

CARACTERIZACIÓN DE LAS DIFERENCIAS DE SEXO EN ESCLEROSIS MÚLTIPLE MEDIANTE ESTRATEGIAS BASADAS EN EL ANÁLISIS MASIVO DE DATOS

Trabajo de fin de máster curso 2020 – 2021

Máster en bioinformática

Autora: Irene Soler Sáez

Tutores: Francisco García García

Zoraida Andreu Martínez

Vicente Arnau Llombart

ÍNDICE

1. INTRODUCCIÓN

2. OBJETIVOS

3. MATERIALES Y MÉTODOS

4. RESULTADOS

5. CONCLUSIONES

INTRODUCCIÓN

OBJETIVOS

MATERIALES Y
MÉTODOS

RESULTADOS

CONCLUSIONES

1. INTRODUCCIÓN

1.1. Esclerosis múltiple

1.2. Tecnologías de alto rendimiento

INTRODUCCIÓN

OBJETIVOS

MATERIALES Y
MÉTODOS

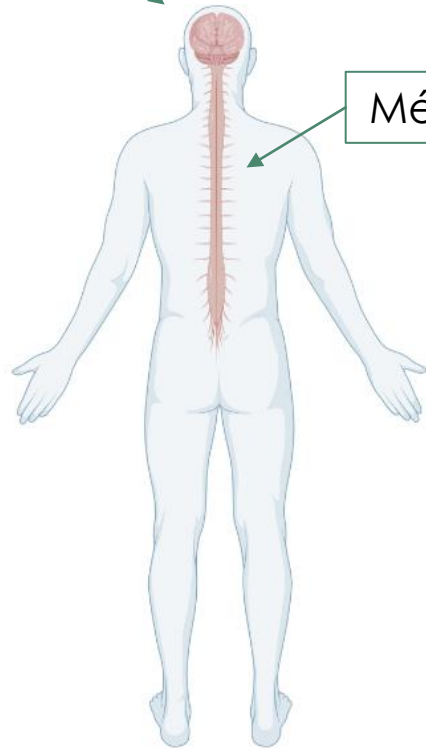
RESULTADOS

CONCLUSIONES

Esclerosis múltiple

Tecnologías de alto rendimiento

Encéfalo



Médula espinal

Sistema nervioso central

INTRODUCCIÓN

OBJETIVOS

MATERIALES Y
MÉTODOS

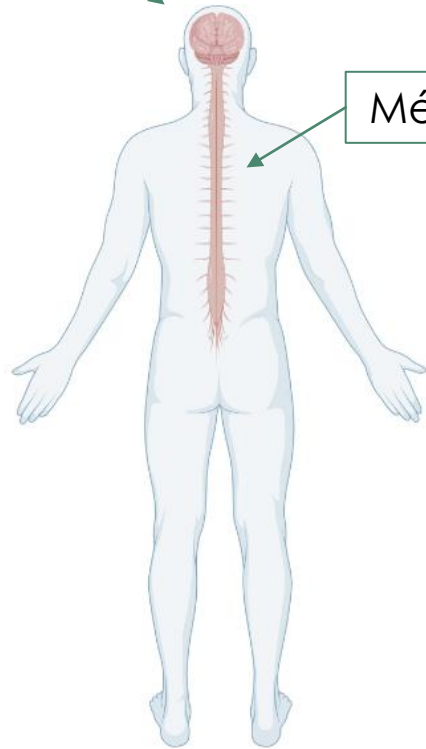
RESULTADOS

CONCLUSIONES

Esclerosis múltiple

Tecnologías de alto rendimiento

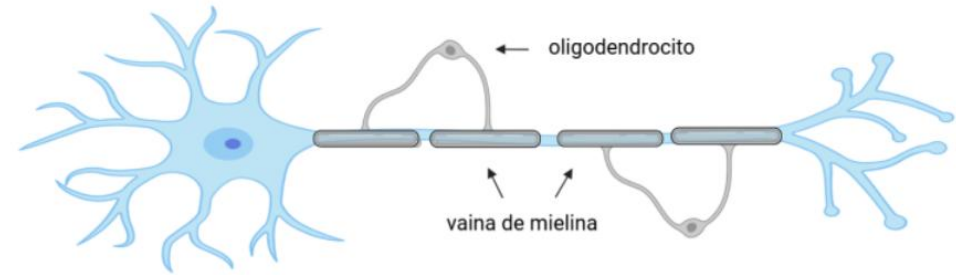
Encéfalo



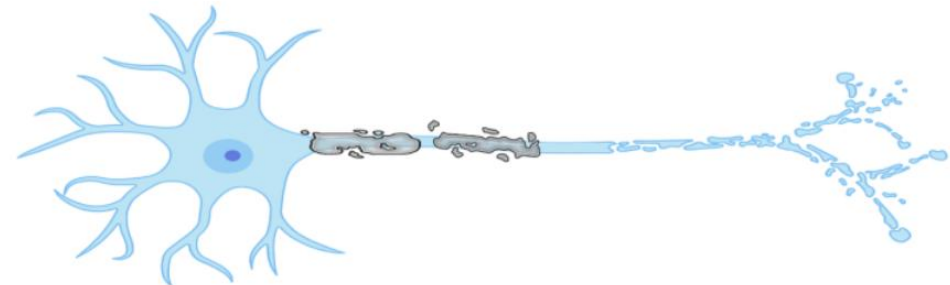
Médula espinal

Sistema nervioso central

Individuo sano



Individuo con esclerosis múltiple



INTRODUCCIÓN

OBJETIVOS

MATERIALES Y
MÉTODOS

RESULTADOS

CONCLUSIONES

Esclerosis múltiple

Tecnologías de alto rendimiento

Diferencias en función del sexo de los individuos

INTRODUCCIÓN

OBJETIVOS

MATERIALES Y
MÉTODOS

RESULTADOS

CONCLUSIONES

Esclerosis múltiple

Tecnologías de alto rendimiento

Diferencias en función del sexo de los individuos

Prevalencia

INTRODUCCIÓN

OBJETIVOS

MATERIALES Y
MÉTODOS

RESULTADOS

CONCLUSIONES

Esclerosis múltiple

Tecnologías de alto rendimiento

Diferencias en función del sexo de los individuos

Prevalencia

Progresión

INTRODUCCIÓN

OBJETIVOS

MATERIALES Y
MÉTODOS

RESULTADOS

CONCLUSIONES

Esclerosis múltiple

Tecnologías de alto rendimiento

