el estrés oxidativo (Nakabeppu et al., 2007).

También se han encontrado sobrerrepresentados en mujeres otros términos GO como GO:0009123, GO:0009126 y GO:0009144 que hacen referencia a procesos metabólicos de nucleósidos, principalmente aquellos con derivados de purinas en su estructura. Esto es interesante dado que Garcia-Esparcia et al. (2015) ya indicaron, después de analizar la expresión de ARNm de genes involucrados en el metabolismo de las purinas en diferentes regiones del cerebro, que algunos de los genes estudiados se encontraban

down-regulated en individuos con EP en comparación con los controles. Pero en este estudio no se demostró correlación entre la expresión génica y el sexo.

En los últimos años se han llevado a cabo numerosos estudios sobre cómo el metabolismo de los nucleósidos está involucrado en la actividad cerebral normal y como su disfuncionalidad está implicada en muchos trastornos neurodegenerativos como es la enfermedad de Parkinson (Boison, 2011).

Por otro lado, los términos GO:0002250, GO:0002694 y GO:0030099 son términos relacionados con el sistema inmune que se encuentran sobreexpresados en hombres. La alteración de algunas funciones del sistema inmune ya se había relacionado con la enfermedad de Parkinson en trabajos anteriores.

Nissen et al., (2019) realizaron un estudio cuyo objetivo era investigar los cambios que se producían en las células inmunes periféricas debido a la EP. A lo largo del trabajo observaron algunas diferencias de sexo a pesar de la distribución equitativa de la cohorte y de la ausencia de diferencias clínicas de sexo significativas. *La frecuencia de las células CD14 + / HLA-DR + fue mayor en los pacientes con EP masculinos que en los femeninos. El porcentaje de la nueva población CD14 + / HLA-DR- en cultivo fue especialmente mayor en pacientes femeninos con EP. Estas observaciones podrían relacionarse con las disparidades en el sistema inmunitario. Mientras que los macrófagos en las mujeres son altamente fagocíticos y presentadores de antígenos más eficientes, los macrófagos en los hombres tienen una mayor producción de citocinas proinflamatorias y expresión de TLR4. Esto podría ser relevante, dado que se sugiere una activación sostenida de TLR4 por α -syn para promover la enfermedad.*

Garretti et al., (2019) comentan resultados similares a los mencionados anteriormente. Su revisión sostiene que los niveles aumentados de las citocinas proinflamatorias se consideran indicativos de un sistema inmune que responde al daño tisular y/o moléculas extrañas. Sus niveles se correlacionan con la etapa clínica de la EP, destacando un papel para la inflamación periférica en su progresión. También defienden que los ratones nulos en el complejo mayor de histocompatibilidad de clase II (MHC-II) son resistentes a la muerte neuronal causada por la intoxicación con MPTP (profármaco que causa síntomas permanentes de la enfermedad de Parkinson al destruir las neuronas dopaminérgicas en la sustancia negra del cerebro) y sus microglías no liberan citocinas proinflamatorias en contraste con los animales salvajes.

Todo esto es debido a la presencia de células T en el cerebro, que diversos estudios indican que pueden ser de naturaleza autorreactiva. Estas células pueden generar una respuesta autoinmune a la proteína α -sinucleína.

Otro resultado interesante que nos ofrece *MetaFun* es que la angiogénesis (GO:0001525) es una función que se encuentra sobreexpresada en hombres. Janelidze et al., (2015) ya demostraron que los factores angiogénicos estaban *up-regulated* en el líquido cefalorraquídeo (LCR) de individuos con EP. Con este trabajo, pudieron asociar la activación angiogénica con dificultades al andar y con disfunciones en la barrera hematoencefálica, lesiones en la sustancia blanca, micro hemorragias cerebrales, neuroinflamación y degeneración axonal en pacientes con EP, cosa que no ocurría en los individuos control.

Este metaanálisis también ha detectado algunas funciones significativas como GO:0030335 a lo largo de los estudios evaluados, que hasta el momento no han sido descritos en otros trabajos de la enfermedad de Parkinson. Esta función concreta está sobrerrepresentada en hombres y hace referencia a la regulación positiva de la

migración celular. Su revisión puede proporcionar nuevas preguntas de investigación para ser abordadas y contestadas en otros trabajos sobre esta enfermedad.

