

**MÁSTER UNIVERSITARIO EN BIOINFORMÁTICA**



VNIVERSITAT  
E VALÈNCIA

**TRABAJO DE FIN DE MÁSTER**

**CARACTERIZACIÓN DE LAS DIFERENCIAS DE SEXO EN  
CÁNCER DE PÁNCREAS DUCTAL MEDIANTE EL ANÁLISIS  
MASIVO DE DATOS TRANSCRIPTÓMICOS**

**AUTORA:  
LUCÍA FANTIN**

**TUTORES:  
FRANCISCO GARCÍA GARCÍA  
VICENTE ARNAU LLOMBART  
RUBÉN GRILLO RISCO  
JOSÉ FRANCISCO CATALÀ SENENT**

**JULIO, 2021**





VNIVERSITAT  
DE VALÈNCIA



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

## **MÁSTER UNIVERSITARIO EN BIOINFORMÁTICA**

### **TRABAJO DE FIN DE MÁSTER**

# **CARACTERIZACIÓN DE LAS DIFERENCIAS DE SEXO EN CÁNCER DE PÁNCREAS DUCTAL MEDIANTE EL ANÁLISIS MASIVO DE DATOS TRANSCRIPTÓMICOS**

**AUTORA:**

**LUCÍA FANTIN**

**TUTORES:**

**FRANCISCO GARCÍA GARCÍA**

**VICENTE ARNAU LLOMBART**

**RUBÉN GRILLO RISCO**

**JOSÉ FRANCISCO CATALÀ SENENT**

---

**TRIBUNAL:**

PRESIDENTE/A:

VOCAL 1:



## Resumen

El cáncer de páncreas representa un importante problema de salud mundial ya que es una de las neoplasias con mayor mortalidad a nivel mundial. El subtipo más frecuente es el adenocarcinoma ductal de páncreas (PDAC). Epidemiológicamente está descrito que esta enfermedad presenta mayor incidencia en hombres que en mujeres, pero todavía no están claros los motivos que generan esta diferencia.

El objetivo de este trabajo es identificar los mecanismos moleculares que expliquen estas diferencias, para poder llevar a cabo una medicina personalizada que permita desarrollar terapias específicas más efectivas y dirigidas que modifiquen el pronóstico de esta enfermedad.

Para lograrlo hemos analizado los datos transcriptómicos de aquellos estudios centrados en PDAC que contaran con información de sexo provenientes de repositorios públicos, obtenidos a través de una revisión sistemática. Para cada uno de ellos realizamos un análisis de expresión diferencial y luego integramos estos datos mediante un metaanálisis de genes. Para caracterizar los resultados obtenidos realizamos un enriquecimiento funcional que nos permite identificar las funciones moleculares afectadas diferencialmente para cada sexo.

**Palabras clave:** cáncer de páncreas, PDAC, sexo, metaanálisis, transcriptómica.

## Abstract

Pancreatic cancer represents a major global health problem because it is one of the neoplasms with the highest rates of mortality worldwide. The most common subtype is pancreatic ductal adenocarcinoma (PDAC). Epidemiologically, it has been described that this disease has higher incidence rates in men than in women, but the reasons that generate this difference are not already clear.

This study aims to identify the molecular mechanisms that explain these differences to develop personalized medicine for more effective and specific therapies that modify the prognosis of this disease.

To achieve this, we have analyzed the transcriptomic data of PDAC studies from public repositories that includes sex information, obtained through a systematic review. For each of them we performed a differential expression analysis and then we integrated these data through a gene meta-analysis. Results were characterized through a functional enrichment that allowed identifying molecular functions differentially affected for each sex.

**Key words:** pancreatic cancer, PDAC, sex, meta-analysis, transcriptomics.



## Agradecimientos



En primer lugar, me gustaría agradecer a todas las personas que hicieron posible este proyecto. A Paco, por su ayuda, paciencia y buena predisposición para enseñar. A Zoraida, Irene y Carla que sin su colaboración hubiera sido mucho más difícil alcanzar esta meta. Al laboratorio de biología molecular del Instituto de Oncología Valenciano, donde se lleva a cabo el proyecto ULISES "Immunological incompatibility as a basis for cancer curing and vaccination" (FETOPEN-01-2018-2019-2020; Proposal number: 899708), centrado en el diseño de una estrategia inmunoterapéutica para PDAC, dentro del cual se ha desarrollado este trabajo.

A nivel personal me gustaría agradecerle a mis compañeros del máster de Bioinformática, en especial a Almu y a Cris por estar ahí siempre, en forma virtual y presencial. A mi familia en Argentina, que apoyó mi decisión de irme lejos para explorar nuevos horizontes. A la "familia" valenciana formada por amigos que saben lo que es estar lejos y siempre están cuando se necesita. A Simón, por acompañarme y venir a socorrerme cuando lo necesito.



# Índice general

|          |   |           |
|----------|---|-----------|
| <b>1</b> | <b>Introducción</b> .....                           | <b>1</b>  |
| 1.1      | Cáncer de páncreas.....                             | 1         |
| 1.1.1    | Epidemiología.....                                  | 1         |
| 1.1.2    | Etiología y Factores de riesgo.....                 | 2         |
| 1.1.3    | Patogénesis.....                                    | 5         |
| 1.1.4    | Bases moleculares del cáncer de páncreas.....       | 5         |
| 1.1.5    | Microambiente tumoral y respuesta inmunitaria.....  | 6         |
| 1.1.6    | Clasificación.....                                  | 6         |
| 1.1.7    | Manifestaciones clínicas.....                       | 7         |
| 1.1.8    | Diagnóstico.....                                    | 7         |
| 1.1.9    | Pronóstico.....                                     | 7         |
| 1.1.10   | Tratamiento.....                                    | 8         |
| 1.2      | Datos ómicos y tecnologías de alto rendimiento..... | 8         |
| 1.2.1    | Transcriptómica.....                                | 9         |
| 1.2.2    | Metaanálisis.....                                   | 10        |
| 1.2.3    | Caracterización funcional.....                      | 10        |
| <b>2</b> | <b>Hipótesis</b> .....                              | <b>13</b> |
| <b>3</b> | <b>Objetivos</b> .....                              | <b>15</b> |
| <b>4</b> | <b>Materiales y métodos</b> .....                   | <b>17</b> |
| 4.1      | Revisión sistemática y selección de estudios.....   | 17        |
| 4.2      | Análisis individual de los estudios.....            | 18        |
| 4.2.1    | Procesamiento de los datos.....                     | 18        |
| 4.2.2    | Análisis exploratorio.....                          | 19        |
| 4.2.3    | Análisis de expresión diferencial.....              | 19        |
| 4.3      | Metaanálisis.....                                   | 20        |
| 4.3.1    | Análisis de enriquecimiento funcional.....          | 22        |
| <b>5</b> | <b>Resultados</b> .....                             | <b>25</b> |
| 5.1      | Revisión sistemática y selección de estudios.....   | 25        |
| 5.2      | Análisis individual de los estudios.....            | 26        |
| 5.2.1    | Procesamiento de los datos.....                     | 26        |
| 5.2.2    | Análisis de expresión diferencial.....              | 38        |
| 5.3      | Metaanálisis de genes.....                          | 39        |
| 5.3.1    | Análisis de enriquecimiento funcional.....          | 42        |

|          |                               |           |
|----------|-------------------------------|-----------|
| <b>6</b> | <b>Discusión.....</b>         | <b>45</b> |
| <b>7</b> | <b>Conclusiones.....</b>      | <b>49</b> |
| <b>8</b> | <b>Bibliografía.....</b>      | <b>51</b> |
|          | <b>Anexo I - Figuras.....</b> | <b>57</b> |
|          | <b>Anexo II - Tablas.....</b> | <b>61</b> |

## Índice de figuras

|  |    |
|--|----|
| Figura 1.1. Anatomía del páncreas y su relación con órganos intestinales .....             | 1  |
| Figura 1.2: Incidencia y mortalidad del cáncer de páncreas por sexo y área geográfica..... | 2  |
| Figura 1.3: Imagen gráfica del dogma central de la biología molecular .....                | 9  |
| Figura 4.1: Flujo de información de una revisión sistemática.....                          | 18 |
| Figura 4.2: Ejemplo de gráfico de embudo.....  | 21 |
| Figura 4.3: Ejemplo de diagrama de bosque.....   | 22 |
| Figura 5.1: Diagrama de la revisión sistemática realizada .....                            | 26 |
| Figura 5.2: Distribución de las muestras para cada uno de los estudios.....                | 27 |
| Figura 5.3: Análisis de componentes principales del estudio GSE15932. ....                 | 28 |
| Figura 5.4: Clustering de las muestras del estudio GSE15932.....                           | 28 |
| Figura 5.5: Análisis de componentes principales del estudio GSE16515. ....                 | 29 |
| Figura 5.6: Clustering de las muestras del estudio GSE16515.....                           | 30 |
| Figura 5.7: Análisis de componentes principales del estudio GSE22780. ....                 | 31 |
| Figura 5.8: Clustering de las muestras del estudio GSE22780.....                           | 31 |
| Figura 5.9: Análisis de componentes principales del estudio GSE18670. ....                 | 32 |
| Figura 5.10: Clustering de las muestras del estudio GSE18670.....                          | 32 |
| Figura 5.11: Análisis de componentes principales del estudio GSE49641. ....                | 33 |
| Figura 5.12: Clustering de las muestras del estudio GSE49641.....                          | 33 |
| Figura 5.13: Análisis de componentes principales del estudio E-EMBL-6.....                 | 34 |
| Figura 5.14: Clustering de las muestras del estudio E-EMBL-6. ....                         | 35 |
| Figura 5.15: Análisis de componentes principales del estudio GSE55643. ....                | 36 |
| Figura 5.16: Clustering de las muestras del estudio GSE55643.....                          | 36 |
| Figura 5.17: Análisis de componentes principales del estudio GSE74629. ....                | 37 |
| Figura 5.18: Clustering de las muestras del estudio GSE74629.....                          | 38 |
| Figura 5.19: Funciones relacionadas a los genes más expresados en mujeres .....            | 42 |
| Figura 5.20: Funciones relacionadas a los genes más expresados en hombres .....            | 42 |
| Figura 5.21: Términos GO BP sobrerrepresentados .....                                      | 43 |
| Figura 5.22: Rutas KEGG sobrerrepresentadas .....  | 44 |
| Figura A 1: Diagrama de cajas del estudio GSE15932.....                                    | 57 |
| Figura A 2: Diagrama de cajas del estudio GSE16515. ....                                   | 57 |
| Figura A 3: Diagrama de cajas del estudio GSE22780. ....                                   | 58 |
| Figura A 4: Diagrama de cajas del estudio GSE18670. ....                                   | 58 |
| Figura A 5: Diagrama de cajas del estudio GSE49641. ....                                   | 59 |
| Figura A 6: Diagrama de cajas del estudio E-EMBL-6.....                                    | 59 |
| Figura A 7: Diagrama de cajas del estudio GSE55643. ....                                   | 60 |
| Figura A 8: Diagrama de cajas del estudio GSE74629. ....                                   | 60 |



## Índice de Tablas

|  |    |
|--|----|
| Tabla 5.1: Estudios seleccionados tras la revisión sistemática.....                | 25 |
| Tabla 5.2: Distribución de muestras por sexo y condición de estudio GSE15932.....  | 28 |
| Tabla 5.3: Distribución de muestras por sexo y condición de estudio GSE16515.....  | 29 |
| Tabla 5.4: Distribución de muestras por sexo y condición de estudio GSE22780.....  | 30 |
| Tabla 5.5: Distribución de muestras por sexo y condición de estudio GSE18670.....  | 32 |
| Tabla 5.6: Distribución de muestras por sexo y condición de estudio GSE49641.....  | 33 |
| Tabla 5.7: Distribución de muestras por sexo y condición de estudio E-EMBL-6.....  | 34 |
| Tabla 5.8: Distribución de muestras por sexo y condición del estudio GSE55643..... | 35 |
| Tabla 5.9: Distribución de muestras por sexo y condición del estudio GSE74629..... | 37 |
| Tabla 5.10: Resumen de los resultados del análisis de expresión diferencial .....  | 39 |
| Tabla 5.11: Listado de los 25 genes más expresados en mujeres .....                | 40 |
| Tabla 5.12: Listado de los 25 genes más expresados en hombres .....                | 41 |
| Tabla 5.13: Resultado del enriquecimiento funcional del metaanálisis.....          | 43 |
| Tabla A 1: Términos GO BP significativos.....                                      | 68 |
| Tabla A 2: Rutas Kegg significativas.....  | 68 |



## Abreviaturas

|                                |  |
|--------------------------------|--|
| <b>ADN</b>                     | ácido desoxirribonucleico.               |
| <b>ARN</b>                     | ácido ribonucleico.                      |
| <b>ARNm</b>                    | ácido ribonucleico mensajero.            |
| <b>BH</b>                      | Benjamini y Hochberg                     |
| <b>BP</b>                      | procesos biológicos.                     |
| <b>CAF</b>                     | fibroblastos asociados al cáncer.        |
| <b>CC</b>                      | componentes celulares.                   |
| <b>CCL2</b>                    | ligando 2 de quimiocina del motivo C-C.  |
| <b>CEA</b>                     | antígeno carcinoembrionario.             |
| <b>CTC</b>                     | células tumorales circulantes.           |
| <b>DL</b>                      | DerSimonian y Laird.                     |
| <b>EGFR</b>                    | factor de crecimiento epidérmico.        |
| <b>EMT</b>                     | transición epitelio-mesenquimatoso.      |
| <b>FGF</b>                     | factor de crecimiento de fibroblastos.   |
| <b>GEO</b>                     | Gene Expression Omnibus.                 |
| <b>GO</b>                      | Gene Ontology.                           |
| <b>GSEA</b>                    | Gene Set Enrichment Analysis             |
| <b>HGF</b>                     | factor de crecimiento de hepatocitos.    |
| <b>IFN-<math>\gamma</math></b> | interferón gamma.                        |
| <b>IL</b>                      | interleuquina.                           |
| <b>IPMN</b>                    | neoplasia mucinosa papilar intraductal.  |
| <b>KEGG</b>                    | Kyoto Encyclopedia of Genes and Genomes. |
| <b>KGF</b>                     | factor de crecimiento de queratinocitos. |
| <b>lncRNA</b>                  | long noncoding RNA.                      |
| <b>logFC</b>                   | logaritmo de fold change.                |
| <b>LOR</b>                     | logaritmo de Odds Ratio.                 |
| <b>M1</b>                      | macrófagos proinflamatorios activados.   |
| <b>MCN</b>                     | neoplasia quística mucinosa.             |

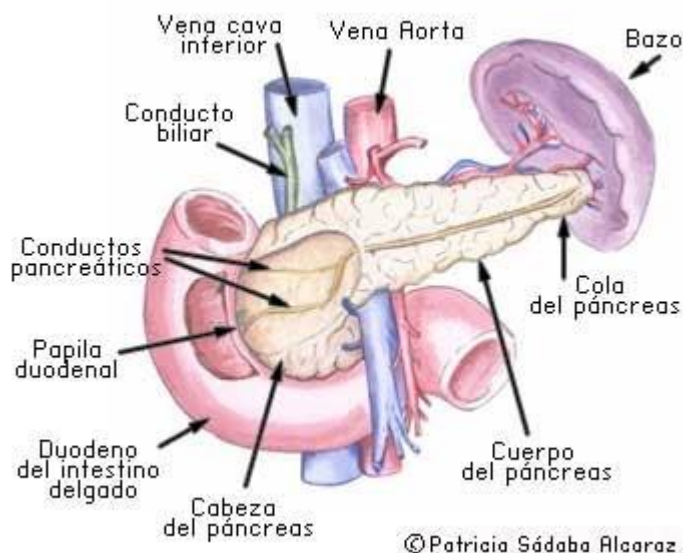
|                                |  |
|--------------------------------|--|
| <b>MEN1</b>                    | neoplasia endocrina múltiple tipo 1.           |
| <b>MeSH</b>                    | Medical Subject Headings.                      |
| <b>MF</b>                      | funciones moleculares.                         |
| <b>mGluR</b>                   | receptor de glutamato.                         |
| <b>MHC</b>                     | complejo mayor de histocompatibilidad.         |
| <b>NCBI</b>                    | National Center for Biotechnology Information. |
| <b>NF1</b>                     | neurofibromatosis tipo 1.                      |
| <b>NHI</b>                     | National Institute of Health                   |
| <b>NK</b>                      | natural killer.                                |
| <b>OMS</b>                     | Organización mundial de la salud.              |
| <b>PanIN</b>                   | neoplasia intraepitelial pancreática.          |
| <b>PDAC</b>                    | adenocarcinoma ductal de páncreas.             |
| <b>PDGF</b>                    | factor estimulante derivado de plaquetas.      |
| <b>PNET</b>                    | tumores neuroendócrinos pancreáticos.          |
| <b>RNAseq</b>                  | secuenciación de ARN.                          |
| <b>SE</b>                      | error estándar.                                |
| <b>TC</b>                      | tomografía computarizada.                      |
| <b>TGF-<math>\beta</math></b>  | factor de crecimiento transformante $\beta$ .  |
| <b>Th1</b>                     | T helper-1.                                    |
| <b>TNF-<math>\alpha</math></b> | factor de necrosis tumoral $\alpha$ .          |
| <b>VHL</b>                     | Von Hippel-Lindau                              |

# 1 Introducción

## 1.1 Cáncer de páncreas

El páncreas es una glándula localizada en el abdomen. Anatómicamente se divide en tres áreas: la cabeza, en contacto con el duodeno; el cuerpo y la cola, que contacta con el bazo. Debido a su localización, el páncreas está en contacto con gran cantidad de vasos sanguíneos y linfáticos, así como de estructuras nerviosas, tanto propias como de los órganos que le rodean (Figura 1.1). Funcionalmente tiene dos componentes: uno exocrino que secreta enzimas para la digestión intestinal y otro endocrino que secreta hormonas como la insulina y el glucagón, que regulan la concentración de glucosa en sangre y la gluconeogénesis, respectivamente.