Diferencias en función del sexo de los individuos

Prevalencia

Progresión

Cromosomas sexuales

INTRODUCCIÓN

OBJETIVOS

MATERIALES Y
MÉTODOS

RESULTADOS

CONCLUSIONES

Esclerosis múltiple

Tecnologías de alto rendimiento

Diferencias en función del sexo de los individuos

Prevalencia

Progresión

Cromosomas sexuales

Hormonas sexuales

INTRODUCCIÓN

OBJETIVOS

MATERIALES Y
MÉTODOS

RESULTADOS

CONCLUSIONES

Esclerosis múltiple

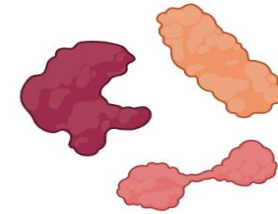
Tecnologías de alto rendimiento



Genómica → ADN



Transcriptómica → ARN



Proteómica → Proteínas

INTRODUCCIÓN

OBJETIVOS

MATERIALES Y
MÉTODOS

RESULTADOS

CONCLUSIONES

Esclerosis múltiple

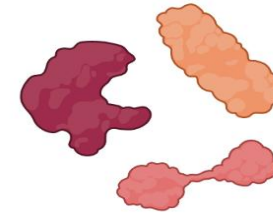
Tecnologías de alto rendimiento



Genómica → ADN



Transcriptómica → ARN



Proteómica → Proteínas

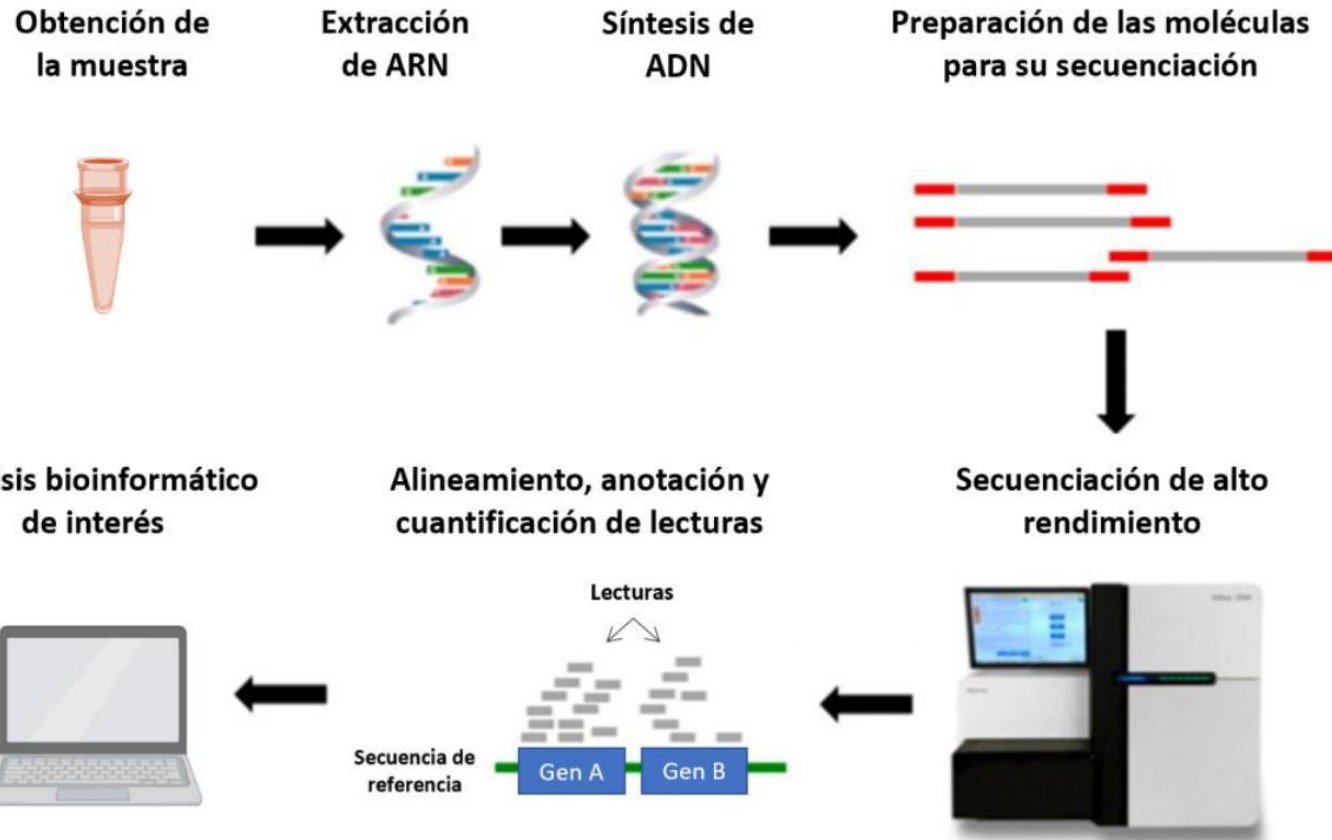
Célula única

Núcleo único

Esclerosis múltiple

Tecnologías de alto rendimiento

Procedimiento de secuenciación de ARN



INTRODUCCIÓN

OBJETIVOS

MATERIALES Y
MÉTODOS

RESULTADOS

CONCLUSIONES

Mejorar la caracterización a nivel molecular de **las diferencias de sexo** presentes en la enfermedad de **esclerosis múltiple** analizando datos procedentes de la **secuenciación de ARN de células y núcleos individuales**

INTRODUCCIÓN

OBJETIVOS

MATERIALES Y
MÉTODOS

RESULTADOS

CONCLUSIONES

Mejorar la caracterización a nivel molecular de **las diferencias de sexo** presentes en la enfermedad de **esclerosis múltiple** analizando datos procedentes de la **secuenciación de ARN de células y núcleos individuales**

OBJETIVOS ESPECÍFICOS:

1. Comprensión de la enfermedad de esclerosis múltiple con perspectiva de sexo y del análisis bioinformático de datos de secuenciación de ARN de célula única
2. Revisión sistemática
3. Análisis primario
4. Integración de los resultados