5. CONCLUSIONES

1. El metaanálisis es un método robusto que nos permite combinar datos obtenidos de distintos estudios a partir de una revisión sistemática exhaustiva.
2. Se ha detectado sobrerrepresentación de funciones relacionadas con la respiración celular en mujeres. Estas alteraciones se han asociado con la EP en estudios anteriores.
3. En los hombres se ha detectado sobrerrepresentación del proceso de angiogénesis, así como una mayor expresión de términos vinculados a la respuesta inmunitaria.
4. *MetaFun* ha confirmado ser una herramienta web útil, accesible y fácil de emplear para la realización de metaanálisis de datos ómicos para la caracterización de las diferencias de sexo.
5. La realización de esta estrategia *in silico* permite detectar diferencias debidas al sexo, abriendo nuevas perspectivas que permitan mejorar las opciones de tratamiento con una futura medicina personalizada y de precisión.

6. REFERENCIAS

ANÁLISIS FUNCIONAL DE GENES | SERVICIO DE BIOINFORMÁTICA (2020). Retrieved 28 April 2020, from http://bioinfo.usal.es/Análisis_Funcional_Genes.html

ARRAYEXPRESS < EMBL-EBI. (2020). Retrieved 18 March 2020, from <https://www.ebi.ac.uk/arrayexpress/>

ASCHERIO, A.; LEWITT, PA.; XU, K.; EBERLY, S.; WATTS, A.; MATSON, WR.; MARRAS, C.; KIEBURTZ, K.; RUDOLPH, A.; BOGDANOV, MB.; SCHWID, SR.; TENNIS, M.; TANNER, CM.; BEAL, MF.; LANG, AE.; OAKES, D.; FAHN, S.; SHOULSO, I.; SCHWARZSCHILD, MA.; PARKINSON STUDY GROUP DATATOP INVESTIGATORS. (2009). Urate as a Predictor of the Rate of Clinical Decline in Parkinson Disease. *Archives Of Neurology*, 66(12), 1460-1468.

ASHBURNER, M.; BALL, C.; BLAKE, J.; BOTSTEIN, D.; BUTLER, H.; CHERRY, JM; DAVIS, AP.; DOLINSKI, K.; DWIGHT, SS.; EPPIG, JT.; HARRIS, MA.; HILL, DP.; ISSEL-TARVER, L.; KASARSKIS, A.; LEWIS, S.; MATESE, JC.; RICHARDSON, JE.; RINGWALD, M.; RUBIN, GM.; SHERLOCK, G. (2000). Gene Ontology: tool for the unification of biology. *Nature Genetics*, 25(1), 25-29.

BAKSHI, R.; MACKLIN, E.; LOGAN, R.; ZORLU, M.; XIA, N.; CROTTY, G.; CROTTY, GF.; ZHANG, E.; CHEN, X.; ASCHERIO, A.; SCHWARZSCHILD, MA. (2019). Higher urate in LRRK2 mutation carriers resistant to Parkinson disease. *Annals Of Neurology*, 85(4), 593-599.

BARRERA, G.; GHIRINGHELLI, P.; MOSHER, S.; CARO-QUINTERO, A.; MASSART, S.; BELAICH, M. (2018). Las ómicas en el control biológico, en: *Control biológico de fitopatógenos, insectos y ácaros: aplicaciones y perspectivas*. Ed. Agrosavia. Colombia, 950-987.

BEDIA, C. (2018). Experimental Approaches in Omic Sciences. *Comprehensive Analytical Chemistry*, 13-36.

BENITO LEÓN, J. (2018). Epidemiología de la enfermedad de Parkinson en España y su contextualización mundial. *Revista De Neurología*, 66(04), 125.

BOISON, D. (2011). Modulators of Nucleoside Metabolism in the Therapy of Brain Diseases. *Current Topics In Medicinal Chemistry*, 11(8), 1068-1086.

CARRETERO COLOMER, M. (2006). Enfermedad de Parkinson. Medicación sintomática. *Elsevier*, 25(8), 116-118.

CIGUDOSA, J. (2004). La revolución de los microarrays en la investigación biosanitaria: tipos de plataformas, usos y perspectivas en oncología. *Anales Del Sistema Sanitario De Navarra*, 27(1).

FABBRINI, G.; BROTCHE, J.; GRANDAS, F.; NOMOTO, M.; GOETZ, C. (2007). Levodopa-induced dyskinesias. *Movement Disorders*, 22(10), 1379-1389.

FERREIRA GONZÁLEZ, I.; URRÚTIA, G.; ALONSO-COELLO, P. (2011). Revisiones sistemáticas y metaanálisis: bases conceptuales e interpretación. *Revista Española De Cardiología*, 64(8), 688-696.