El cáncer de páncreas se origina en la glándula pancreática a partir de cualquiera de los tipos celulares que la forman y puede afectar a cualquier región del órgano, aunque lo más frecuente es que se localicen en la cabeza del páncreas y que surjan de las células que forman parte de los conductos del componente exocrino (Adenocarcinoma ductal)<sup>1</sup>. De hecho, más del 90 % de los tumores pancreáticos corresponden al adenocarcinoma ductal de páncreas (PDAC), altamente agresivos y de mal pronóstico.



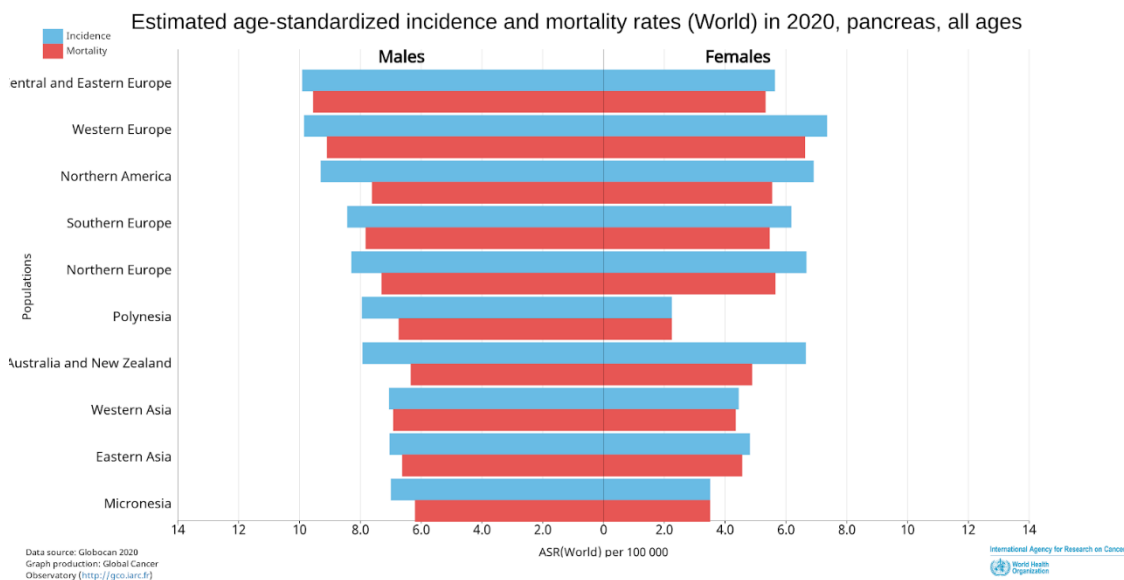
**Figura 1.1. Anatomía del páncreas y su relación con órganos intestinales. Fuente: Asociación española contra el cáncer<sup>2</sup>**

### 1.1.1 Epidemiología

El cáncer de páncreas representa un importante problema de salud mundial. Según la Organización Mundial de la Salud (OMS) en el año 2020 fue el doceavo tipo de cáncer diagnosticado y el séptimo más mortal a nivel mundial. En España en el mismo año ocupó

el séptimo lugar en frecuencia y el tercero en mortalidad<sup>3</sup>. Existe una tendencia general a tasas de incidencia más altas en los países desarrollados en comparación con los países en desarrollo tanto en hombres como en mujeres<sup>4</sup>.

Este cáncer afecta mayoritariamente a mayores de 55 años, siendo más frecuente a partir de los 65 años. Cuando se diagnostica en personas más jóvenes hay que sospechar alguna asociación con una alteración genética familiar<sup>5</sup>. La incidencia es ligeramente mayor en hombres que en mujeres (tasa estandarizada por edad de 5,5 en los hombres frente a 4,0 en las mujeres). Se cree que estas diferencias podrían deberse a diferentes exposiciones ambientales o a factores genéticos<sup>4</sup>, aunque en la literatura también describen diferencias sexuales en el páncreas exocrino normal y en los mecanismos que operan en la homeostasis y estados de enfermedad en ambos sexos<sup>6</sup>. Con respecto a la etnia, la raza afroamericana es más susceptible de padecerlo que la caucásica y también podría deberse a la exposición a los factores de riesgo.



**Figura 1.2: Incidencia y mortalidad del cáncer de páncreas por sexo y área geográfica. Gráfico de barras mostrando la tasa de incidencia y mortalidad por sexo y región en cáncer de páncreas en 2020 para todas las edades<sup>3</sup>**

### 1.1.2 Etiología y Factores de riesgo

No se conoce la etiología exacta de este tipo de tumor, pero existen diversos factores de riesgo que favorecen su aparición y que pueden agruparse como: i) debidos al ambiente y que pueden ser modificables (tabaco, dieta y obesidad, alcohol, infecciones, o fármacos como la insulina o ii) debidos al propio hospedador (edad, género, etnia, grupo sanguíneo, factores genéticos, pancreatitis crónica y diabetes mellitus), que suelen ser no modificables<sup>7</sup>. A continuación, se describe brevemente la importancia de estos factores en el PDAC.

#### Factores ambientales (modificables)

### Tabaco

Es el factor de riesgo modificable más importante<sup>4</sup>. Se asocia con un riesgo dos a cuatro veces mayor de padecer cáncer de páncreas en fumadores<sup>5</sup> y un 20% más de riesgo en los ex fumadores<sup>4</sup>. El aumento de la intensidad y la dosis acumulada aumentan el riesgo de padecer la enfermedad. También se ha demostrado que la exposición ambiental al humo del tabaco aumenta un 20% el riesgo de cáncer de páncreas<sup>8</sup>. Algunos estudios han demostrado que el riesgo permanece entre 10 o 20 años tras dejar de fumar<sup>4</sup>.

### Obesidad

Se asocia con una tasa de muerte de 20 a 40% más alta por cáncer de páncreas. La tumorigénesis aumenta por el exceso de tejido adiposo<sup>9</sup>. El consumo de grasas genera hipertrofia e hiperplasia pancreática en respuesta a la secreción de lipasa mediada por colecistoquinina por la presencia de grasa en el duodeno, lo que aumenta el riesgo de carcinogénesis de las glándulas exocrinas. Además, la hiperglucemia, los niveles anormales de glucosa y la resistencia a la insulina se asocian con un mayor riesgo de cáncer de páncreas<sup>8</sup>.

### Factores dietéticos

Los factores dietéticos también contribuyen al desarrollo del tumor: la mantequilla, las grasas saturadas, las carnes rojas y los alimentos procesados aumentan el riesgo de padecerlo. Mientras que una ingesta alta de frutas y ácido fólico podría reducir el riesgo<sup>9</sup>.

### Alcohol

Los metabolitos resultados del metabolismo del alcohol pueden causar lesiones tanto en la parte exocrina como endocrina del páncreas, fibrosis y la liberación de moléculas proinflamatorias, contribuyendo a la aparición de carcinogénesis pancreática<sup>7</sup>.

### Pancreatitis crónica

La pancreatitis crónica genera un estado inflamatorio del páncreas que conduce a fibrosis y pérdida de células acinares e islotes<sup>4</sup>. Aproximadamente el 5% de los enfermos de pancreatitis crónica generará un cáncer de páncreas en alrededor de 10 a 20 años<sup>4,8</sup>. La causa más común de pancreatitis crónica es el consumo de alcohol<sup>8,9</sup>.

### Diabetes mellitus

La diabetes es un factor de riesgo bien establecido para el cáncer de páncreas. Se ha demostrado que los enfermos de diabetes tanto de tipo uno como de tipo dos tienen un riesgo dos veces mayor de padecer cáncer de páncreas. Sin embargo, además de ser un factor de riesgo la diabetes puede ser un síntoma de la enfermedad cuando es de reciente aparición<sup>4,10</sup>.

### Infecciones

La infección por *Helicobacter pylori*, hepatitis B y el virus de la inmunodeficiencia humana están relacionados con un aumento en el riesgo relativo de cáncer de páncreas<sup>4,9</sup>.

### **Factores del hospedador (no modificables)**

#### Edad

El cáncer de páncreas afecta mayormente a adultos mayores y es extremadamente raro que los jóvenes menores de 30 años desarrollen la enfermedad. Aproximadamente el 90% de los pacientes recién diagnosticados tienen más de 55 años, la mayoría de ellos entre 70 y 80 años<sup>4,11</sup>.

#### Sexo

Los hombres tienen una mayor incidencia y menor supervivencia de adenocarcinoma ductal de páncreas a nivel mundial. En los países desarrollados, esta brecha es aún más pronunciada<sup>4,11,12</sup>. Existen diferencias dependientes de sexo en la etiología y pronóstico de las enfermedades pancreáticas, entre ellas el cáncer, aunque todavía pobremente conocidas. Estas diferencias son atribuidas a los cromosomas sexuales y a las hormonas sexuales, principalmente. Algunos autores han observado un efecto protector de los estrógenos sobre las mujeres<sup>13</sup> mientras que el crecimiento de las células normales del páncreas, pero también de las tumorales sería promovido por la acción de los andrógenos, aunque esta asociación está todavía en discusión. Otra de las explicaciones es que esta diferencia estaría más relacionada con una diferencia de sexo en combinación con diferentes exposiciones en factores ambientales, como el tabaco o el alcohol que predominan más en hombres, o genéticos<sup>4</sup>.

#### Etnia

La incidencia de cáncer de páncreas varía en diferentes partes del mundo. En los Estados Unidos, la incidencia en los afroamericanos es mayor que la de los caucásicos, mientras que los estadounidenses de origen asiático y los isleños del Pacífico tienen la incidencia más baja<sup>4,11</sup>. Se propone que las tasas de incidencia más altas dentro de la población afroamericana están relacionadas con una mayor exposición a otros factores de riesgo de cáncer de páncreas, como el tabaquismo, el consumo de alcohol, el índice de masa corporal elevado y una mayor incidencia de diabetes. Algunos estudios consideran también que existen factores genéticos que contribuyen a esta diferencia<sup>4</sup>.

#### Grupo sanguíneo

Se ha demostrado que las personas con grupo sanguíneo A, B y AB presentan mayor riesgo de desarrollar un cáncer de páncreas que las personas del grupo O. Se cree que esto puede deberse a mecanismos de regulación inflamatoria<sup>4,11</sup>.

### Antecedentes familiares y factores genéticos

Tener antecedentes familiares de cáncer de páncreas aumenta nueve veces el riesgo de padecerlo, y si se han diagnosticado previamente tres o más familiares de primer grado, el riesgo aumenta 32 veces<sup>4</sup>. Entre las alteraciones genéticas más importantes, destacan: sobreactivación del oncogén *KRAS*, inactivación de los genes supresores *CDKN2A* y *P53*, alteración de *SMAD4*, mutación del *BRCA2*, mutaciones en el gen *PRSSI* (implicado en la pancreatitis familiar), alteraciones en los genes implicados en la reparación del ácido desoxirribonucleico (ADN) como: *MLH1* o *MSH2*, *MLH3*, *MSH6*, *TGBR2*, *PMS1*, y *PMS2*. Alteraciones del gen *VHL* (Síndrome de von Hippel-Lindau), del gen *NF1* (implicado en la neurofibromatosis tipo 1) y que aumenta la incidencia de cánceres endocrinos como el somatostatina o del gen *MEN1* (que suele producir neoplasia endocrina múltiple tipo 1) pueden también contribuir a la aparición de cáncer de páncreas<sup>14</sup>.

### **1.1.3 Patogénesis**

Una serie de mutaciones escalonadas de la mucosa normal del páncreas generan lesiones precursoras específicas que progresan a una neoplasia maligna invasiva. Los tres precursores mejor caracterizados de esta neoplasia son la neoplasia intraepitelial pancreática (PanIN), las neoplasias mucinosas papilares intraductales (IPMN) y las neoplasias quísticas mucinosas (MCN)<sup>4,8</sup>.

### **1.1.4 Bases moleculares del cáncer de páncreas**

PanIN es el precursor más común del adenocarcinoma de páncreas. Está demostrado que estas lesiones tienen anomalías genéticas que son comunes al adenocarcinoma de páncreas y la progresión histológica de PanIN es paralela a la acumulación de anomalías moleculares. Las lesiones de PanIN de bajo grado tienen mutaciones en el oncogén *KRAS* y muestran un acortamiento de los telómeros. Las mutaciones en p16, *CDNK27*, p53 y *SMAD4* aparecen más tarde y están presentes en PanIN de alto grado y en el adenocarcinoma de páncreas. La tasa de mutación de *KRAS* también aumenta en relación con el grado de PanIN<sup>4</sup>.

Las anomalías genéticas más frecuentes del adenocarcinoma de páncreas son la activación del oncogén *KRAS*, la inactivación de genes supresores de tumores como *CDKN2A*, *TP53*, *SMAD4*, y *BRCA2*, pérdidas cromosómicas generalizadas y acortamiento de telómeros. Menos frecuentemente ocurren mutaciones en oncogenes como *BRAF*, *MYB*, *AKT2* y *EGFR*, y en genes supresores de tumores como *MAP2K4*, *STK11*, *TGFBR1*, *TGFBR2*, *ACVR1B*, *ACVR2A*, *FBXW7* y *EP300*<sup>15</sup>.

### 1.1.5 Microambiente tumoral y respuesta inmunitaria

El estroma circundante o microambiente tumoral está compuesto por una matriz extracelular alterada formada de colágeno, integrina, laminina, fibronectina, glicosaminoglicanos, entre otros, y una amplia variedad de tipos de células, incluidas células epiteliales, fibroblastos, células de músculo liso, células inmunes e inflamatorias y células endoteliales linfáticas y vasculares. Una población de fibroblastos circundantes conocidos como fibroblastos asociados al cáncer (CAF) desempeñan un papel fundamental en la progresión tumoral al estimular las vías de señalización de supervivencia y proliferación a través de la señalización heterotípica con células cancerosas. Los CAF son activados por el factor de crecimiento transformante  $\beta$  (TGF- $\beta$ ), factor de necrosis tumoral  $\alpha$  (TNF- $\alpha$ ), factor estimulante derivado de plaquetas (PDGF), y las interleuquinas (IL)- 1, -6, y - 10. También son estimulados por otros mediadores, como el ligando 2 de quimiocina del motivo C-C (CCL2), el factor de crecimiento de hepatocitos (HGF) y el factor de crecimiento de fibroblastos (FGF). Los CAF inducen la desmoplasia a través de la secreción de colágeno de tipo I y III, fibronectina, proteoglicanos y glucosaminoglicanos, lo que conduce a un aumento de presión mecánica en la matriz extracelular, que puede promover la migración de las células cancerosas e inhibir la vascularización. A su vez proporcionan nutrición a las células cancerosas en condiciones de baja glucosa y contribuyen a la transición epitelio-mesenquimatosa (EMT), la invasión del cáncer, angiogénesis y metástasis<sup>16,17</sup>.

La respuesta inmunitaria tiene un papel fundamental en el desarrollo del cáncer de páncreas. Tanto la respuesta inmune innata como la específica actúan contra las células cancerígenas. La función anticancerígena eficaz del sistema inmunológico requiere células T CD8 citotóxicas, células T helper-1 (Th1), células dendríticas maduras, macrófagos proinflamatorios activados (M1) y células NK (Natural Killer). Al igual que otros tipos de cáncer, el cáncer de páncreas es capaz de evadir al sistema inmunitario a través de una variedad de mecanismos inmunosupresores que favorecen la progresión y diseminación del cáncer. Entre estos mecanismos destacan la disminución de la expresión de las moléculas de clase I del complejo mayor de histocompatibilidad (MHC), que son las encargadas de presentar los antígenos a los que se unen los linfocitos T CD8+ para activarse, o la expresión de ligandos que actúan como puntos de control del sistema inmune, produciendo un agotamiento inmunológico. A su vez, el cáncer de páncreas secreta factores inmunosupresores solubles como IL -10 y TGF- $\beta$  que estimulan la entrada y aumentan la función de células inmunitarias supresoras del sistema inmune<sup>18,19</sup>. Actualmente se están desarrollando diversas estrategias immuno-terapéuticas que permitan estimular al sistema inmune para que sea capaz de reconocer y destruir las células tumorales de forma más eficaz.