INTRODUCCIÓN

OBJETIVOS

**MATERIALES Y
MÉTODOS**

RESULTADOS

CONCLUSIONES

3. MATERIALES Y MÉTODOS

3.1. Revisión sistemática

3.2. Análisis primario

3.3. Integración de los resultados

INTRODUCCIÓN

OBJETIVOS

**MATERIALES Y
MÉTODOS**

RESULTADOS

CONCLUSIONES

Revisión sistemática

Análisis primario

Integración de los resultados

Directrices PRISMA



1. IDENTIFICACIÓN

Criterios de inclusión



Bases de datos
públicas

INTRODUCCIÓN

OBJETIVOS

**MATERIALES Y
MÉTODOS**

RESULTADOS

CONCLUSIONES

Revisión sistemática

Análisis primario

Integración de los resultados

Directrices PRISMA



1. IDENTIFICACIÓN



2. REVISIÓN



Criterios de inclusión



Bases de datos
públicas

Criterios de exclusión:

1. Metodología
2. Patología evaluada
3. Diseño experimental
4. Sexo
5. Tamaño muestral

INTRODUCCIÓN

OBJETIVOS

**MATERIALES Y
MÉTODOS**

RESULTADOS

CONCLUSIONES

Revisión sistemática

Análisis primario

Integración de los resultados

Directrices PRISMA



1. IDENTIFICACIÓN

Criterios de inclusión



Bases de datos
públicas



2. REVISIÓN

Criterios de exclusión:

1. Metodología
2. Patología evaluada
3. Diseño experimental
4. Sexo
5. Tamaño muestral



3. ELEGIBILIDAD

Criterios de exclusión

INTRODUCCIÓN

OBJETIVOS

**MATERIALES Y
MÉTODOS**

RESULTADOS

CONCLUSIONES

Revisión sistemática

Análisis primario

Integración de los resultados

Directrices PRISMA



1. IDENTIFICACIÓN

Criterios de inclusión



Bases de datos
públicas



2. REVISIÓN

Criterios de exclusión:

1. Metodología
2. Patología evaluada
3. Diseño experimental
4. Sexo
5. Tamaño muestral



3. ELEGIBILIDAD

Criterios de exclusión



4. INCLUSIÓN

Descarga de datos

INTRODUCCIÓN

OBJETIVOS

**MATERIALES Y
MÉTODOS**

RESULTADOS

CONCLUSIONES

Revisión sistemática

Análisis primario

Integración de los resultados

Preprocesamiento

Control de calidad

Normalización

Selección genes altamente variables

Reducción de la dimensionalidad

Agrupamiento por identidad celular

Anotación de los tipos celulares

Evaluación de los genes marcadores

Análisis bioestadísticos

INTRODUCCIÓN

OBJETIVOS

**MATERIALES Y
MÉTODOS**

RESULTADOS

CONCLUSIONES

Revisión sistemática

Análisis primario

Integración de los resultados

Preprocesamiento

Control de calidad

Normalización

Selección genes altamente variables

Reducción de la dimensionalidad

Agrupamiento por identidad celular

Anotación de los tipos celulares

Evaluación de los genes marcadores

Análisis bioestadísticos

Matriz de conteos

SingleCellExperiment

Metadatos de células

Metadatos de genes

INTRODUCCIÓN

OBJETIVOS

**MATERIALES Y
MÉTODOS**

RESULTADOS

CONCLUSIONES

Revisión sistemática

Análisis primario

Integración de los resultados

Preprocesamiento

Control de calidad

Normalización

Selección genes altamente variables

Reducción de la dimensionalidad

Agrupamiento por identidad celular

Anotación de los tipos celulares

Evaluación de los genes marcadores

Análisis bioestadísticos

① Cálculo de indicadores

② Representación gráfica

③ Establecimiento de puntos de corte

④ Filtrado

INTRODUCCIÓN

OBJETIVOS

**MATERIALES Y
MÉTODOS**

RESULTADOS

CONCLUSIONES

Revisión sistemática

Análisis primario

Integración de los resultados

Preprocesamiento

Control de calidad

Normalización

Selección genes altamente variables

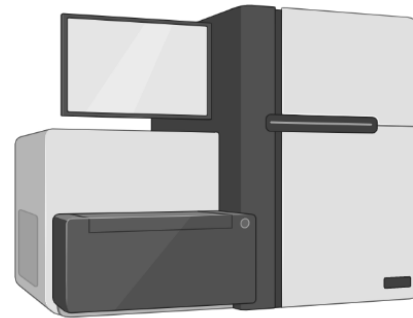
Reducción de la dimensionalidad

Agrupamiento por identidad celular

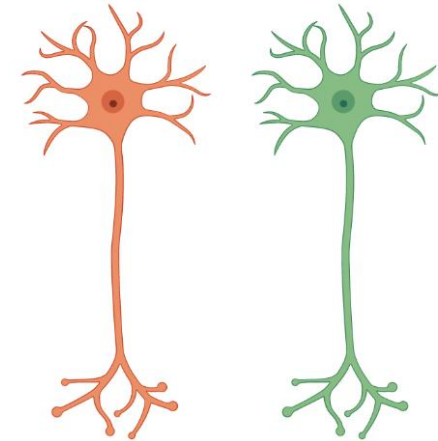
Anotación de los tipos celulares

Evaluación de los genes marcadores

Análisis bioestadísticos



**Variabilidad
técnica**



**Variabilidad
biológica**

INTRODUCCIÓN

OBJETIVOS

**MATERIALES Y
MÉTODOS**

RESULTADOS

CONCLUSIONES

Revisión sistemática

Análisis primario

Integración de los resultados

Preprocesamiento

Control de calidad

Normalización

Selección genes altamente variables

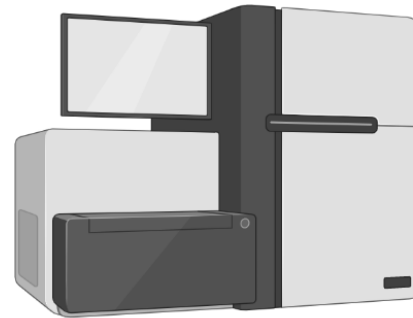
Reducción de la dimensionalidad

Agrupamiento por identidad celular

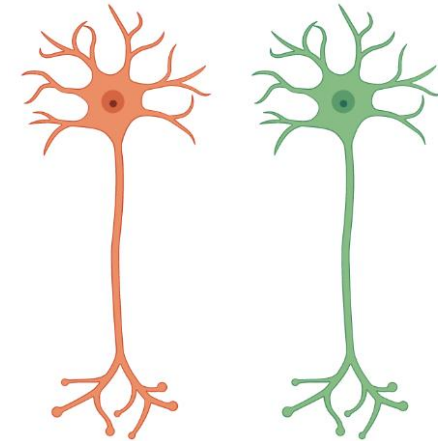
Anotación de los tipos celulares

Evaluación de los genes marcadores

Análisis bioestadísticos



Variabilidad
técnica



**Variabilidad
biológica**

INTRODUCCIÓN

OBJETIVOS

**MATERIALES Y
MÉTODOS**

RESULTADOS

CONCLUSIONES

Revisión sistemática

Análisis primario

Integración de los resultados

Preprocesamiento

Control de calidad

Normalización

Selección genes altamente variables

Reducción de la dimensionalidad

Agrupamiento por identidad celular

Anotación de los tipos celulares

Evaluación de los genes marcadores

Análisis bioestadísticos

Gen 1

Gen 4

Gen 10

Gen 5

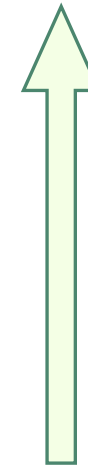
Gen 2

Gen 18

Gen 9

Gen 7

...



Variabilidad biológica

INTRODUCCIÓN

OBJETIVOS

**MATERIALES Y
MÉTODOS**

RESULTADOS

CONCLUSIONES

Revisión sistemática

Análisis primario

Integración de los resultados

Preprocesamiento

Control de calidad

Normalización

Selección genes altamente variables

Reducción de la dimensionalidad

Agrupamiento por identidad celular

Anotación de los tipos celulares

Evaluación de los genes marcadores

Análisis bioestadísticos

20%

Gen 1

Gen 4

Gen 10

Gen 5

Gen 2

Gen 18

Gen 9

Gen 7

...

Variabilidad biológica

INTRODUCCIÓN

OBJETIVOS

**MATERIALES Y
MÉTODOS**

RESULTADOS

CONCLUSIONES

Revisión sistemática

Análisis primario

Integración de los resultados

Preprocesamiento

Control de calidad

Normalización

Selección genes altamente variables

Reducción de la dimensionalidad

Agrupamiento por identidad celular

Anotación de los tipos celulares

Evaluación de los genes marcadores

Análisis bioestadísticos

Resumir la variabilidad

PCA (*Principal Component Analysis*)

Visualización

tSNE (*t-distributed Stochastic Neighbor Embedding*)

UMAP (*Uniform Manifold Approximation and Projection*)