GARCIA-ESPARCIA, P.; HERNÁNDEZ-ORTEGA, K.; ANSOLEAGA, B.; CARMONA, M.; FERRER, I. (2015). Purine metabolism gene deregulation in Parkinson's disease. *Neuropathology And Applied Neurobiology*, 41(7), 926-940.

GARCÍA-GARCÍA, F. (2016) Methods of functional enrichment analysis in genomic studies. Tesis Doctoral del Departamento de Bioquímica y Biología Molecular. Programa de Doctorado de Biomedicina y Biotecnología. Universitat de València. 190 pp.

GARRETTI, F.; AGALLIU, D.; LINDESTAM ARLEHAMN, C.; SETTE, A.; SULZER, D. (2019). Autoimmunity in Parkinson's Disease: The Role of α -Synuclein-Specific T Cells. *Frontiers In Immunology*, 10.

GRANT, G.; MANDUCHI, E.; STOECKERT, C. (2007). Analysis and Management of Microarray Gene Expression Data. *Current Protocols In Molecular Biology*, 77(1).

HARIZ, G.; LINDBERG, M.; HARIZ, M.; BERGENHEIM, A. (2003). Does the ADL part of the unified Parkinson's disease rating scale measure ADL? An evaluation in patients after pallidotomy and thalamic deep brain stimulation. *Movement Disorders*, 18(4), 373-381.

HENRY, V.; BANDROWSKI, A.; PEPIN, A.; GONZALEZ, B.; DESFEUX, A. (2014). OMICtools: an informative directory for multi-omic data analysis. *Database*, 2014, bau069.

HOFFMAN, J. (2015). Meta-analysis, en: *Biostatistics for Medical and Biomedical Practitioners*. Ed. Elsevier. California, USA: 645-653

HOME - GEO - NCBI. (2020). Retrieved 17 March 2020, from <https://www.ncbi.nlm.nih.gov/geo/>

HUNG, J.; YANG, T.; HU, Z.; WENG, Z.; DELISI, C. (2011). Gene set enrichment analysis: performance evaluation and usage guidelines. *Briefings In Bioinformatics*, 13(3), 281-291.

JANELIDZE, S.; LINDQVIST, D.; FRANCARDO, V.; HALL, S.; ZETTERBERG, H.; & BLENNOW, K.; Adler, CH.; Beach, TG.; Serrano, GE.; van Westen, D.; Londos, E.; Cenci, MA.; Hansson, O. (2015). Increased CSF biomarkers of angiogenesis in Parkinson disease. *Neurology*, 85(21), 1834-1842.

JESÚS, S., PÉREZ, I., CÁCERES-REDONDO, M., CARRILLO, F., CARBALLO, M., GÓMEZ-GARRE, P., & MIR, P. (2012). Low serum uric acid concentration in Parkinson's disease in southern Spain. *European Journal Of Neurology*, 20(1), 208-210.

KALIA, L.; LANG, A. (2015). Parkinson's disease. *The Lancet*, 386(9996), 896-912.

KANEHISA, M. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research*, 28(1), 27-30.