### 1.1.6 Clasificación

Las neoplasias pancreáticas se pueden clasificar en endocrinas o exocrinas según el componente de la glándula que afecten. El 95% de las neoplasias pancreáticas tienen origen exocrino, y de ellas el adenocarcinoma ductal es el más frecuente, representando alrededor del 80% de los tumores pancreáticos<sup>9</sup>. Otros tipos menos frecuentes de tumores exocrinos son: carcinoma de célula acinar, carcinoma adenoescamoso, carcinoma coloidal, tumor de células gigantes, carcinoma hepatoide, neoplasias quísticas mucinosas, pancreatoblastoma,

cistadenoma seroso, carcinoma de células en anillo de sello, tumores sólidos y pseudopapilares, carcinoma de células escamosas y carcinoma no diferenciado. Los tumores endocrinos y neuroendocrinos pancreáticos (PNET) representan el 1% de las neoplasias pancreáticas. Pueden ser tumores funcionales o no funcionales. En caso de ser funcionales se los denomina según la hormona que producen: Insulinoma, Glucagonoma, Gastrinoma, Somatostatina, VIPomas, PPomas<sup>20</sup>.

### **1.1.7 Manifestaciones clínicas**

La clínica de este tumor aparece tardíamente y se produce por el efecto de masa que genera en los órganos circundantes o por efecto de las metástasis. La manifestación más común es la ictericia, que aparece cuando los tumores se localizan en la cabeza de la glándula (ictericia silenciosa), al comprimir el conducto colédoco. Cuando se localizan en el cuerpo o en la cola las manifestaciones son más tardías. Otras manifestaciones pueden ser dolor abdominal o lumbar, vómitos, pérdida de peso, esteatorrea, diabetes de reciente aparición<sup>9,15</sup>. Ocasionalmente, la obstrucción del conducto pancreático puede provocar ataques de pancreatitis. La trombosis venosa profunda y superficial no es inusual y podría ser un signo de presentación de una enfermedad maligna.

### **1.1.8 Diagnóstico**

El diagnóstico de este tipo de cáncer es problemático sobre todo en las etapas tempranas de desarrollo, dada la ausencia de marcadores tumorales fiables. Actualmente, los biomarcadores más utilizados en la práctica clínica para el cáncer de páncreas son el antígeno CA 19-9 y el antígeno carcinoembrionario (CEA), menos utilizado que el primero, pero ninguno de ellos tiene la suficiente sensibilidad y especificidad para la diagnosis de este tipo de cáncer<sup>21</sup>. Por otra parte, los estudios de imágenes resultan útiles para visualizar la ocupación del páncreas por una masa y para estadificar. La tomografía computarizada (TC) con protocolo pancreático trifásico es la mejor prueba de diagnóstico inicial para el cáncer de páncreas<sup>15</sup>. La ecografía endoscópica también es muy precisa para el diagnóstico y permite también realizar una toma de muestra para citología en el mismo momento. El diagnóstico citológico generalmente se puede hacer con ecografía endoscópica o aspiración con aguja fina guiada por TC. No se necesita una muestra de biopsia para la resección quirúrgica cuando la sospecha de cáncer es alta; en general, la resección proporcionará un beneficio terapéutico. Retrasar sustancialmente la cirugía para confirmar un diagnóstico podría retrasar el comienzo de un tratamiento eficaz<sup>15</sup>.

### **1.1.9 Pronóstico**

Una vez diagnosticado, la esperanza de vida es de aproximadamente 5% a los 5 años<sup>9</sup>. Los factores que influyen en la supervivencia incluyen la edad, el sexo, la calidad de la atención médica, la presencia de comorbilidades y hábitos de estilo de vida. El principal factor que influye en el resultado de la enfermedad es el estadio del tumor en el momento del diagnóstico. Sólo el 20% de los pacientes con cáncer de páncreas tienen enfermedad

resecable quirúrgicamente en el momento de la presentación. En pacientes que pueden someterse a una resección quirúrgica satisfactoria, la supervivencia a 5 años se calcula como 27%, mientras que, si el paciente tiene enfermedad localmente avanzada o metastásica, la supervivencia media es de seis a once meses y de dos y seis meses, respectivamente<sup>4</sup>.

### **1.1.10 Tratamiento**

El tratamiento del cáncer de páncreas requiere de un equipo multidisciplinario formado por oncólogos, cirujanos, radiólogos, gastroenterólogos, oncólogos radioterapeutas, patólogos, expertos en manejo del dolor, trabajadores sociales, dietistas y (cuando sea apropiado) expertos en cuidados paliativos<sup>15</sup>.

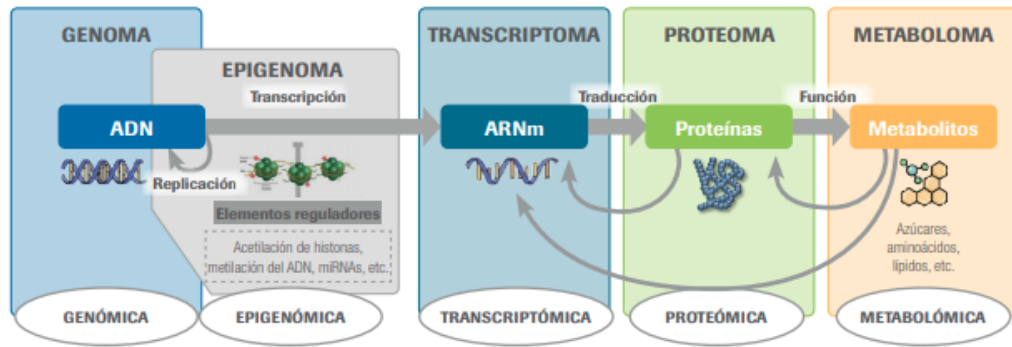
El único tratamiento curativo hasta el momento es la resección quirúrgica del tumor, pero menos del 20% de los diagnósticos se realizan en la etapa en que el tumor es totalmente resecable<sup>9</sup>. La adición de quimioterapia y radioterapia en el contexto adyuvante mejoran las tasas de supervivencia<sup>4,15</sup>. En caso de enfermedad localmente invasiva el tratamiento es la quimioterapia. La adición de radioterapia en estos pacientes es controvertida<sup>15</sup>. La enfermedad metastásica se trata con quimioterapia, aunque pueden también requerir cirugías que alivien los síntomas. Es importante que estos pacientes cuenten con apoyo de cuidados paliativos enfocados a la prevención y alivio del dolor mejorando su calidad de vida<sup>9</sup>.

## **1.2 Datos ómicos y tecnologías de alto rendimiento**

La información genética de los seres vivos se encuentra en el ácido desoxirribonucleico o ADN. Esta molécula está conformada por una parte central compuesta por una desoxirribosa y un fosfato, a la que se enlazan las bases nitrogenadas que pueden ser de 4 tipos: Adenina, Citosina, Timina y Guanina. El orden en que se combinan las bases nitrogenadas es lo que codifica la información genética<sup>22</sup>.

Para poder llevar a cabo las funciones que se encuentran codificadas en el ADN es necesario primero realizar un proceso de transcripción donde se genera una molécula de ácido ribonucleico (ARN) tomando una parte de ADN como molde. Esta molécula está conformada por una cadena simple de ribonucleótidos, cada uno de ellos formado por una ribosa, un fosfato y una base nitrogenada. A diferencia de la molécula de ADN, en este caso la Timina se sustituye por Uracilo. El ARN mensajero (ARNm) contiene la información necesaria para codificar una proteína. Para esto es necesario que se lleve a cabo el proceso de traducción, donde se forma la cadena de aminoácidos que dará lugar a la formación de la proteína encargada de llevar a cabo la función determinada<sup>22</sup>.

La transcripción y traducción, junto con la replicación y la transcripción reversa, conforman lo que es conocido como el dogma central de la biología molecular. Este dogma resume los pasos necesarios a seguir para las síntesis de proteínas encargadas de llevar a cabo las funciones contenidas en el genoma, necesarias para la supervivencia del organismo<sup>23</sup>. Desde su formulación en 1970, el interés por estudiar los pasos descritos en él ha ido en constante aumento<sup>24</sup>.



**Figura 1.3: Imagen gráfica del dogma central de la biología molecular y las ciencias ómicas que estudian cada uno de sus pasos<sup>25</sup>.**

Las ciencias ómicas son las disciplinas encargadas de estudiar cada uno de los pasos del dogma central de la biología molecular. La primera en describirse fue la genómica como el estudio del conjunto completo de genes de un organismo. El desarrollo de nuevas herramientas de trabajo permitió desarrollar también áreas como la transcriptómica, que estudia el conjunto de ARN generado por un organismo, y la proteómica, que estudia las proteínas del organismo. Incluso, actualmente se han desarrollado ciencias ómicas que están más allá del dogma central de la biología molecular, como la metabolómica y la interactómica<sup>24,26</sup>.

El avance de las nuevas tecnologías facilita la obtención de información cada vez más precisa acerca de los mecanismos moleculares de un organismo. Conocer más a fondo estos procesos nos permite abordar de forma más certera diferentes enfermedades, en el marco de la Medicina Personalizada.

### 1.2.1 Transcriptómica

El transcriptoma es el conjunto completo de ARN transcrito por el genoma de una célula y la transcriptómica la ciencia que lo estudia. El estudio del transcriptoma permite conocer el nivel de expresión de una región determinada del genoma bajo ciertas condiciones. Las células poseen mecanismos que regulan la expresión génica, de manera que ante unas condiciones se expresan unos genes y bajo otras condiciones otros. Estudiar y comparar los transcritos producidos bajo diferentes circunstancias (como en condiciones normales o patológicas) nos permitirá conocer funciones sobreexpresadas que orienten al desarrollo de una patología<sup>27</sup>.

El desarrollo de la transcriptómica ha sido posible gracias al desarrollo de plataformas como los chips de ADN y la secuenciación de ARN (RNA-seq), a través de los cuales podemos conocer los niveles de transcripción de una secuencia de interés. Los chips o arrays de ADN están conformados por sondas de ADN que se adhieren a un soporte sólido en una determinada disposición. El transcriptoma de la muestra de interés se acopla con una molécula fluorescente o radiactiva y luego se deposita sobre dicho soporte, hibridando con las sondas del chip. La intensidad de señal emitida por una sonda permite identificar la cantidad en que se ha transcrito una parte del genoma<sup>28</sup>.

La técnica de RNA-seq requiere primero convertir el ARN de la muestra de interés en ADN por transcripción inversa para obtener la secuencia de nucleótidos mediante técnicas de secuenciación masiva. Esta técnica permite conocer los transcritos de la célula y cuantificarlos<sup>29</sup>.

Los datos obtenidos de estos estudios se muestran en forma de matriz donde se indica el nivel de expresión para cada sonda. A partir de esta información pueden llevarse a cabo diferentes estrategias de análisis, según el tipo de datos que se utilice y el interés del estudio<sup>30</sup>.

Algunas de estas son<sup>31</sup>:

- **Análisis de expresión diferencial:** permite identificar genes o transcritos que varían sus niveles de expresión entre diferentes condiciones experimentales.
- **Predicción de clases:** a partir de un conjunto de datos de entrenamiento, con perfiles transcriptómicos que varían entre las diferentes condiciones, es posible realizar una nueva clasificación supervisada de nuevos perfiles.
- **Análisis cluster o clasificación no supervisada:** agrupación de muestras o grupos de genes según su similitud a nivel transcriptómico.

### **1.2.2 Metaanálisis**

El metaanálisis es un diseño de estudio que combina los resultados de diferentes estudios individuales pero relacionados, para obtener mayor evidencia científica<sup>32</sup>. La integración de la información de diferentes estudios permite obtener un análisis con mayor potencia estadística que si se realiza sobre un único estudio<sup>33,34</sup>.

La revisión sistemática integra toda la información disponible sobre un tema específico de interés o una pregunta de investigación. El metaanálisis es la metodología estadística que permite combinar resultados sobre los efectos procedentes de diversos estudios individuales obtenidos a través de esta revisión<sup>30</sup>. Permite incluso combinar resultados de estudios con resultados contradictorios estimando un efecto promedio.

La versatilidad de esta técnica también hace posible realizarla tanto a nivel de transcrito, con los resultados de un análisis de expresión diferencial, como a nivel funcional, con los resultados de un análisis de enriquecimiento funcional<sup>30</sup>.

### **1.2.3 Caracterización funcional**

El análisis de expresión diferencial identifica una lista de diferentes elementos biológicos de interés, con sus respectivos niveles de expresión para cada una de las condiciones que se comparan<sup>35</sup>. Sin embargo, este listado no nos aporta información acerca de la función biológica de estos elementos y es necesario combinar estos elementos con la información contenida en las bases de datos disponibles. De este modo se obtiene un significado biológico de las funciones que resultan significativas estadísticamente y se enriquece la interpretación de los resultados. Las bases de datos de mayor interés son aquellas que aportan información de funciones, rutas o procesos en las que están implicados los genes<sup>30</sup> como:

- Gene Ontology (GO), que contiene información sobre funciones biológicas específicas para diferentes genes. Esta información está clasificada en tres ontologías: procesos biológicos (BP), funciones moleculares (MF) y componentes celulares (CC)<sup>36</sup>.
- Kyoto Encyclopedia of Genes and Genomes (KEGG), que contiene información acerca de las rutas metabólicas en las que participan los genes<sup>37</sup>.

Una vez añadida esta información a los genes considerados en el estudio es posible llevar a cabo análisis de enriquecimiento funcional. Una de las aproximaciones utilizadas es Gene Set Enrichment Analysis (GSEA)<sup>30,38</sup>. Esta metodología actúa sobre una lista de genes ordenados por su nivel de expresión diferencial (o cualquier otra ordenación de interés biológico), y a continuación determina si los genes que codifican para una determinada función presentan un patrón de expresión común (sobrexpresados o infraexpresados), otorgándole significado biológico a los resultados obtenidos.



## **2 Hipótesis**

El cáncer de páncreas es una de las neoplasias con mayor mortalidad a nivel mundial. Epidemiológicamente está descrito que esta enfermedad presenta mayor incidencia en hombres que en mujeres. Nuestra hipótesis es que existen diferencias entre sexos a nivel molecular relacionadas con esta patología. Su estudio permitiría desarrollar estrategias diagnósticas más precisas y terapias mejor enfocadas y más efectivas.



### 3 Objetivos

El objetivo principal de este trabajo es identificar los mecanismos moleculares alterados en cáncer de páncreas, específicamente en PDAC, que afectan de forma diferencial a hombres y mujeres, a través de un abordaje *in silico*. Para ello, se han propuesto los siguientes objetivos específicos:

- Revisión sistemática y selección de estudios de cáncer de páncreas con datos de microarray o RNA-seq en los repositorios GEO y ArrayExpress.
- Análisis individual de los estudios seleccionados mediante un análisis de expresión diferencial y enriquecimiento funcional.
- Metaanálisis de genes para identificar aquellos que presenten un perfil diferencial por sexo.



## 4 Materiales y métodos

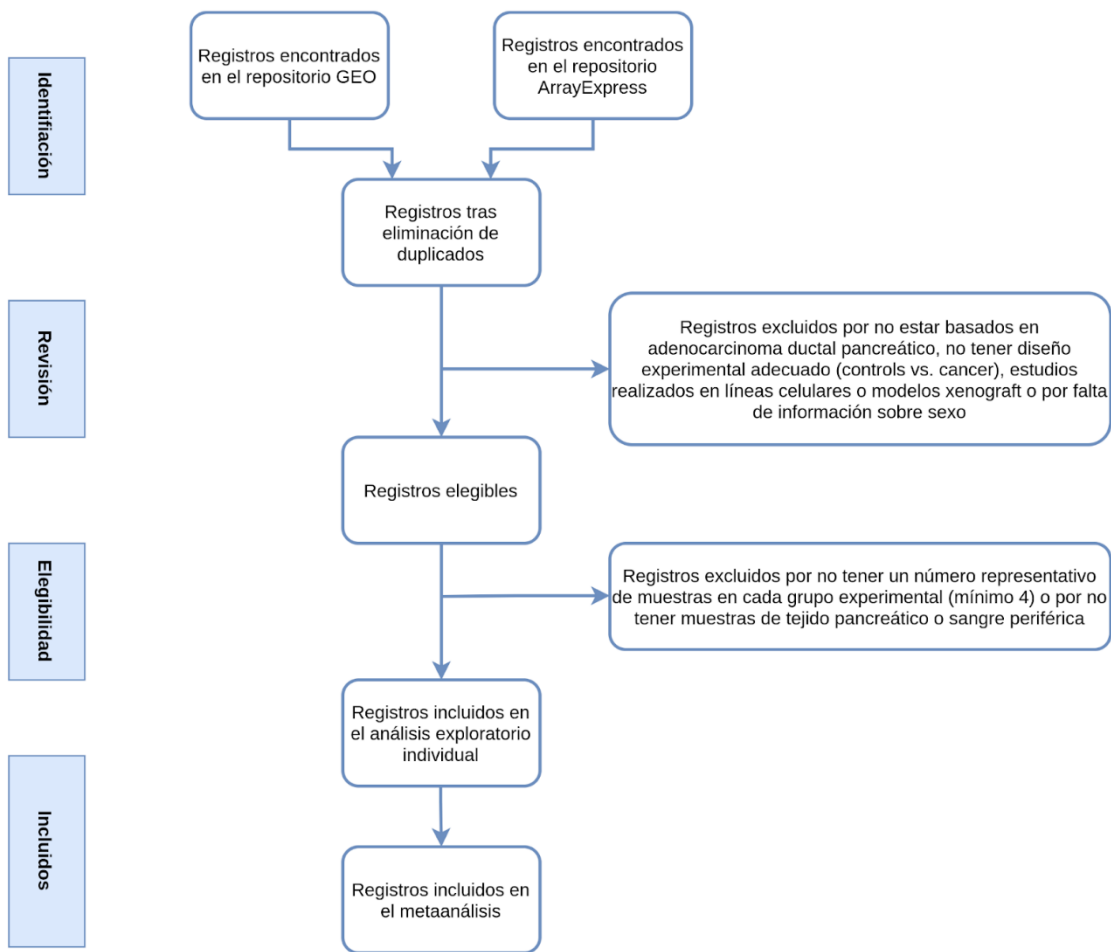
Todo el análisis bioinformático ha sido realizado en el lenguaje de programación R versión 4.0.4<sup>39</sup>.