INTRODUCCIÓN

OBJETIVOS

**MATERIALES Y
MÉTODOS**

RESULTADOS

CONCLUSIONES

Revisión sistemática

Análisis primario

Integración de los resultados

Preprocesamiento

Control de calidad

Normalización

Selección genes altamente variables

Reducción de la dimensionalidad

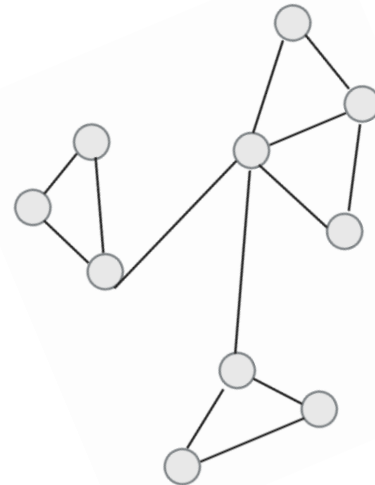
Agrupamiento por identidad celular

Anotación de los tipos celulares

Evaluación de los genes marcadores

Análisis bioestadísticos

**Construcción
de un grafo**



INTRODUCCIÓN

OBJETIVOS

**MATERIALES Y
MÉTODOS**

RESULTADOS

CONCLUSIONES

Revisión sistemática

Análisis primario

Integración de los resultados

Preprocesamiento

Control de calidad

Normalización

Selección genes altamente variables

Reducción de la dimensionalidad

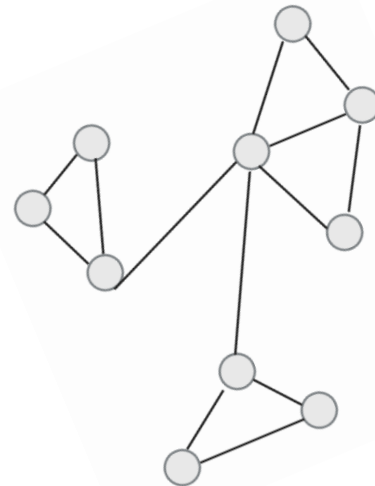
Agrupamiento por identidad celular

Anotación de los tipos celulares

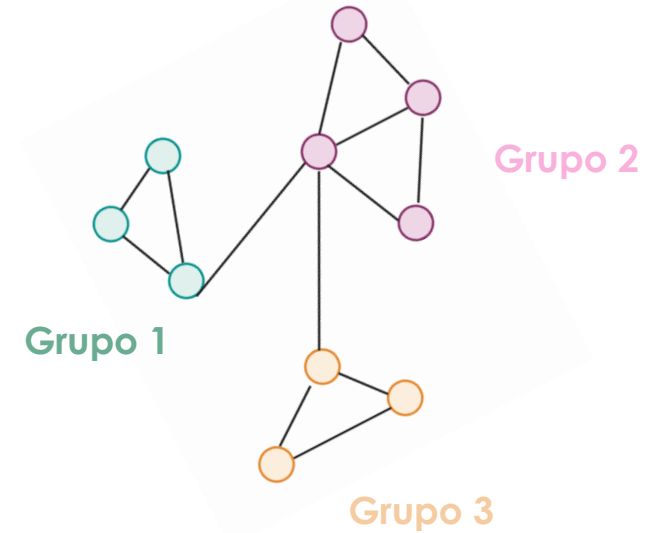
Evaluación de los genes marcadores

Análisis bioestadísticos

**Construcción
de un grafo**



**Búsqueda de
comunidades**



INTRODUCCIÓN

OBJETIVOS

**MATERIALES Y
MÉTODOS**

RESULTADOS

CONCLUSIONES

Revisión sistemática

Análisis primario

Integración de los resultados

Preprocesamiento

Control de calidad

Normalización

Selección genes altamente variables

Reducción de la dimensionalidad

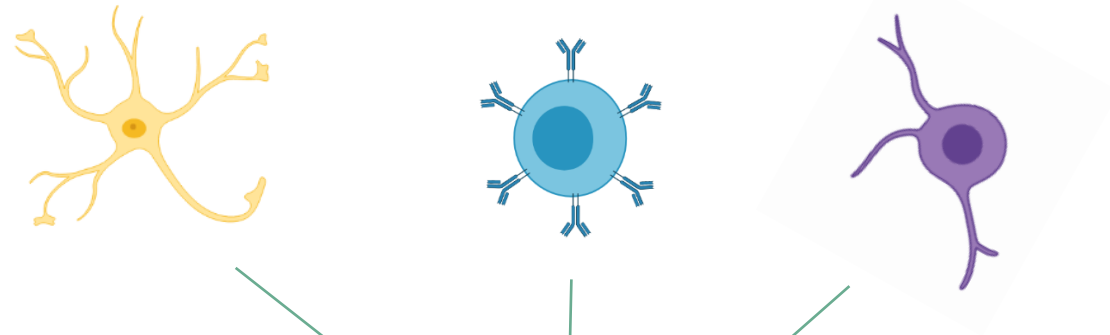
Agrupamiento por identidad celular

Anotación de los tipos celulares

Evaluación de los genes marcadores

Análisis bioestadísticos

Células de referencia



Célula analizada

INTRODUCCIÓN

OBJETIVOS

**MATERIALES Y
MÉTODOS**

RESULTADOS

CONCLUSIONES

Revisión sistemática

Análisis primario

Integración de los resultados

Preprocesamiento

Control de calidad

Normalización

Selección genes altamente variables

Reducción de la dimensionalidad

Agrupamiento por identidad celular

Anotación de los tipos celulares

Evaluación de los genes marcadores

Análisis bioestadísticos

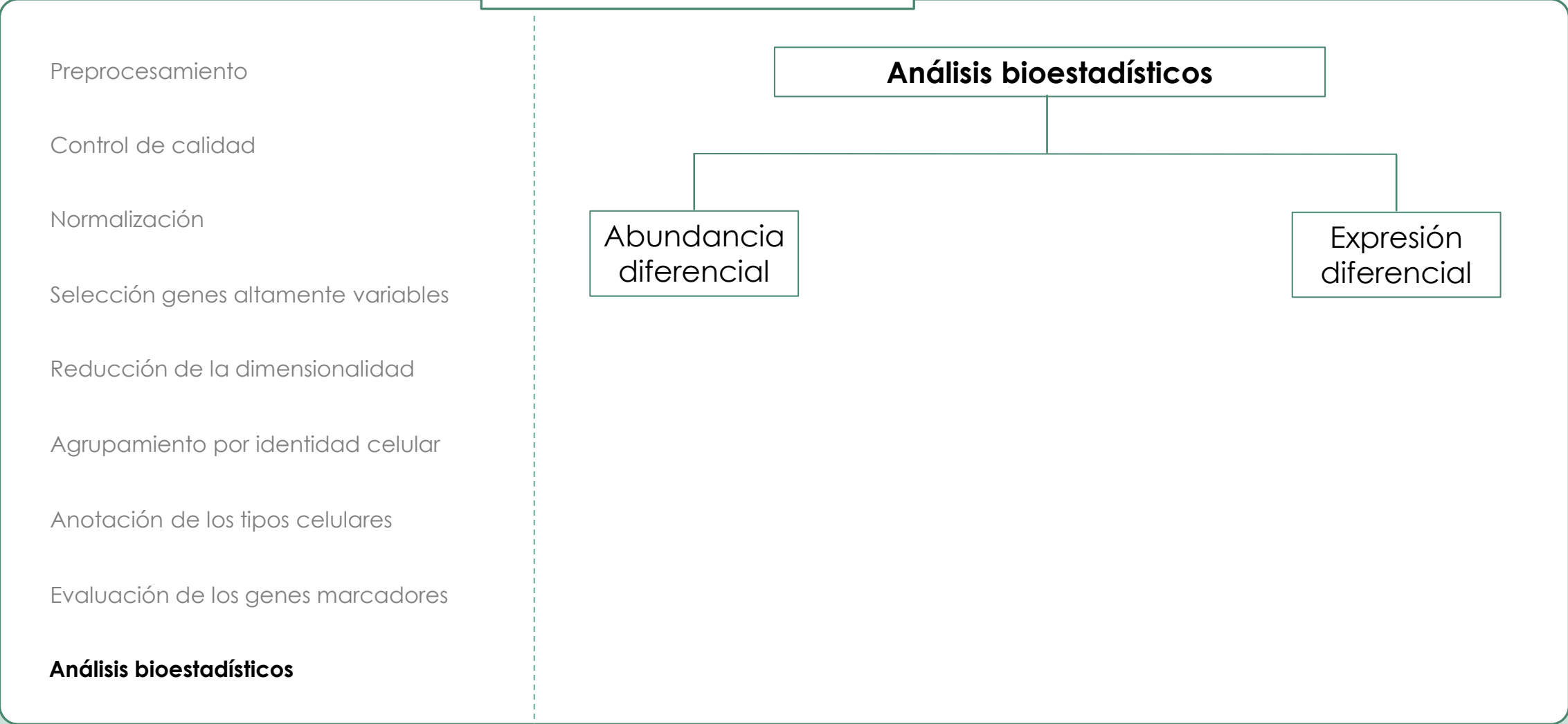
- 1 Expresión diferencial entre tipos celulares
- 2 Búsqueda bibliográfica