- KOROTKEVICH, G.; SUKHOV, V.; SERGUSHICHEV, A. (2016). Fast gene set enrichment analysis. *bioRxiv*
- KUMAGAI, T.; NAGAYAMA, H.; OTA, T.; NISHIYAMA, Y.; MISHINA, M.; UEDA, M. (2014). Sex Differences in the Pharmacokinetics of Levodopa in Elderly Patients With Parkinson Disease. *Clinical Neuropharmacology*, 37(6), 173-176.
- LAVIN, K.; SEALFON, S.; MCDONALD, M.; ROBERTS, B.; WILK, K.; NAIR, V.; Ge, Y.; LAKSHMAN KUMAR, P.; WINDHAM, ST.; BAMMAN, MM. (2020). Skeletal muscle transcriptional networks linked to type I myofiber grouping in Parkinson's disease. *Journal Of Applied Physiology*, 128(2), 229-240.
- LESNICK, TG.; PAPAPETROPOULOS, S.; MASH, D.; FFRENCH-MULLEN, J.; SHEHADEH, L.; DE ANDRADE, M.; HENLEY, JR.; ROCCA, WA.; AHLKOG, JE.; MARAGANORE, DM. (2007). A Genomic Pathway Approach to a Complex Disease: Axon Guidance and Parkinson Disease. *Plos Genetics*, 3(6), e98.
- MACKENZIE, R.J. (2018) RNA-seq: Basics, Applications and Protocol. Genomics Research form Technology Networks
- MAITI, P.; MANNA, J.; DUNBAR, G. (2017). Current understanding of the molecular mechanisms in Parkinson's disease: Targets for potential treatments. *Translational Neurodegeneration*, 6(28).
- MALMIERCA, P. (2019) Metafun: herramienta web para la integración y caracterización funcional de estudios ómicos con técnicas de metaanálisis. Trabajo Final de Máster. Máster Universitario en Bioinformática, Universitat de València. pp 60.
- MANTEROLA, C.; ASTUDILLO, P.; ARIAS, E.; CLAROS, N. (2013). Revisiones sistemáticas de la literatura. Qué se debe saber acerca de ellas. *Cirugía Española*, 91(3), 149-155.
- MANZONI, C.; KIA, D.; VANDROVCOVA, J.; HARDY, J.; WOOD, N.; LEWIS, P.; FERRARI, R. (2016). Genome, transcriptome and proteome: the rise of omics data and their integration in biomedical sciences. *Briefings In Bioinformatics*, 19(2), 286-302.
- MARTIN, I.; KIM, J.; DAWSON, V.; DAWSON, T. (2016). LRRK2 pathobiology in Parkinson's disease - virtual inclusion. *Journal Of Neurochemistry*, 139, 75-76.
- MIRAGLIA, F.; COLLA, E. (2019). Microbiome, Parkinson's Disease and Molecular Mimicry. *Cells*, 8(3), 222.
- MORENO, V.; SOLÉ, X. (2004). Uso de chips de ADN (microarrays) en medicina: fundamentos técnicos y procedimientos básicos para el análisis estadístico de resultados. *Medicina Clínica*, 122(1), 73-79.
- NAKABEPPU, Y.; TSUCHIMOTO, D.; YAMAGUCHI, H.; SAKUMI, K. (2007). Oxidative damage in nucleic acids and Parkinson's disease. *Journal Of Neuroscience Research*, 85(5), 919-934.

NICOLETTI, A.; VASTA, R.; MOSTILE, G.; NICOLETTI, G.; ARABIA, G.; ILICETO, G.; LAMBERTI, P.; MARCONI, R.; MORGANTE, L.; BARONE, P.; QUATTRONE, A.; ZAPPIA, M. (2017). Gender effect on non-motor symptoms in Parkinson's disease: are men more at risk?. *Parkinsonism & Related Disorders*, 35, 69-74.

NISSEN, S.; SHRIVASTAVA, K.; SCHULTE, C.; OTZEN, D.; GOLDECK, D.; BERG, D. MØLLER, H.J.; MAETZLER, W.; ROMERO-RAMOS, M. (2019). Alterations in Blood Monocyte Functions in Parkinson's Disease. *Movement Disorders*, 34(11), 1711-1721.

PICILLO, M.; NICOLETTI, A.; FETONI, V.; GARAVAGLIA, B.; BARONE, P.; PELLECCIA, M. (2017). The relevance of gender in Parkinson's disease: a review. *Journal Of Neurology*, 264(8), 1583-1607.

RODRIGO RUIZ, D.; PÉREZ GONZÁLEZ, J.; CEJUDO, J. (2017). Dificultades de reconocimiento emocional facial como déficit primario en niños con trastorno por déficit de atención/hiperactividad: revisión sistemática. *Revista De Neurología*, 65(04), 145.

RUSSO, M.W (2007). How to Review a Meta-analysis. *Gastroenterology & hepatology*, 3(8), 637-642.

SCHWARZSCHILD, M.; SCHWID, SR.; MAREK, K.; WATTS, A.; LANG, A. E.; OAKES, D.; SHOULSON, I.; ASCHERIO, A.; PARKINSON STUDY GROUP PRECEPT INVESTIGATORS; HYSON, C.; GORBOLD, E.; RUDOLPH, A.; KIEBURTZ, K.; FAHN, S.; GAUGER, L.; GOETZ, C.; SEIBYL, J.; FORREST, M.; ONDRASIK, J. (2008). Serum Urate as a Predictor of Clinical and Radiographic Progression in Parkinson Disease. *Archives Of Neurology*, 65(6), 716-723.

TIPNEY, H.; HUNTER, L. (2010). An introduction to effective use of enrichment analysis software. *Human Genomics*, 4(3), 202-206.

WANG, Z.; GERSTEIN, M.; SNYDER, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics*, 10(1), 57-63.

ZHANG, Y.; JAMES, M.; MIDDLETON, F.; DAVIS, R. (2005). Transcriptional analysis of multiple brain regions in Parkinson's disease supports the involvement of specific protein processing, energy metabolism, and signaling pathways, and suggests novel disease mechanisms. *American Journal Of Medical Genetics Part B: Neuropsychiatric Genetics*, 137B(1), 5-16.