### 4.1 Revisión sistemática y selección de estudios

La revisión sistemática es un método que integra toda la información disponible sobre un tema de interés. La selección de estos estudios se realiza a través de una búsqueda en bases de datos seleccionadas, a partir de unos criterios de selección establecidos. En este trabajo la revisión se realizó en los repositorios Gene Expression Omnibus (GEO)<sup>40</sup> y ArrayExpress<sup>41</sup>, siguiendo las directrices PRISMA para el desarrollo de revisiones sistemáticas y metaanálisis<sup>34</sup> (Figura 4.1). El término utilizado en la búsqueda fue “Pancreatic cancer” (Medical Subject Headings o MeSH). Los resultados obtenidos los filtramos a aquellos estudios que fueran realizados en humanos, que contaran con 8 o más muestras, cuya tecnología utilizada fuera Microarray o RNA-seq y que no hubieran sido realizados en líneas celulares.

De los estudios obtenidos a partir de la búsqueda, se seleccionaron aquellos que estuvieran centrados en PDAC, que contuvieran información sobre el sexo y que hubieran usado muestras de tejido pancreático o sangre periférica. Se excluyeron aquellos estudios que no tuvieran muestras control o tuvieran un tamaño muestral por grupo experimental menor a cuatro sujetos.

Los datos de los estudios seleccionados fueron obtenidos utilizando el paquete GEOquery<sup>42</sup>.



**Figura 4.1: Flujo de información a través de las diferentes fases de una revisión sistemática.**

## 4.2 Análisis individual de los estudios

Una vez obtenidos los datos normalizados se completó el procesamiento de los datos en los estudios que lo requirieron y se realizó un análisis exploratorio de cada uno de los estudios, la detección de genes diferencialmente expresados, y su caracterización funcional.

### 4.2.1 Procesamiento de los datos

Se llevó a cabo una estandarización de los nombres de los grupos experimentales en cada uno de los estudios que garantizara una nomenclatura común. Además, se realizó una estandarización de los datos de anotación de los genes.

- Estandarización de grupos experimentales:

Según si los sujetos sufren o no la enfermedad se los clasificó en:

- Grupo control: pacientes sanos

- Grupo case: Pacientes con diagnóstico de PDAC

Según sexo:

- Grupo male: muestras pertenecientes a pacientes masculinos
- Grupo female: muestras pertenecientes a pacientes femeninas

- Estandarización de la anotación:

Es necesario realizar una estandarización de la anotación de los estudios porque estos, al utilizar diferentes modelos y pertenecer a distintas plataformas comerciales, utilizan identificadores de sondas diferentes. Es necesario modificar los identificadores de sonda de cada uno de los estudios (ProbeID) por un identificador común. En este caso utilizamos los identificadores Entrez, que identifica cada uno de los genes en la base de datos del National Center for Biotechnology Information (NCBI)<sup>43</sup>. Dado que en los chips de ADN existen diferentes sondas que pueden hibridar en diferentes regiones de un mismo gen, fue necesario realizar la mediana de los niveles de expresión para esas sondas para obtener un único valor de expresión. Para esto se utilizó la función medianReps.

#### **4.2.2 Análisis exploratorio**

Se realizó un análisis exploratorio de cada uno de los estudios seleccionados y anotados con el objetivo de evaluar la distribución de las muestras en grupos de interés y detectar comportamientos anómalos y evaluar la calidad de los datos. Se representaron mediante diagramas de caja los datos de expresión de las muestras de cada uno de los estudios para comprobar su normalización y detectar valores anómalos. Además, mediante un análisis de agrupamiento o clustering y un análisis de componentes principales (PCA), se evaluó la forma en que se distribuyen las muestras de los grupos experimentales.

#### **4.2.3 Análisis de expresión diferencial**

Para realizar el análisis de expresión diferencial y conocer la diferencia en la expresión de los genes entre mujeres y hombres que padecen PDAC, fue necesario crear una nueva variable que combine la información del sexo (Male o Female) con el estado de la enfermedad (Case o Control). A partir de esto se planteó el siguiente contraste de interés:

*(Female case - Female control) - (Male case - Male control)*

Utilizando las funciones de la librería de R limma<sup>44</sup> se creó una matriz de contrastes y luego se ajustó, para cada gen, un modelo lineal. A partir de la matriz de contrastes y el modelo lineal se calcularon los coeficientes estimados, errores estándar y los estadísticos de la

expresión diferencial. Los p-valores de todos los genes fueron ajustados usando el método de Benjamini y Hochberg (BH)<sup>45</sup>.

### 4.3 Metaanálisis

Para determinar qué genes están diferencialmente expresados en el conjunto de los estudios seleccionados de PDAC, se integraron los datos de expresión diferencial obtenidos individualmente para cada uno de los estudios en forma de metaanálisis.

A partir de los logaritmos de fold change (logFC) y el error estándar (SE) se crearon 2 matrices donde las filas representan los genes y las columnas indican el estudio al que pertenecen.

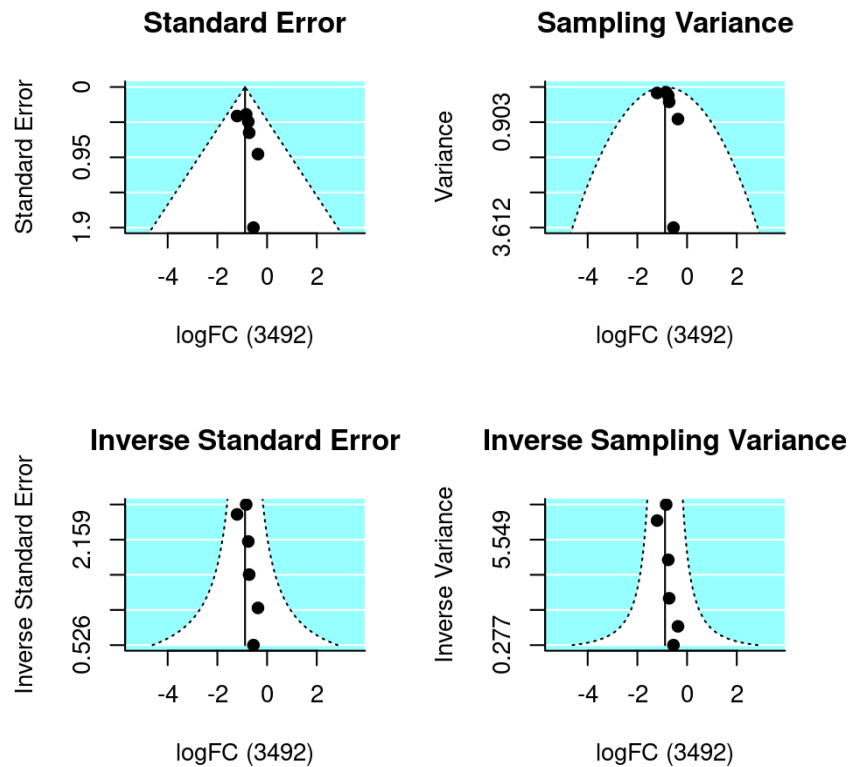
#### **Determinación de la medida combinada del efecto:**

A partir de estas matrices generadas se realizó el metaanálisis para cada uno de los genes de los estudios. Para realizarlo se empleó el paquete metafor de R<sup>46</sup> utilizando el método de efectos aleatorios de DerSimonian y Laird (DL)<sup>47</sup>. Como resultado del metaanálisis se obtuvo una tabla con los diferentes estimadores de la medida combinada del efecto de los diferentes genes e indicadores de la heterogeneidad del metaanálisis<sup>30</sup>:

- QE y QEp representan el estadístico de contraste y el p-valor respectivamente de la prueba de DL, utilizada para detectar la presencia de heterogeneidad entre los estudios. La hipótesis nula indica la presencia de homogeneidad entre los estudios.
- Logaritmo de Odds Ratio (LOR) es la estimación del efecto combinado de todos los estudios. El signo indica en qué grupo se encuentra más sobreexpresada una función: el signo positivo indica mayor presencia de genes con nivel alto de expresión en el primer grupo de la comparación con respecto al segundo y el negativo mayor expresión en el segundo grupo.
- Intervalo de confianza del LOR (95%) construido con la variabilidad estimada en el modelo seleccionado. La no inclusión del 0 en este intervalo indica que el LOR es significativo.
- P-valor informa del nivel de significación de un efecto combinado nulo. El p-valor no contempla la multiplicidad del análisis realizado por lo que es necesario realizar una corrección utilizando el método BH.

#### **Evaluación de la heterogeneidad:**

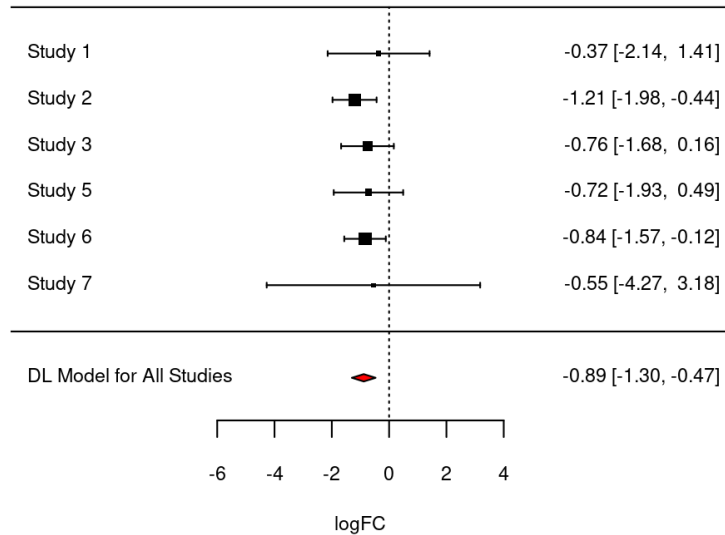
Dado que el test Q del método DL presenta una baja potencia cuando el número de estudios es pequeño, es necesario acompañar el análisis de la heterogeneidad entre estudios por gráficos de embudo. En este tipo de gráfico se representa la magnitud del efecto medido en el eje X y una medida de precisión en el eje Y. Cada punto del gráfico representa un estudio. Evaluando este tipo de gráficos podemos determinar sesgos entre los estudios<sup>30</sup>.



**Figura 4.2: Ejemplo de gráfico de embudo para evaluar la heterogeneidad de los estudios seleccionados. En el eje X se representa la magnitud del efecto medido en el eje y en eje Y se muestra una medida de precisión en el eje Y. Cada punto del gráfico representa un estudio.**

### Representación de resultados:

Además de presentar los resultados del metaanálisis, también se realizaron gráficos de bosque para cada uno de los genes. En estos se representan en el extremo izquierdo los estudios incluidos en el metaanálisis y en el derecho la estimación de la medida resumen de cada uno, con su intervalo de confianza al 95%. En la parte central se representa la medida del efecto (cuadrado en color negro) cuyo tamaño es proporcional a la precisión de las estimaciones. El cuadrado está ubicado dentro de un segmento que representa los extremos de su intervalo de confianza. En la parte inferior, el resultado global del metaanálisis se representa con un rombo en color verde. Su posición respecto a la línea de efecto nulo nos informa sobre la significación estadística del resultado global, mientras que su anchura nos proporciona una idea de su precisión (su intervalo de confianza). También en la parte inferior derecha de esta Figura se indica el valor de significación de los intervalos de confianza (habitualmente 95 %) y a la izquierda el modelo de análisis de datos que se ha utilizado. Este tipo de gráficos nos permiten evaluar los resultados globales y el peso en el resultado de cada función de cada estudio individual<sup>30</sup> (Figura 4.3).



**Figura 4.3: Ejemplo de diagrama de bosque para representar los resultados del metaanálisis. Para cada uno de los estudios incluidos se representa la medida del efecto y su intervalo de confianza. La línea punteada representa la línea de efecto nulo. La posición de la medida del efecto con respecto a esta nos informa sobre la significación estadística del resultado global.**

### Selección de genes significativos

A partir de los resultados del metaanálisis seleccionamos los 25 genes con mayor señal biológica en hombres y en mujeres. De los genes con logFC mayor a 0 (mayor expresión en mujeres que en hombres), se seleccionaron los 25 genes con un p-valor menor. De forma análoga, de los genes con logFC menor a 0 (más expresados en hombres que en mujeres) se seleccionaron los 25 genes con un menor p-valor menor. A continuación, se realizó un análisis de sobrerrepresentación de los procesos biológicos utilizando PANTHER<sup>48</sup>, para identificar las funciones en las que participan estos genes.

#### 4.3.1 Análisis de enriquecimiento funcional

A partir de los valores de logFC obtenidos en el metaanálisis es posible realizar un análisis de enriquecimiento funcional mediante GSEA<sup>38</sup> que nos permita identificar las funciones moleculares más sobrerrepresentadas para cada sexo. Para esto se utilizó la librería mdgsa<sup>49</sup>. Se utilizaron las bases de datos Gene Ontology y KEGG para obtener la información funcional. Este enfoque incluye el conjunto de todos los genes identificados en el estudio, ordenados por su nivel de expresión diferencial y no solo una selección de los genes con

mayor expresión diferencial. Son abordajes complementarios que ofrecen diferentes niveles de caracterización funcional.

Para llevar a cabo el enriquecimiento funcional con términos GO fue necesario obtener la información de la anotación mediante el paquete `org.Hs.eg.db`<sup>46</sup> y luego realizar una correspondencia entre estos y los identificadores Entrez de nuestras muestras, teniendo en cuenta la ontología de procesos biológico de GO (GO BP). Se realizó la propagación de la anotación para asegurar que los genes anotados hereden la anotación de los términos ancestros.

En el caso del enriquecimiento funcional con los términos KEGG, se realizó utilizando el paquete `org.Hs.eg.db`<sup>46</sup>. En este caso no fue necesario realizar una propagación de la anotación.

Para ambas bases de datos se filtraron las anotaciones que fueran muy específicas (número de genes menor a 10) o muy generales (número de genes mayor a 500).

A partir de los resultados del enriquecimiento funcional y las anotaciones funcionales se llevó a cabo un GSEA. Se consideraron significativas las funciones que tuvieran un p-valor ajustado por el método BH menor a 0.05.



## 5 Resultados

En este apartado se muestran los resultados obtenidos siguiendo la metodología descrita anteriormente.