Revisión sistemática

Análisis primario

Integración de los resultados



INTRODUCCIÓN

OBJETIVOS

**MATERIALES Y
MÉTODOS**

RESULTADOS

CONCLUSIONES

Revisión sistemática

Análisis primario

Integración de los resultados

Preprocesamiento

Control de calidad

Normalización

Selección genes altamente variables

Reducción de la dimensionalidad

Agrupamiento por identidad celular

Anotación de los tipos celulares

Evaluación de los genes marcadores

Análisis bioestadísticos

Análisis bioestadísticos

Abundancia
diferencial

Expresión
diferencial

Comparaciones:

- ✓ EM_Mujer – Control_Mujer
- ✓ EM_Hombre – Control_Hombre
- ✓ $(EM_Mujer - Control_Mujer) - (EM_Hombre - Control_Hombre)$

EM: esclerosis múltiple

INTRODUCCIÓN

OBJETIVOS

**MATERIALES Y
MÉTODOS**

RESULTADOS

CONCLUSIONES

Revisión sistemática

Análisis primario

Integración de los resultados

Genes significativos comunes:

① por estudio en todos los tipos celulares

② por tipo celular en todos los estudios

INTRODUCCIÓN

OBJETIVOS

**MATERIALES Y
MÉTODOS**

RESULTADOS

CONCLUSIONES

Revisión sistemática

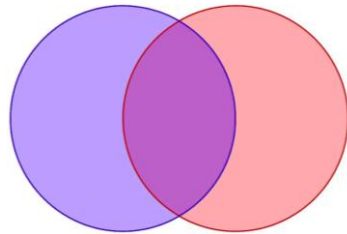
Análisis primario

Integración de los resultados

Genes significativos comunes:

① por estudio en todos los tipos celulares

② por tipo celular en todos los estudios



Intersección



Open Targets



GENEONTOLOGY
Unifying Biology

INTRODUCCIÓN

OBJETIVOS

MATERIALES Y
MÉTODOS

RESULTADOS

CONCLUSIONES

4. RESULTADOS

4.1. Revisión sistemática

4.2. Análisis primario

4.3. Integración de los resultados

INTRODUCCIÓN

OBJETIVOS

MATERIALES Y
MÉTODOS

RESULTADOS

CONCLUSIONES

Revisión sistemática

Análisis primario

Integración de los resultados

IDENTIFICACIÓN

Estudios identificados en
GEO (n = 24)

Estudios identificados en
ArrayExpress (n = 2)

Estudios identificados en
UCSC Cell Browser (n = 2)

Estudios identificados en el
buscador Google (n = 5)

REVISIÓN

Estudios tras la eliminación
de duplicados (n = 31)

Estudios excluidos:
- Metodología (n = 14)
- Enfermedad evaluada (n = 4)
- Diseño experimental (n = 0)
- Sexo (n = 0)
- Tamaño muestral (n = 6)

ELEGIBILIDAD

Estudios elegibles (n = 7)

Estudios excluidos:
- Presencia de pseudo-controles (n = 3)
- No cumplir con el tamaño muestral tras la solicitud de los datos (n = 1)
- No disponibilidad de los datos (n = 1)

INCLUSIÓN

Estudios incluidos en el
análisis exploratorio (n = 2)

Estudios incluidos en el
análisis (n = 2)

INTRODUCCIÓN

OBJETIVOS

MATERIALES Y
MÉTODOS

RESULTADOS

CONCLUSIONES

Revisión sistemática

Análisis primario

Integración de los resultados

Estudio	Metodología	Tipo de muestra
<i>Multiple sclerosis</i>	núcleo único	tejido nervioso
Cohorte 1 GSE144744	célula única	sangre periférica
Cohorte 3 GSE144744	célula única	sangre periférica

INTRODUCCIÓN

OBJETIVOS

MATERIALES Y
MÉTODOS

RESULTADOS

CONCLUSIONES

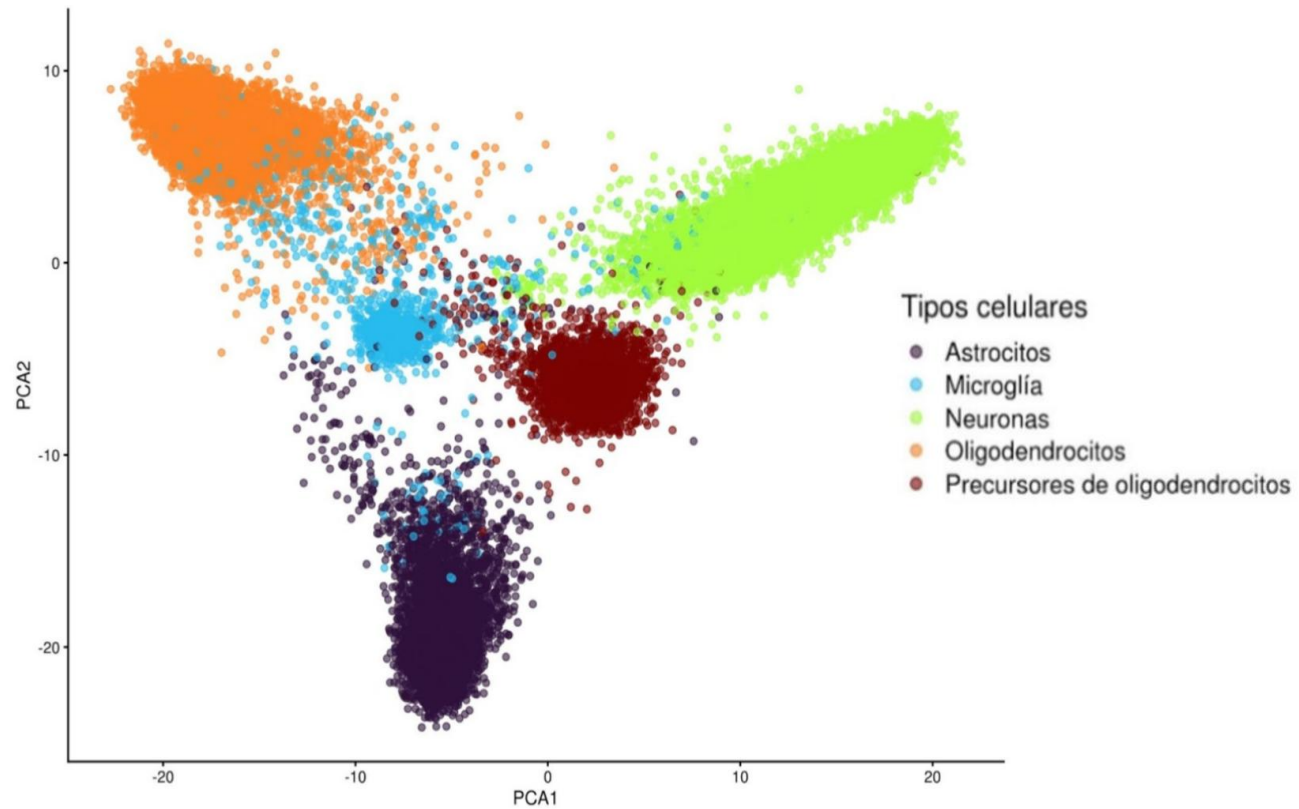
Revisión sistemática

Análisis primario

Integración de los resultados

Anotación de los tipos celulares

Tejido nervioso:



INTRODUCCIÓN

OBJETIVOS

MATERIALES Y
MÉTODOS

RESULTADOS

CONCLUSIONES

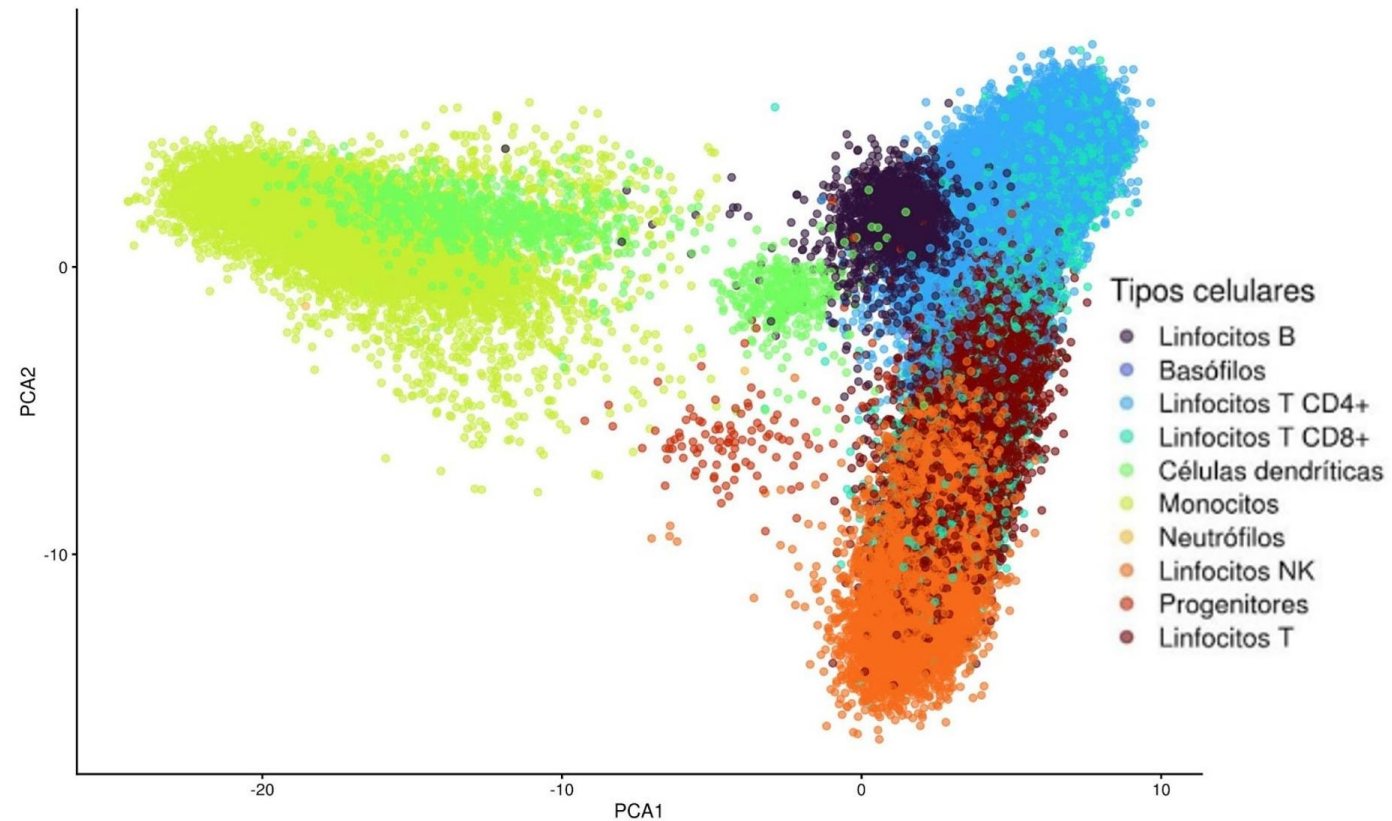
Revisión sistemática

Análisis primario

Integración de los resultados

Anotación de los tipos celulares

Sangre periférica:



INTRODUCCIÓN

OBJETIVOS

MATERIALES Y
MÉTODOS

RESULTADOS

CONCLUSIONES

Revisión sistemática

Análisis primario

Integración de los resultados

Análisis bioestadísticos

Comparaciones:

- ✓ EM_Mujer – Control_Mujer
- ✓ EM_Hombre – Control_Hombre
- ✓ (EM_Mujer – Control_Mujer) – (EM_Hombre – Control_Hombre)

EM: esclerosis múltiple

INTRODUCCIÓN

OBJETIVOS

MATERIALES Y
MÉTODOS

RESULTADOS

CONCLUSIONES

Revisión sistemática

Análisis primario

Integración de los resultados

Análisis bioestadísticos

Comparaciones:

✓ EM_Mujer – Control_Mujer

✓ EM_Hombre – Control_Hombre

✓ **(EM_Mujer – Control_Mujer) – (EM_Hombre – Control_Hombre)**

EM: esclerosis múltiple

INTRODUCCIÓN

OBJETIVOS

MATERIALES Y
MÉTODOS

RESULTADOS

CONCLUSIONES

Revisión sistemática

Análisis primario

Integración de los resultados

Abundancia diferencial

Estudio	Método	Tipos celulares (signo logFC)
<i>Multiple sclerosis</i>	1	-
	2	oligodendrocitos (+)
Cohorte 1 GSE144744	1	-
	2	-
Cohorte 3 GSE144744	1	linfocitos NK (-)
	2	linfocitos NK (-)