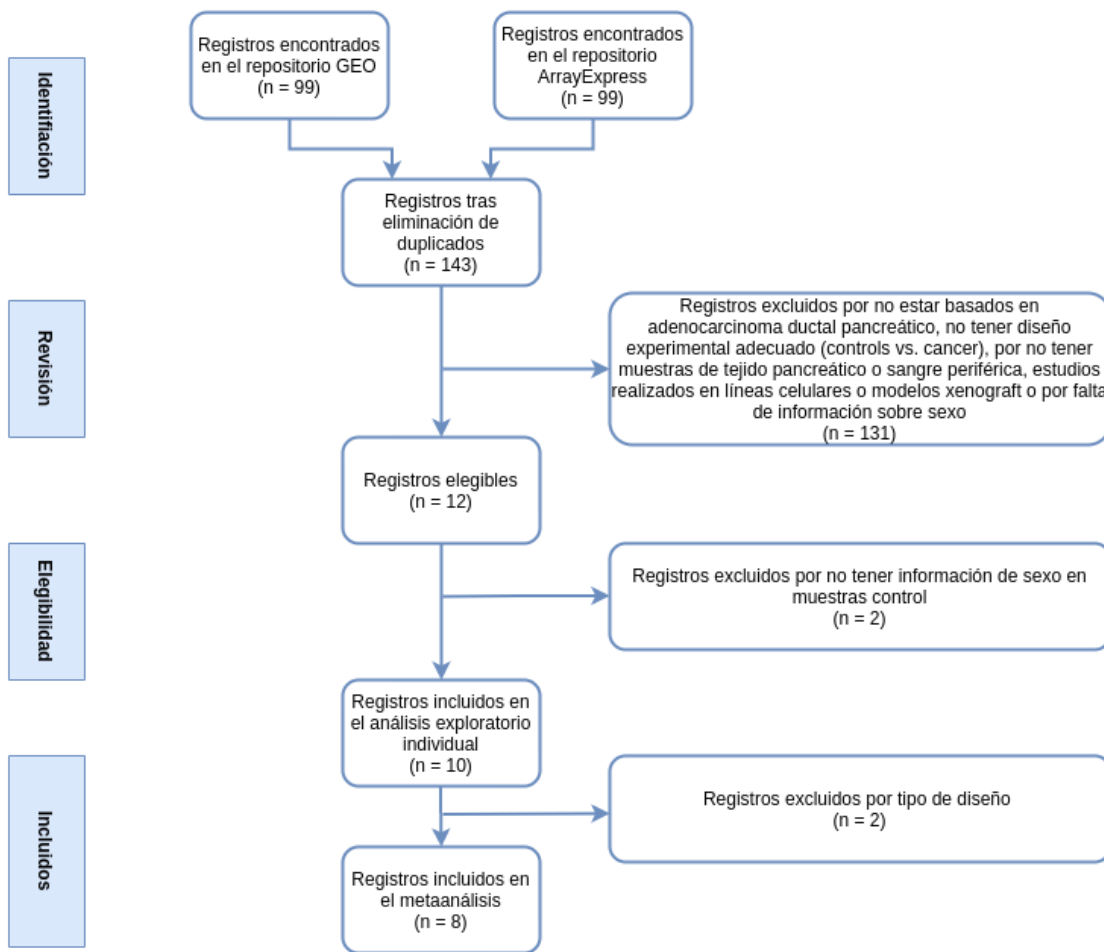
### 5.1 Revisión sistemática y selección de estudios

A partir de la revisión sistemática obtuvimos 99 estudios en GEO y 99 en ArrayExpress. De ellos, 55 aparecían en ambas bases de datos, por lo que finalmente nos quedamos con 143 estudios, los cuales fueron revisados de forma independiente. 131 estudios fueron descartados por no estar centrados en PDAC, no tener diseño experimental adecuado, por haber sido realizados en modelos xenoinjerto o en líneas celulares y por carecer de información de sexo de los pacientes. De los 12 estudios restantes, 2 fueron excluidos por no contar con información de sexo para las muestras de control (GSE43795 y GSE56560). Se intentó contactar con los responsables de ambos estudios para que nos brindaran esta información, pero no obtuvimos respuesta. Los 10 estudios seleccionados fueron descargados para realizar el análisis exploratorio, pero 2 ellos fueron descartados: GSE102238 por tener una tecnología de microarray con la anotación centrada en secuencias long noncoding RNA (lncRNA) y el GSE39751 por tratarse de un array comparativo de dos colores, con gran número de NA y una normalización que centra la media de todas las muestras en 0 con una distribución de entre -5 y +5.

Finalmente 8 estudios fueron seleccionados para realizar el metaanálisis.

| Estudios | Plataforma  | Publicación |
|----------|---|-------------|
| GSE15932 | Affymetrix Human Genome U133 Plus 2.0 Array               | No presenta |
| GSE16515 | Affymetrix Human Genome U133 Plus 2.0 Array               | 50-52       |
| GSE22780 | Affymetrix Human Genome U133 Plus 2.0 Array               | No presenta |
| GSE18670 | Affymetrix Human Genome U133 Plus 2.0 Array               | 53          |
| GSE49641 | Affymetrix Human Gene 1.0 ST Array                        | 54,55       |
| E-EMBL-6 | Affymetrix GeneChip Human Genome U95Av2                   | 56          |
| GSE55643 | Agilent-014850 Whole Human Genome Microarray 4x44K G4112F | 57          |
| GSE74629 | Illumina HumanHT-12 V4.0 expression beadchip              | 54          |

**Tabla 5.1: Estudios seleccionados tras la revisión sistemática. Se indica el código del estudio, la plataforma en que fue realizado y su publicación asociada, si posee.**



**Figura 5.1: Diagrama de la revisión sistemática realizada siguiendo la declaración PRISMA.**

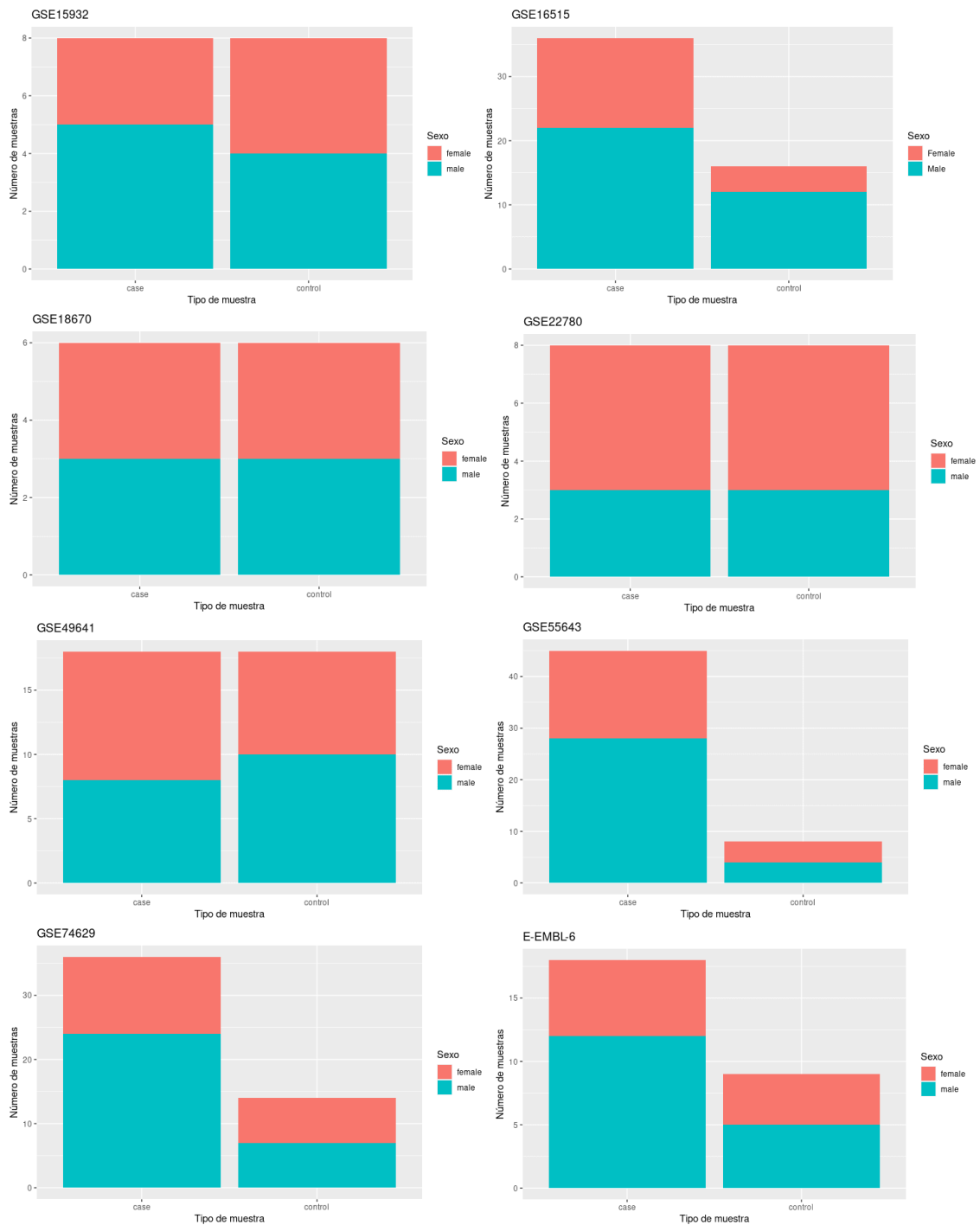
## 5.2 Análisis individual de los estudios

A continuación, se detalla el análisis individual para cada uno de los estudios seleccionados.

### 5.2.1 Procesamiento de los datos

Tras descargar los datos y estandarizar la nomenclatura, se procedió a explorar los datos de cada estudio. En el caso de los estudios GSE15932, GSE16515, GSE22780 y E-EMBL-6 fue necesario realizar una transformación logarítmica en base 2 de los datos normalizados para conseguir una escala homogénea entre todos los estudios. Los diagramas de cajas obtenidos se muestran en el Anexo I.

La distribución de muestras por estudio se muestra en la Figura 5.2.



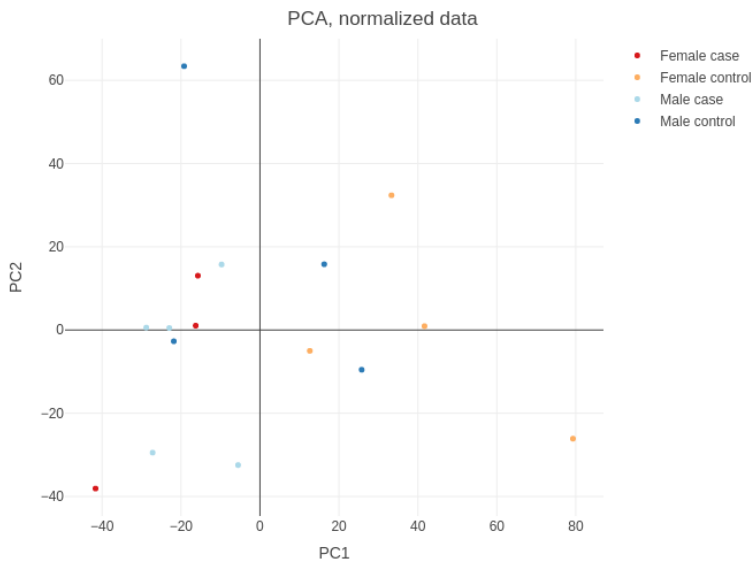
**Figura 5.2: Distribución de las muestras para cada uno de los estudios analizados. Case: pacientes con diagnóstico de PDAC, control: pacientes sin enfermedad.**

### GSE15932

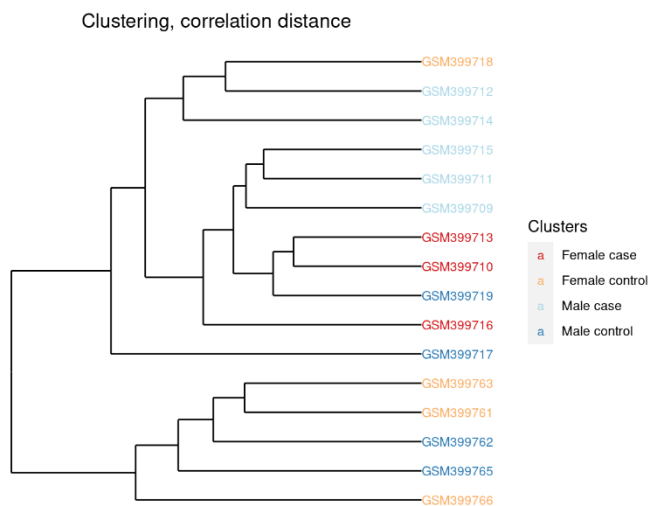
En este estudio se analizan biomarcadores sanguíneos de cáncer de páncreas en asociación con diabetes mellitus, estudiando los perfiles de expresión génica en muestras de sangre periférica. Para nuestro trabajo seleccionamos únicamente las muestras de pacientes con diagnóstico de PDAC sin diabetes mellitus y los controles sanos, sin diagnóstico de diabetes. En el siguiente cuadro se muestra la distribución de las muestras.

|         | Male | Female | Total |
|---------|------|--------|-------|
| Case    | 5    | 3      | 8     |
| Control | 4    | 4      | 8     |
| Total   | 9    | 7      | 16    |

**Tabla 5.2: Distribución de muestras por sexo y condición de estudio GSE15932.**



**Figura 5.3: Análisis de componentes principales del estudio GSE15932.**



**Figura 5.4: Clustering de las muestras del estudio GSE15932.**

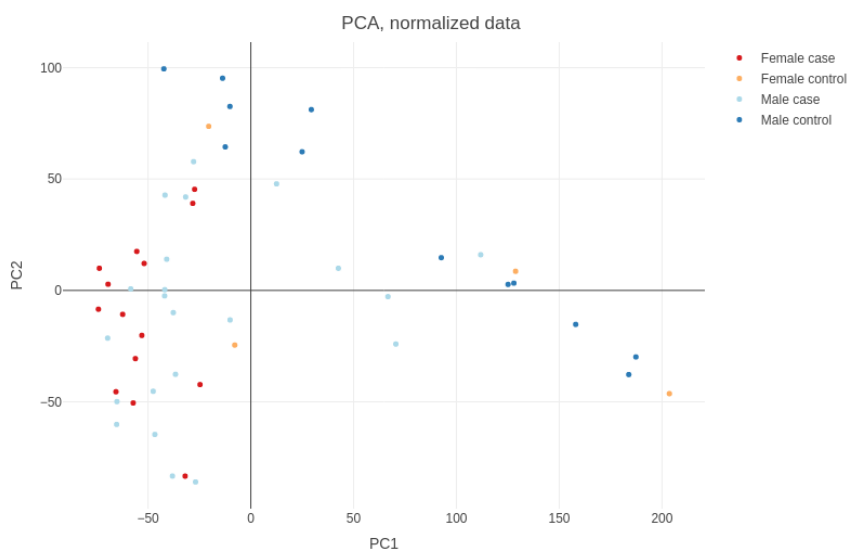
Tanto en el análisis de componentes principales como el en clustering se observa una separación de las muestras por condición de la enfermedad más que por sexo.

## GSE16515

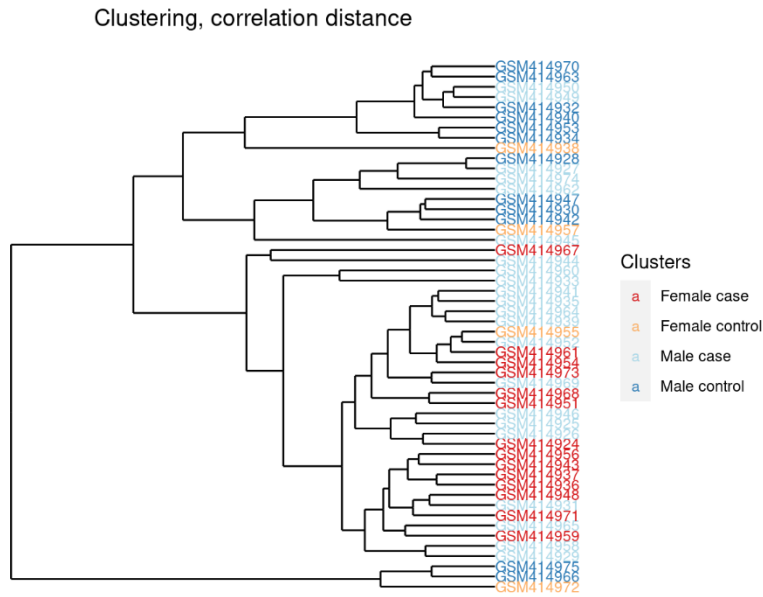
En este estudio se comparan los niveles de expresión del gen *FKBP5* en muestras de tejido pancreático tumoral y tejido sano control. A continuación se muestra la distribución de las muestras del estudio.

|         | Male | Female | Total |
|---------|------|--------|-------|
| Case    | 22   | 14     | 36    |
| Control | 12   | 4      | 16    |
| Total   | 34   | 18     | 52    |

**Tabla 5.3: Distribución de muestras por sexo y condición de estudio GSE16515.**



**Figura 5.5: Análisis de componentes principales del estudio GSE16515.**



**Figura 5.6: Clustering de las muestras del estudio GSE16515.**

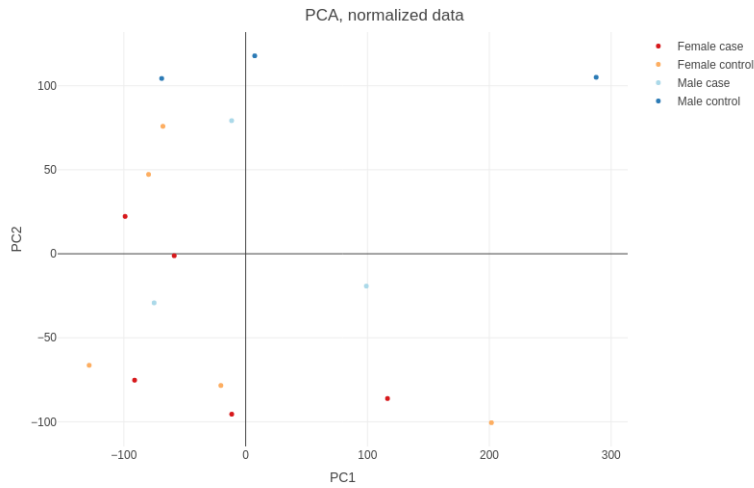
Tanto en el análisis de componentes principales como el en clustering se observa un mayor agrupamiento de las muestras por condición de la enfermedad más que por sexo.

### GSE22780

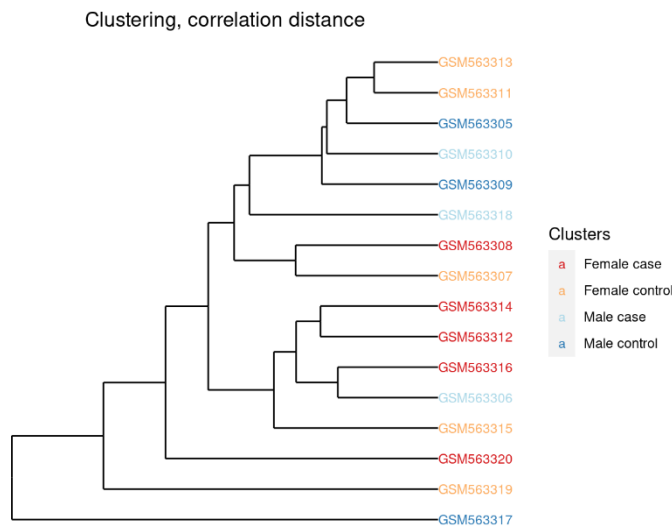
Este estudio se centra en identificar genes de la región cromosómica 3p12 que participen en la vía de tumorigénesis de cáncer de páncreas, estudiando la diferencia de expresión entre muestras de cáncer de páncreas y tejido sano. A continuación se muestra la distribución de las muestras de este estudio.

|         | Male | Female | Total |
|---------|------|--------|-------|
| Case    | 3    | 5      | 8     |
| Control | 3    | 5      | 8     |
| Total   | 6    | 10     | 16    |

**Tabla 5.4: Distribución de muestras por sexo y condición de estudio GSE22780.**



**Figura 5.7: Análisis de componentes principales del estudio GSE22780.**



**Figura 5.8: Clustering de las muestras del estudio GSE22780.**

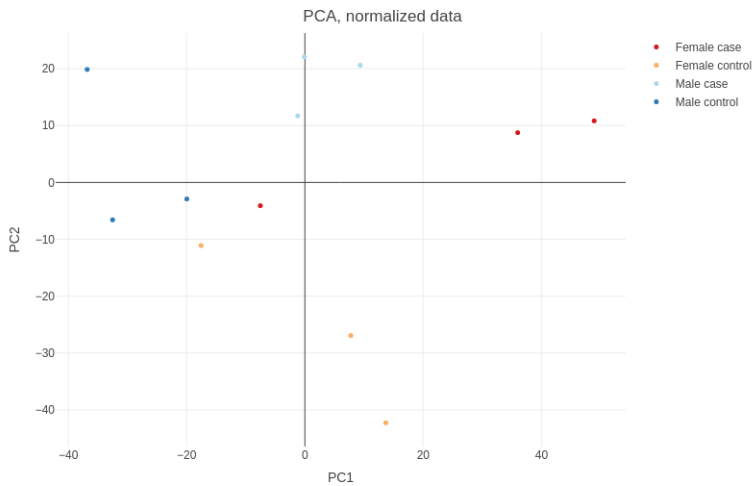
No se observa agrupación de las muestras por condición ni por sexo en el análisis de componentes principales ni en el clustering.

### **GSE18670**

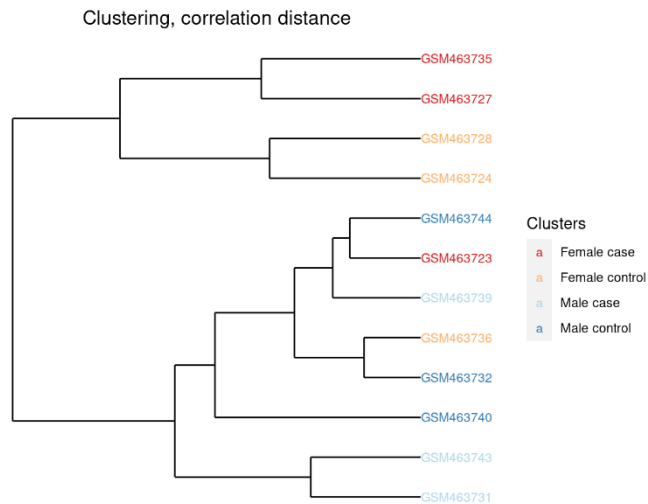
El objetivo de este estudio es conocer la expresión génica de las células tumorales circulantes (CTC), que se desprenden del cáncer primario y desarrollan metástasis, y evaluar su relevancia en el pronóstico de la enfermedad luego de realizar una cirugía de resección de PDAC. Para esto se compararon los perfiles de expresión de muestras de CTC, células hematológicas, tumor y tejido sano control. Para nuestro trabajo seleccionamos las muestras de tumor y de tejido sano. A continuación se muestra la distribución de las muestras del estudio.

|         | Male | Female | Total |
|---------|------|--------|-------|
| Case    | 3    | 3      | 6     |
| Control | 3    | 3      | 6     |
| Total   | 6    | 6      | 12    |

**Tabla 5.5: Distribución de muestras por sexo y condición de estudio GSE18670.**



**Figura 5.9: Análisis de componentes principales del estudio GSE18670.**



**Figura 5.10: Clustering de las muestras del estudio GSE18670.**

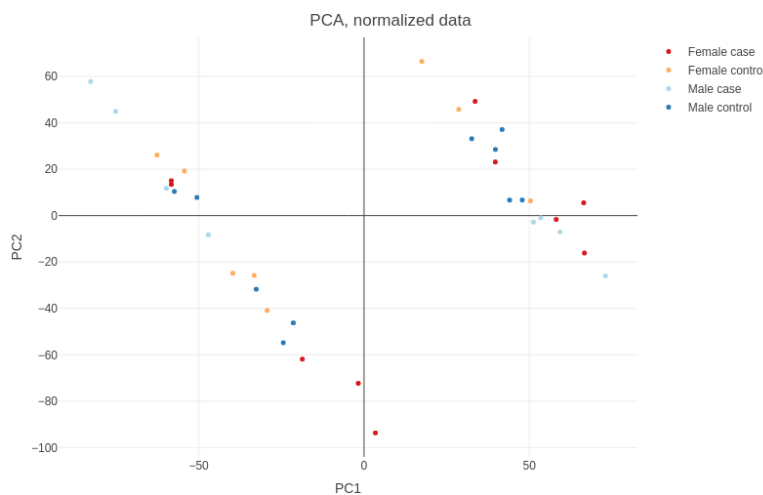
Tanto en el análisis de componentes principales como el en clustering se observa un mayor agrupamiento de las muestras por sexo que por condición de la enfermedad.

**GSE49641**

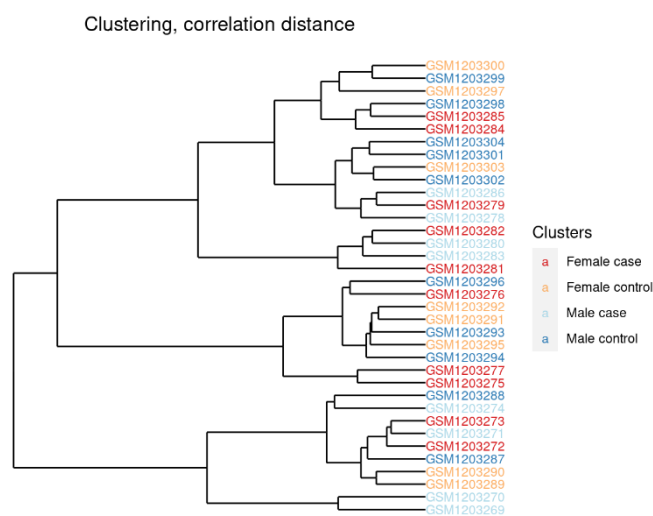
El objetivo de este estudio es identificar en sangre periférica nuevos marcadores que diferencian a pacientes con PDAC de controles sanos, para facilitar la detección rápida de la enfermedad, a partir del estudio del perfil de expresión génica de células mononucleares de sangre periférica. A continuación se muestra la distribución de las muestras para este estudio.

|         | Male | Female | Total |
|---------|------|--------|-------|
| Case    | 8    | 10     | 18    |
| Control | 10   | 8      | 18    |
| Total   | 18   | 18     | 32    |

**Tabla 5.6: Distribución de muestras por sexo y condición de estudio GSE49641.**



**Figura 5.11: Análisis de componentes principales del estudio GSE49641.**



**Figura 5.12: Clustering de las muestras del estudio GSE49641.**

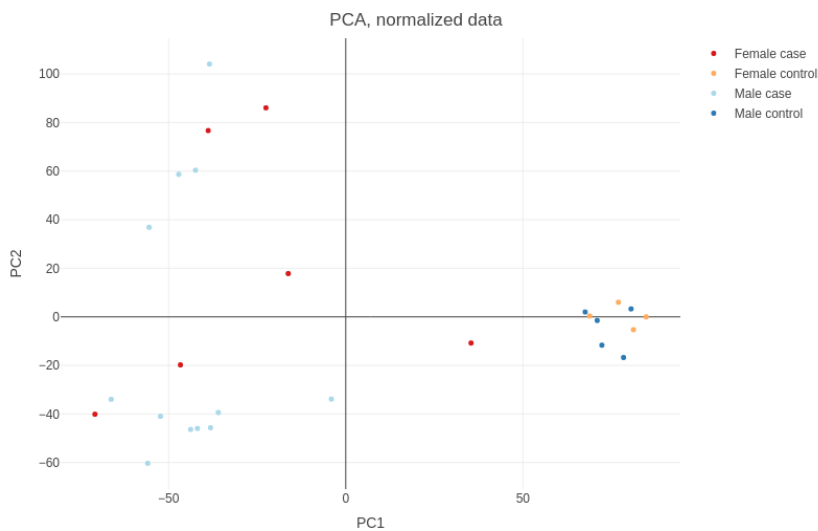
En el análisis de componentes principales se observa un agrupamiento de las muestras que no es debido al sexo ni a la condición de la enfermedad, sino que se debe a un factor no detectado (efecto batch). Esto fue tenido en cuenta en los análisis posteriores.

### E-EMBL-6

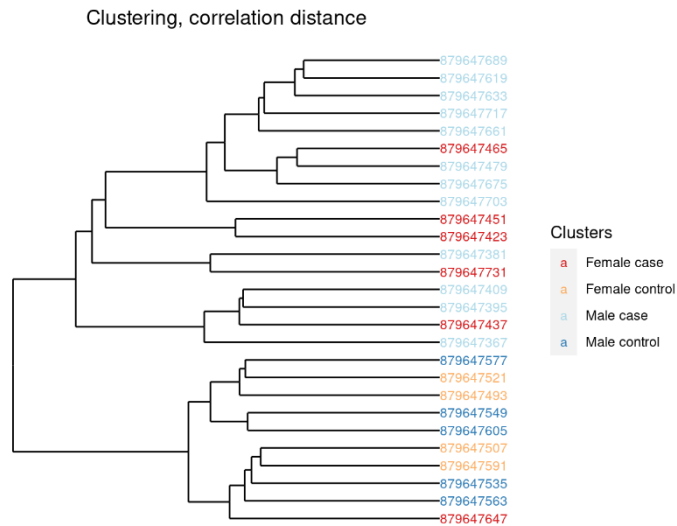
En este estudio se estudian los perfiles de expresión génica de muestras de cáncer de páncreas localizado, cáncer de páncreas metastásico, pancreatitis crónica y tejido sano. Para nuestro trabajo descartamos las muestras de pancreatitis crónica. A continuación se muestra la distribución de las muestras para este estudio.

|         | Male | Female | Total |
|---------|------|--------|-------|
| Case    | 8    | 12     | 20    |
| Control | 10   | 5      | 15    |
| Total   | 18   | 17     | 35    |

**Tabla 5.7: Distribución de muestras por sexo y condición de estudio E-EMBL-6.**



**Figura 5.13: Análisis de componentes principales del estudio E-EMBL-6.**



**Figura 5.14: Clustering de las muestras del estudio E-EMBL-6.**

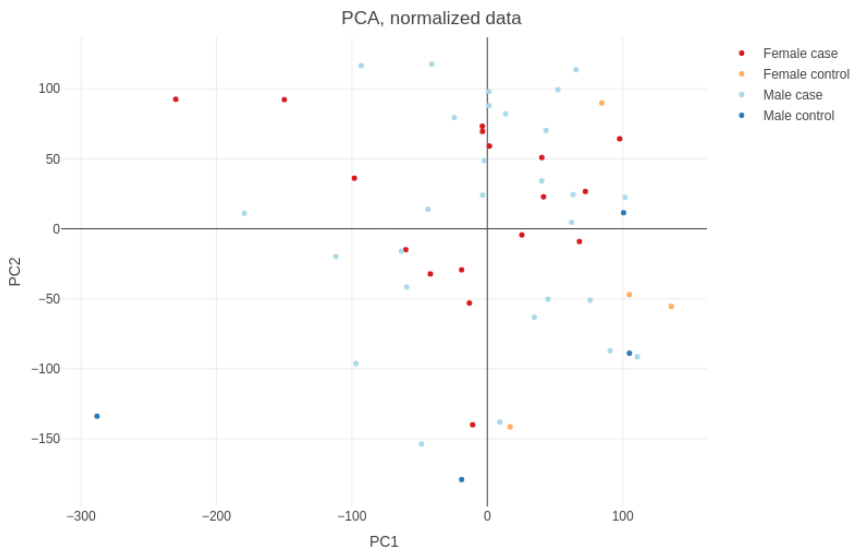
Tanto en el análisis de componentes principales como en el clustering se puede observar un agrupamiento de las muestras por condición de la enfermedad, a excepción de una de las muestras.

### GSE55643

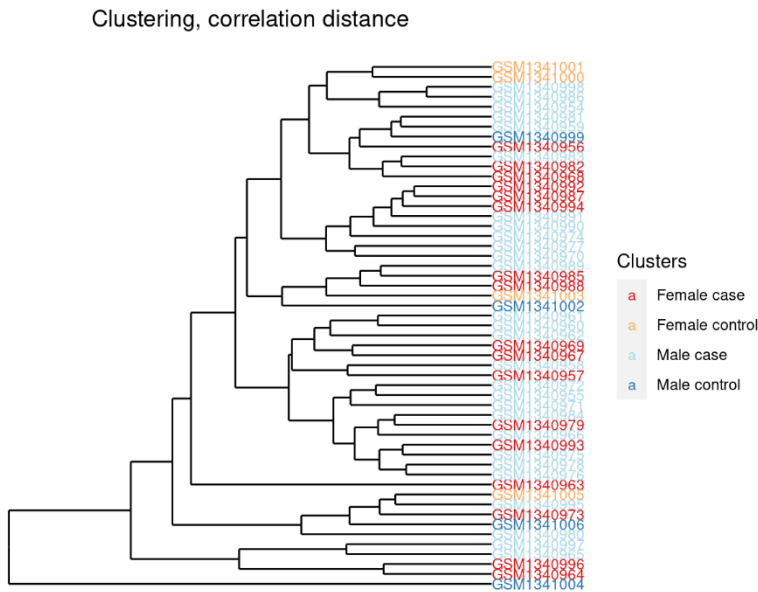
En este estudio se analiza la expresión génica del PDAC en muestras de tejido pancreático. A continuación se muestra la distribución de las muestras del estudio.

|         | Male | Female | Total |
|---------|------|--------|-------|
| Case    | 28   | 17     | 45    |
| Control | 4    | 4      | 8     |
| Total   | 32   | 21     | 53    |

**Tabla 5.8: Distribución de muestras por sexo y condición del estudio GSE55643.**



**Figura 5.15: Análisis de componentes principales del estudio GSE55643.**



**Figura 5.16: Clustering de las muestras del estudio GSE55643.**

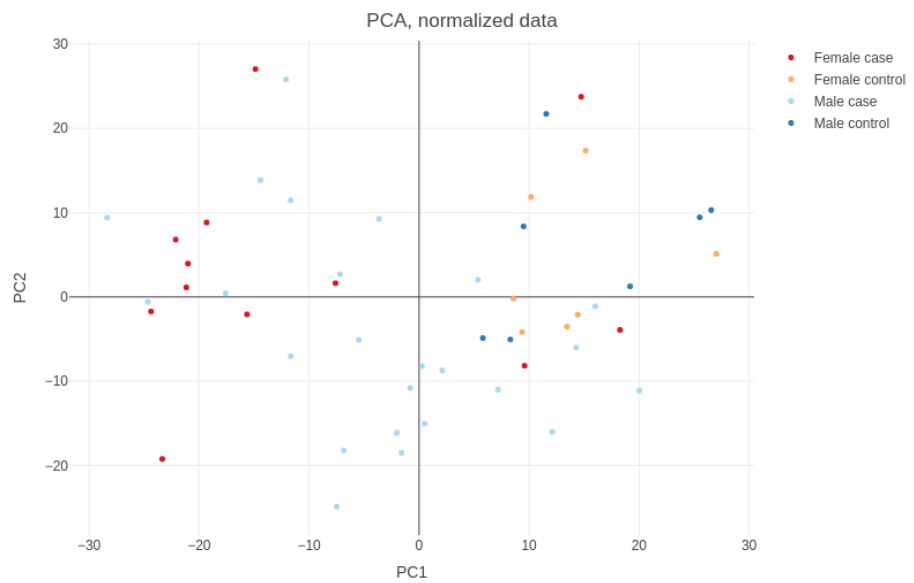
No se observa agrupación de las muestras por condición ni por sexo en el análisis de componentes principales ni en el clustering.