INTRODUCCIÓN

OBJETIVOS

MATERIALES Y
MÉTODOS

RESULTADOS

CONCLUSIONES

Revisión sistemática

Análisis primario

Integración de los resultados

Expresión diferencial

Astroцитos	Microglía	Neuronas
742	428	4903

Oligodendrocitos	Precursores de oligodendrocitos
286	216

INTRODUCCIÓN

OBJETIVOS

MATERIALES Y
MÉTODOS

RESULTADOS

CONCLUSIONES

Revisión sistemática

Análisis primario

Integración de los resultados

Integración por estudio

Estudio	logFC	Genes comunes
<i>Multiple sclerosis</i>	+	<i>CADM2, PCDH9, UBA6-AS1</i>

- *Adherens junction organization*
- *Homophilic cell adhesion via plasma membrane adhesion molecules*
- *Cell-cell junction organization*
- *Cell-cell adhesion via plasma-membrane adhesion molecules*

INTRODUCCIÓN

OBJETIVOS

MATERIALES Y
MÉTODOS

RESULTADOS

CONCLUSIONES

Revisión sistemática

Análisis primario

Integración de los resultados

Integración por estudio

Estudio	logFC	Genes comunes
<i>Multiple sclerosis</i>	-	<i>ACTB, BEX1, GAPDH, LINGO1, NDRG4, NRGN, TMSB10, TMSB4X, TUBA1B, TUBA4A, VSNL1</i>

➤ *Sequestering of actin monomers*

INTRODUCCIÓN

OBJETIVOS

MATERIALES Y
MÉTODOS

RESULTADOS

CONCLUSIONES

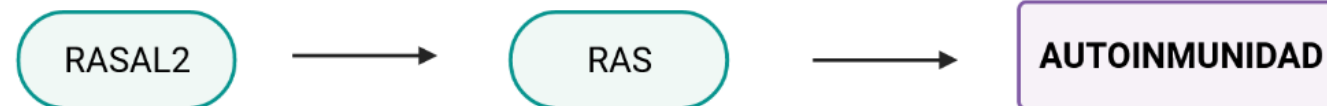
Revisión sistemática

Análisis primario

Integración de los resultados

Integración por estudio

Estudio	logFC	Genes comunes
Cohorte 1 GSE144744	+	RASAL2
	-	-



INTRODUCCIÓN

OBJETIVOS

MATERIALES Y
MÉTODOS

RESULTADOS

CONCLUSIONES

Revisión sistemática

Análisis primario

Integración de los resultados

Integración por estudio

Estudio	logFC	Genes comunes
Cohorte 3 GSE144744	+	<i>EGR1</i>
	-	<i>FAM118A, HLA-A, HLA-DRB5</i>

INTRODUCCIÓN

OBJETIVOS

MATERIALES Y
MÉTODOS

RESULTADOS

CONCLUSIONES

Revisión sistemática

Análisis primario

Integración de los resultados

Integración por estudio

Estudio	logFC	Genes comunes
Cohorte 3 GSE144744	+	<i>EGR1</i>
	-	<i>FAM118A, HLA-A, HLA-DRB5</i>

- *Regulation of leukocyte proliferation*
- *Antigen processing and presentation*
- *Cytokine production involved in immune response*

INTRODUCCIÓN

OBJETIVOS

MATERIALES Y
MÉTODOS

RESULTADOS

CONCLUSIONES

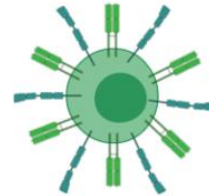
Revisión sistemática

Análisis primario

Integración de los resultados

Integración por tipo celular

LINFOCITOS T CD4+
(linfocitos T cooperadores)



logFC	Genes comunes	Genes comunes asociados	Funciones biológicas
+	138	47	14
-	207	56	0

INTRODUCCIÓN

OBJETIVOS

MATERIALES Y
MÉTODOS

RESULTADOS

CONCLUSIONES

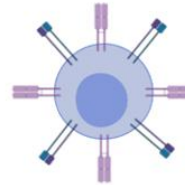
Revisión sistemática

Análisis primario

Integración de los resultados

Integración por tipo celular

LINFOCITOS T CD8+
(linfocitos T citotóxicos)



logFC	Genes comunes	Genes comunes asociados	Funciones biológicas
+	44	21	5
-	36	8	0

INTRODUCCIÓN

OBJETIVOS

MATERIALES Y
MÉTODOS

RESULTADOS

CONCLUSIONES

1. El resultado de la revisión sistemática ha puesto de manifiesto la necesidad de incluir suficiente tamaño muestral para poder analizar la **perspectiva de sexo** en los estudios biomédicos.
2. En base a la bibliografía revisada, este es el **primer trabajo** que evalúa las diferencias de sexo en la enfermedad de esclerosis múltiple utilizando datos de secuenciación de ARN de células y núcleos individuales con un enfoque integrativo.
3. Se ha identificado un mayor número de **oligodendrocitos** (en el sistema nervioso) y un menor número de **linfocitos NK** (en sangre periférica) en mujeres enfermas respecto a hombres afectados.
4. Todos los tipos celulares evaluados presentan patrones de **expresión diferencial** en función del sexo del individuo que sufre la enfermedad.

5. En los tipos celulares de **sistema nervioso**, las mujeres presentan sobrerrepresentadas funciones involucradas en la adhesión celular. Por su parte, un proceso biológico relacionado con la actina se encuentra enriquecido en hombres. Asimismo, los hombres disponen de genes diferencialmente expresados descritos previamente en la bibliografía, como *LINGO1*.
6. Los linfocitos T CD4+ de las mujeres presentan sobrerrepresentadas funciones implicadas en el procesamiento de moléculas de ARNm, mientras que los linfocitos T CD8+ en procesos biológicos relacionados con la respuesta a estrés. No se han encontrado funciones significativas para estos tipos celulares en hombres, pero al evaluar el conjunto de **células inmunitarias** se ha identificado el proceso de presentación de antígenos (en base a los genes *HLA-A*, *HLA-DRB5*).

*¡ MUCHAS GRACIAS POR VUESTRA
ATENCIÓN !*