## GSE74629

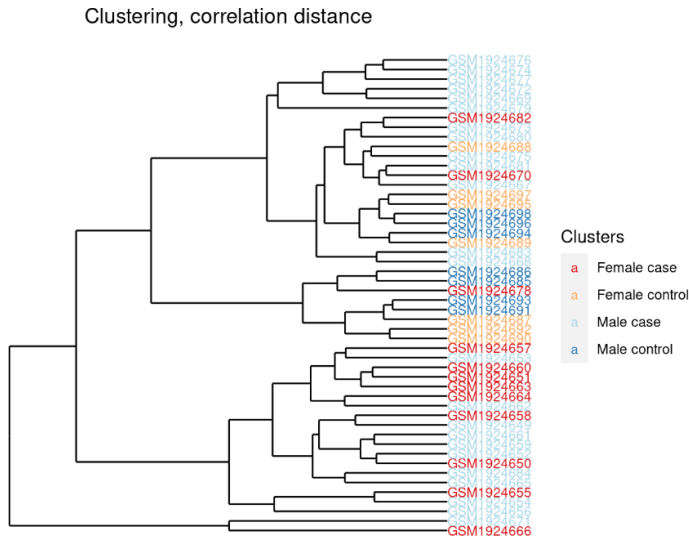
En este estudio se realiza un análisis del perfil de expresión de células de sangre periférica de pacientes con PDAC y de pacientes sanos. A continuación se muestra la distribución de las muestras para este estudio.

|         | Male | Female | Total |
|---------|------|--------|-------|
| Case    | 24   | 12     | 36    |
| Control | 7    | 7      | 14    |
| Total   | 31   | 19     | 50    |

**Tabla 5.9: Distribución de muestras por sexo y condición del estudio GSE74629.**



**Figura 5.17: Análisis de componentes principales del estudio GSE74629.**



**Figura 5.18: Clustering de las muestras del estudio GSE74629.**

No se observa agrupación de las muestras por condición ni por sexo en el análisis de componentes principales ni en el clustering.

### 5.2.2 Análisis de expresión diferencial

A continuación, se muestran los resultados obtenidos del análisis de expresión diferencial para el contraste de interés para cada uno de los estudios (Tabla 5.10). También se muestran los resultados de los contrastes de mujeres con diagnóstico de PDAC y mujeres control y de hombres con PDAC y hombres control, realizados con fines exploratorios.

Los resultados se consideran significativos cuando presentan un p-valor ajustado por BH menor a 0.05. Aquellos que presentan logFC mayor a cero fueron agrupados como Up, indicando que esos genes se encuentran sobreexpresados en mujeres en relación con los hombres, y down cuando el logFC es menor a 0.

|          |      | (FP-FC) - (MP-MC) | FP - FC | MP - MC |
|----------|------|-------------------|---------|---------|
| GSE15932 | Up   | 0                 | 0       | 0       |
|          | Down | 0                 | 0       | 0       |
| GSE16515 | Up   | 0                 | 1635    | 2226    |
|          | Down | 0                 | 2517    | 2751    |
| GSE22780 | Up   | 0                 | 0       | 0       |
|          | Down | 0                 | 0       | 0       |

|          |      |   |     |     |
|----------|------|---|-----|-----|
| GSE18670 | Up   | 0 | 0   | 1   |
|          | Down | 0 | 0   | 0   |
| GSE49641 | Up   | 0 | 37  | 2   |
|          | Down | 0 | 61  | 1   |
| E-EMBL-6 | Up   | 0 | 387 | 804 |
|          | Down | 0 | 287 | 610 |
| GSE55643 | Up   | 0 | 3   | 18  |
|          | Down | 0 | 0   | 30  |
| GSE74629 | Up   | 0 | 175 | 81  |
|          | Down | 1 | 211 | 212 |

**Tabla 5.10: Resumen de los resultados del análisis de expresión diferencial. FP: female PDAC, FC: female control, MP: male PDAC, MC: male control. Up: genes que presentan logFC mayor a 0 y Down: genes con logFC menor a 0.**

Para la comparación de interés obtuvimos únicamente 1 gen que se encuentra sobreexpresado en hombres, en el estudio GSE74629. Este gen es el *KDM5D* que codifica para una proteína (Lysine Demethylase 5D) que contiene dominios de dedos de zinc. Un péptido derivado de esta proteína es un antígeno de histocompatibilidad menor que puede conducir al rechazo del injerto de células de donantes masculinos en una receptora femenina.

### 5.3 Metaanálisis de genes

En el metaanálisis de genes no se obtuvieron genes significativos con un p-valor ajustado por BH menor a 0.05. De la lista de genes que obtuvimos una vez realizado el metaanálisis, seleccionamos los 25 genes más expresados en hombres y en mujeres.

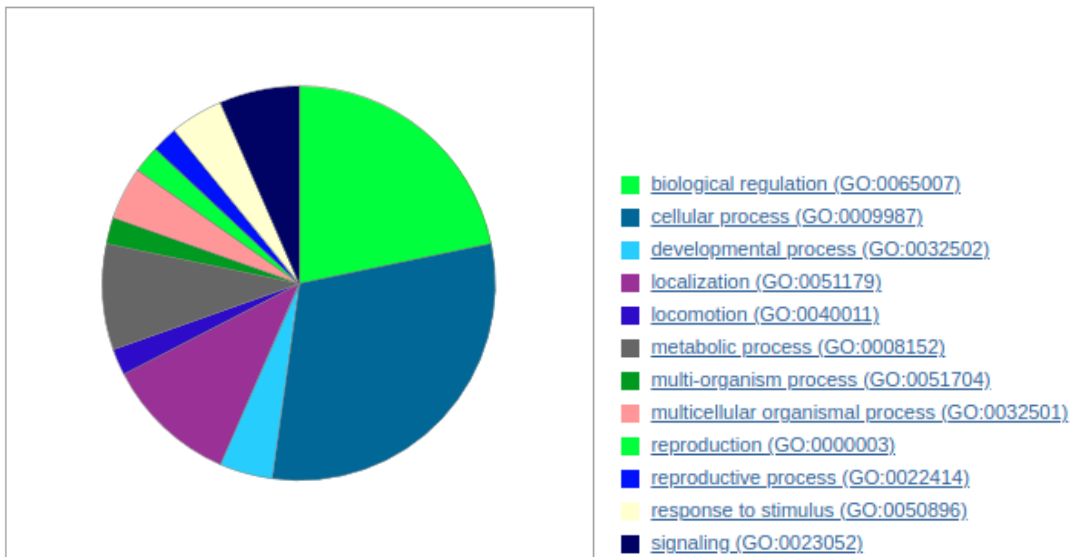
| ENTREZID  | LogFC | p-valor | SYMBOL      |
|-----------|-------|---------|-------------|
| 1299      | 0.480 | 0.000   | COL9A3      |
| 162427    | 0.226 | 0.000   | RETREG3     |
| 23416     | 0.250 | 0.000   | KCNH3       |
| 23584     | 0.100 | 0.000   | VSIG2       |
| 55323     | 0.117 | 0.000   | LARP6       |
| 80148     | 0.368 | 0.000   | SLC66A2     |
| 100127888 | 0.430 | 0.001   | SLCO4A1-AS1 |
| 10279     | 0.101 | 0.001   | PRSS16      |
| 170712    | 0.077 | 0.001   | COX7B2      |
| 2245      | 0.342 | 0.001   | FGD1        |
| 23623     | 0.302 | 0.001   | RUSC1       |
| 2897      | 0.057 | 0.001   | GRIK1       |
| 50855     | 0.230 | 0.001   | PARD6A      |
| 65267     | 0.213 | 0.001   | WNK3        |
| 10654     | 0.210 | 0.002   | PMVK        |
| 140578    | 0.074 | 0.002   | CHODL       |
| 149699    | 0.102 | 0.002   | GTSF1L      |
| 155465    | 0.446 | 0.002   | AGR3        |
| 610       | 0.096 | 0.002   | HCN2        |
| 6795      | 0.055 | 0.002   | AURKC       |
| 8226      | 0.125 | 0.002   | PUDP        |
| 90874     | 0.365 | 0.002   | ZNF697      |
| 25858     | 0.370 | 0.003   | CATSPERZ    |
| 2911      | 0.078 | 0.003   | GRM1        |
| 51042     | 0.209 | 0.003   | ZNF593      |

**Tabla 5.11: Listado de los 25 genes más expresados en mujeres. Se muestra el identificador ENTREZ, el logaritmo de fold change (logFC), el p-valor y el nombre del gen (SYMBOL).**

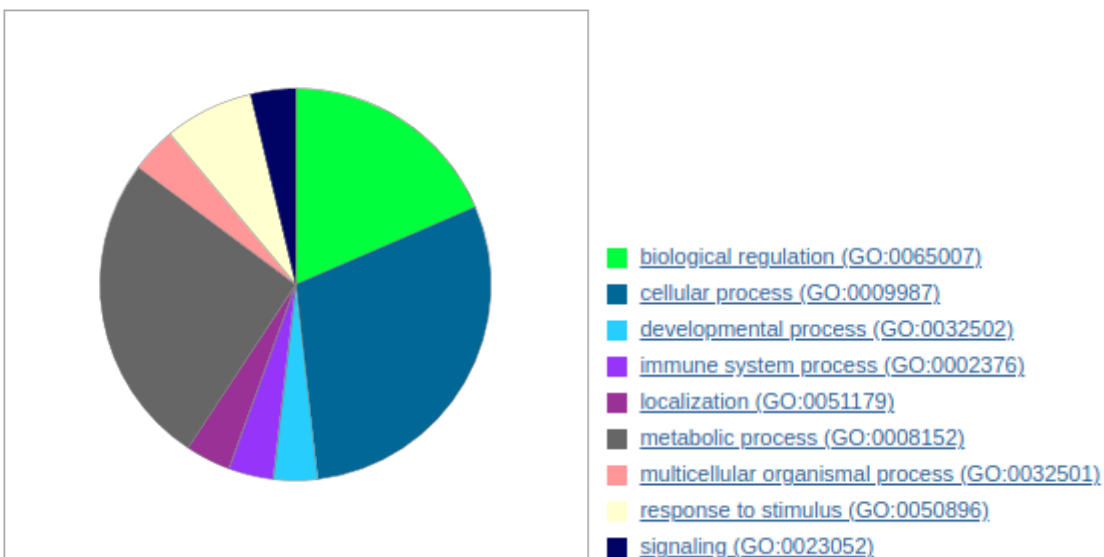
| ENTREZID  | LogFC  | p-valor | SYMBOL       |
|-----------|--------|---------|--------------|
| 100505635 | -0.430 | 0.000   | DAAM2-AS1    |
| 100506047 | -0.595 | 0.000   | LINC02580    |
| 101926960 | -0.557 | 0.000   | LOC101926960 |
| 202559    | -0.112 | 0.000   | KHDRBS2      |
| 22973     | -0.492 | 0.000   | LAMB2P1      |
| 2947      | -0.472 | 0.000   | GSTM3        |
| 340184    | -0.819 | 0.000   | LOC340184    |
| 3492      | -0.886 | 0.000   | IGH          |
| 100506305 | -0.461 | 0.001   | LINC00958    |
| 100506472 | -0.486 | 0.001   | LOC100506472 |
| 101927699 | -0.351 | 0.001   | LOC101927699 |
| 101929549 | -0.654 | 0.001   | GAU1         |
| 246269    | -0.229 | 0.001   | AFG1L        |
| 28902     | -0.554 | 0.001   | IGKV1D-13    |
| 338755    | -0.119 | 0.001   | OR2AG2       |
| 9298      | -0.578 | 0.001   | SNORD31      |
| 9562      | -0.096 | 0.001   | MINPP1       |
| 130399    | -0.078 | 0.002   | ACVR1C       |
| 3841      | -0.076 | 0.002   | KPNA5        |
| 5704      | -0.148 | 0.002   | PSMC4        |
| 100131303 | -0.426 | 0.003   | DPEP2NB      |
| 10961     | -0.070 | 0.003   | ERP29        |
| 163081    | -0.074 | 0.003   | ZNF567       |
| 201626    | -0.132 | 0.003   | PDE12        |
| 444       | -0.088 | 0.003   | ASPH         |

**Tabla 5.12: Listado de los 25 genes más expresados en hombres. Se muestra el identificador ENTREZ, el logaritmo de fold change (logFC), el p-valor y el nombre del gen (SYMBOL).**

El análisis realizado con PANTHER de estos genes nos muestra las funciones relacionadas con ellos para cada uno de los sexos (Figuras 5.19 y 5.20).



**Figura 5.19: Funciones relacionadas a los genes más expresados en mujeres obtenidas con PANTHER.**



**Figura 5.20: Funciones relacionadas a los genes más expresados en hombres obtenidas con PANTHER.**

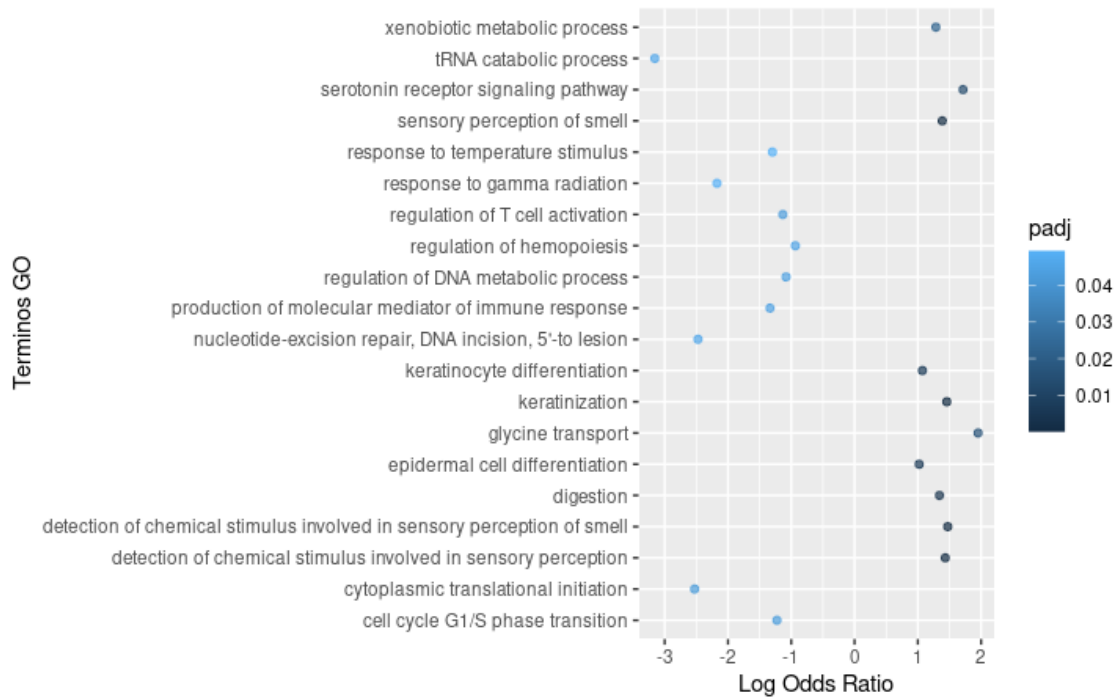
### 5.3.1 Análisis de enriquecimiento funcional

Los resultados del análisis de GSEA para términos GO de la ontología BP y rutas KEGG para los resultados del metaanálisis se resumen en la Tabla 5.13.

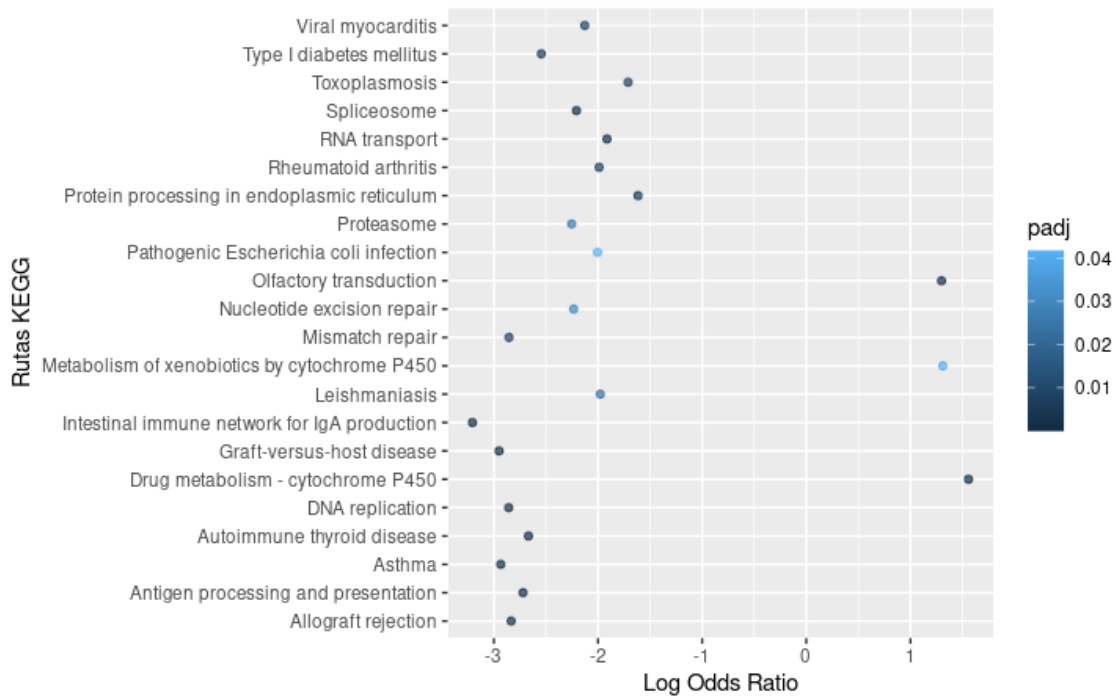
|       | Up | Down | Total |
|-------|----|------|-------|
| GO BP | 16 | 141  | 157   |
| KEGG  | 2  | 15   | 17    |

**Tabla 5.13: Resumen de resultados de enriquecimiento funcional de los resultados del metaanálisis. Up: funciones que presentan logFC mayor a 0 y están sobrerrepresentadas en mujeres y down: funciones que presentan logFC menor a 0 y están sobrerrepresentadas en hombres.**

A continuación, en las Figuras 5.21 y 5.22, se muestran los resultados más significativos para la ontología BP de GO y las rutas KEGG. Los resultados completos se muestran en el Anexo II (Tablas A 1 y A 2).



**Figura 5.21: Términos GO BP sobrerrepresentados. Aquellos términos con Log Odds Ratio menor a 0 corresponden a funciones más representadas en hombres y los términos con Log Odds Ratio mayor a 0 a funciones más representadas en mujeres.**



**Figura 5.22: Rutas KEGG sobrerrepresentadas. Aquellas rutas con Log Odds Ratio menor a 0 corresponden a rutas más expresadas en hombres y las rutas con Log Odds Ratio mayor a 0 a rutas más expresadas en mujeres.**

## 6 Discusión

Está descrito en la literatura que el cáncer de páncreas afecta en mayor medida a hombres que a mujeres, pero todavía se desconocen las razones que lo producen, aunque se cree que podría tener relación con factores hormonales y ambientales. Conocer los motivos que generan estas diferencias es imprescindible para llevar a cabo una medicina personalizada ya que permitiría desarrollar terapias específicas más efectivas y dirigidas que modifiquen el pronóstico de esta enfermedad.

Actualmente, los estudios *in silico* constituyen herramientas rápidas y muy rentables que pueden ayudar a predecir y validar los efectos de una variable sobre la salud humana. Concretamente en este trabajo se ha evaluado la variable sexo en PDAC que relativamente hace poco tiempo ha sido reconocida por el instituto de salud nacional (NHI, National Institute of Health) como una variable biológica a incluir en las investigaciones<sup>58</sup>. Para el desarrollo del presente trabajo ha sido necesario utilizar datos con perspectivas de sexo de otros estudios disponibles en repositorios públicos. Sin embargo, cuando revisamos estas bases de datos, nos encontramos con que la mayoría de los estudios no incluyen la información del sexo de sus muestras. No contar con estos datos dificulta la identificación de biomarcadores que expliquen la diferencia en la incidencia de PDAC entre hombres y mujeres, así como la detección de asociaciones que puedan favorecer estas diferencias. Por ello, es importante incluir en el diseño de los estudios una perspectiva de sexo, además de contar con datos estandarizados que permitan su reutilización. Con este objetivo se crearon los principios FAIR que regulan las publicaciones científicas. Estos ofrecen un conjunto de cualidades precisas y medibles que una publicación de datos debería seguir para que los datos sean Encontrables, Accesibles, Interoperables y Reutilizables (del inglés FAIR – Findable, Accessible, Interoperable, and Reusable)<sup>59</sup>. Actualmente se exige cada vez más a las investigaciones que cumplan con estos requisitos. De esta manera se podrá contar con mejor información, facilitará su acceso y permitirá el desarrollo de investigaciones con mejor evidencia.

Con el objetivo de encontrar mecanismos moleculares que expliquen las diferencias entre hombres y mujeres que sufren cáncer de páncreas, se han analizado los datos de aquellos estudios centrados en PDAC que contaran con información de sexo provenientes de repositorios públicos mediante una estrategia de metaanálisis, un método robusto que permite integrar información de diferentes estudios individuales para obtener mayor potencia estadística<sup>30,32</sup>. Este método resultó el más adecuado para este abordaje debido a la disparidad de datos no estandarizados, provenientes de distintas plataformas y estudios.

En la literatura hemos encontrado estudios que buscan determinar biomarcadores en PDAC a través de una estrategia de metaanálisis de datos transcriptómicos<sup>54,60</sup> analizando datos seleccionados de repositorios públicos, pero en ninguno de ellos se incluye la perspectiva de sexo.

A través del metaanálisis de genes realizado no se han detectado genes que presentaran un p-valor ajustado por BH menor a 0.05, es decir significativos. Sin embargo, este perfil transcriptómico de la expresión diferencial en el conjunto de todos los genes evaluados, permitió identificar la relevancia funcional en el cáncer de páncreas considerando la perspectiva del sexo, mediante el enriquecimiento funcional sobre este perfil transcriptómico consenso.

El análisis de enriquecimiento funcional caracteriza en forma funcional los perfiles de expresión génica metaanalizados. y ha mostrado que tanto para términos GO como para rutas KEGG se han encontrado una mayor cantidad de funciones significativas en hombres que en mujeres.

Este análisis nos muestra que en mujeres están sobreexpresados términos relacionados con la detección de estímulos químicos implicados en la percepción sensorial (GO:0050907) y específicamente en la percepción olfatoria (GO:0050911, GO:0007608, map04740). Existen estudios que evidencian que las neuronas sensitivas tienen un rol fundamental en el desarrollo de enfermedades inflamatorias del páncreas como la pancreatitis, enfermedad de riesgo para el desarrollo de cáncer o PanIN<sup>61</sup>. La ablación de estas neuronas demostró un efecto protector frente al cáncer de páncreas.

Otras funciones que resultan representativas en mujeres son las relacionadas con el desarrollo del microambiente tumoral, como el crecimiento de células epiteliales (GO:0009913), diferenciación de queratinocitos (GO:0031424, GO:0030216), desarrollo de epidermis y de piel (GO:0008544, GO:0043588). Está estudiado que en tejido tumoral pancreático existe sobreexpresión del factor de crecimiento de queratinocitos (KGF)<sup>62,63</sup> y del factor de crecimiento epidérmico (EGFR) que juegan un papel importante en la estimulación del factor de crecimiento endotelial vascular (VEGF) e IL-8, que favorecen la proliferación de células tumorales, la angiogénesis y metástasis<sup>64</sup>.

También se encuentran sobrerrepresentados en mujeres términos asociados el transporte de aminoácidos como alanina (GO:0015808) o glicina (GO:0015816), transmisión de señales de neurotransmisores como serotonina (GO:0007210, GO:0007218) y con respuesta celular a estímulos xenobióticos (GO:0006805, GO:0009410, GO:0071466, map00982 y map00980). Recientemente, se ha publicado que los axones periféricos liberan serina para favorecer el crecimiento de células de PDAC dependientes de serina exógena durante la privación de serina y glicina en un ambiente pobre en nutrientes<sup>65</sup>.

Cuando observamos las funciones que resultan más representativas en los hombres nos encontramos con aquellas que están en relación con la regulación de la respuesta inmune. Los términos GO:0042110, GO:0050870, GO:0001912, GO:0002708, GO:0002696, GO:0050864, GO:0042113 y map05332 están relacionados con la regulación de linfocitos tanto T como B. Como describimos anteriormente, los linfocitos T CD8 + son componentes importantes de la inmunidad adaptativa celular específica del tumor. Su activación a través de la presentación de antígenos es crucial para efectuar su actividad antitumoral. Encontramos sobreexpresadas funciones relacionadas con la presentación de antígenos (GO:0048002, GO:0019886, GO:0019882, GO:0002478, GO:0019884, GO:0050851, GO:0002495, GO:0002504). Los términos GO:0002717 y GO:0045954 hacen referencia a la regulación de las células NK, la actividad de estas células está disminuida en el cáncer de páncreas, ya que en condiciones normales estas células atacan a las células tumorales<sup>19</sup>. Además de las funciones relacionadas con la activación de las células inmunes también encontramos términos relacionados con la regulación del sistema inmune a través de la producción de mediadores como el interferón gamma (IFN- $\gamma$ ) o activación de inmunoglobulinas (GO:0002440, GO:0032729, GO:0002381, GO:0034341, GO:0002455, GO:0002460, GO:0002377). En definitiva, una respuesta inmune crónica y exagerada contribuye y acelera los procesos neoplásicos. Se ha descrito también que la respuesta inmune innata puede contribuir al estado de caquexia en PDAC<sup>66</sup>.

Otras funciones que aparecen más representadas en los hombres son aquellas relacionadas con la regulación del ciclo celular (GO:0044843, GO:0045930, GO:1901988, GO:1901991,

GO:0000082). En las lesiones precursoras de PDAC se observan alteraciones genéticas que generan displasia y atipia celular que progresan hasta generar las alteraciones del cáncer. La adquisición secuencial de mutaciones en el protooncogén *KRAS* y los supresores de tumores *INK4A*, *TP53* y *DPC4/SMAD4* conduce a una profunda alteración en la regulación del ciclo celular, un sello distintivo del cáncer de páncreas<sup>67</sup>. Estas alteraciones conllevan daños en el ADN, lo que explica que también encontremos sobreexpresadas funciones relacionadas con la reparación del ADN (GO:0000731, GO:0006296, GO:0042769, map03420, map03430, map03030). La incapacidad para reparar el ADN dañado produce acumulación de defectos genéticos que generan una transformación maligna posterior con progresión al cáncer<sup>68</sup>.

Por último y no menos importante, también encontramos sobrerrepresentadas en hombres funciones relacionadas con el metabolismo de ARN (GO:0016078, GO:0009451, GO:0006406, GO:0071427, GO:0050658, GO:0051236, map 03013) y de respuesta a estímulos externos (GO:0009266, GO:0010332, GO:0010212, GO:0002831, GO:0002833) como estímulos bióticos, radiación o calor que podrían sugerir, por ejemplo, un papel en la respuesta a tratamiento.

Utilizamos la herramienta Open Target<sup>69</sup> para caracterizar los genes más representativos tanto para hombres como para mujeres. Buscamos aquellos para los que se conociera la relación con cáncer de páncreas. De los 50 genes analizados encontramos únicamente 7 descritos en esta patología, 5 corresponden a genes sobreexpresados en hombres y 2 en mujeres.

Los genes *GRM-1* y *GRIK1* están expresados diferencialmente en mujeres. Ambos codifican para un receptor de glutamato (mGluR). El glutamato es un neurotransmisor excitador del sistema nervioso central. Hay subtipos de mGluR que se expresan típicamente de forma selectiva en tejidos de diferentes tipos de cáncer, incluido el de páncreas, lo que indica la participación de este receptor en la tumorigénesis, aunque todavía no está estudiada su participación en PDAC<sup>70</sup>.

En hombres encontramos el gen *IGH* que codifica para una cadena pesada de inmunoglobulinas, lo que tiene relación con el aumento de las funciones relacionadas con el sistema inmune antes descritas. El gen *KHDRBS2* codifica para una proteína de unión al ARN. Mientras que el gen *GSTM3* codifica para una glutatión-s-transferasa, una enzima que actúa en la desintoxicación de compuestos electrofílicos, incluidos algunos carcinógenos, fármacos terapéuticos, toxinas ambientales y productos del estrés oxidativo, por conjugación con glutatión. *ACVR1C*, también conocido como *ALK7* codifica para un receptor de activina A tipo 1C que se une a TGF- $\beta$  y se ha relacionado con la carcinogénesis, metástasis, invasión y pronóstico de PDAC<sup>71</sup>. *ERP29* codifica una proteína que se localiza en la luz del retículo endoplásmico y ha sido relacionada con la sensibilidad a gemcitabine<sup>72</sup>.

El resto de los genes encontrados en este estudio con expresión diferencial tanto para hombres como para mujeres no relacionados con cáncer de páncreas representan una novedad. Para determinar las funciones implicadas en estos genes realizamos un análisis con PANTHER donde encontramos que las funciones de regulación de funciones biológicas, procesos celulares, localización de elementos celulares y reproducción están más representadas en mujeres y las relacionadas con procesos metabólicos, respuesta a estímulos y procesos del sistema inmune en hombres.

En nuestro estudio hemos determinado que existen diferencias a nivel transcriptómico para ambos sexos en cáncer de páncreas, especialmente a nivel del sistema inmune y del microambiente tumoral. Dada la alta mortalidad que presenta esta enfermedad, no solo por

sus características malignas propias, sino también porque carecemos actualmente de métodos de diagnóstico precoz y tratamientos efectivos, será necesario desarrollar nuevos estudios con el objetivo de caracterizar las diferencias encontradas y poder utilizarlas en beneficio del desarrollo de nuevas técnicas diagnósticas y terapéuticas. La falta de información sobre el sexo en los estudios dificulta este tipo de estudios, por eso es imprescindible que se incluya esta información para poder avanzar sobre el tratamiento personalizado del cáncer de páncreas.

## 7 Conclusiones

1. El metaanálisis es un método robusto para la evaluación e integración de datos provenientes de una revisión sistemática de estudios transcriptómicos, potenciando la reutilización de datos de otros estudios y cumpliendo con los criterios FAIR.
2. La revisión sistemática y selección de estudios en cáncer de páncreas confirma la insuficiente información del sexo en los estudios biomédicos y su necesaria inclusión para garantizar resultados de calidad que puedan ser útiles al conjunto de la población.
3. En mujeres hemos detectado funciones relacionadas con el desarrollo de células y componentes del microambiente tumoral y con el desarrollo de enfermedades inflamatorias.
4. Las funciones más representativas en hombres son las relacionadas con el sistema inmune y con las alteraciones a nivel del ciclo celular.
5. Los genes y funciones identificados pueden ser útiles para guiar nuevos estudios y estrategias terapéuticas.
6. La estrategia propuesta en este trabajo permite detectar y caracterizar diferencias debidas al sexo en cáncer de páncreas, incorporando la necesaria perspectiva de la medicina personalizada y de precisión en los estudios biomédicos.



## 8 Bibliografía

1. Macarulla, T. Digestivo - Páncreas - SEOM: Sociedad Española de Oncología Médica © 2019. [https://seom.org/131-Información al Público - Patologías/Digestivo - Páncreas](https://seom.org/131-Información%20al%20Público%20-%20Patologías/Digestivo%20-%20Páncreas) (2019).
2. Asociación Española Contra el Cáncer. Anatomía del Páncreas | AECC. *Cáncer de páncreas* <https://www.aecc.es/es/todo-sobre-cancer/tipos-cancer/cancer-pancreas/anatomia>.
3. World Health Organization. Cancer Today. *Estimated age-standardized incidence rates (World) in 2020, pancreas, all ages* <https://gco.iarc.fr>.
4. McGuigan, A. *et al.* Pancreatic cancer: A review of clinical diagnosis, epidemiology, treatment and outcomes. *World Journal of Gastroenterology* vol. 24 (2018).
5. Kuper, H., Boffetta, P. & Adami, H. O. Tobacco use and cancer causation: Association by tumour type. *Journal of Internal Medicine* vol. 252 (2002).
6. Wang, M., Gorelick, F. & Bhargava, A. Sex Differences in the Exocrine Pancreas and Associated Diseases. *Cell. Mol. Gastroenterol. Hepatol.* (2021) doi:10.1016/j.jcmgh.2021.04.005.
7. Midha, S., Chawla, S. & Garg, P. K. Modifiable and non-modifiable risk factors for pancreatic cancer: A review. *Cancer Letters* vol. 381 (2016).
8. Becker, A. E., Hernandez, Y. G., Frucht, H. & Lucas, A. L. Pancreatic ductal adenocarcinoma: Risk factors, screening, and early detection. *World Journal of Gastroenterology* vol. 20 (2014).
9. Ducreux, M. *et al.* Cancer of the pancreas: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann. Oncol.* **26**, (2015).
10. Barone, E., Corrado, A., Gemignani, F. & Landi, S. Environmental risk factors for pancreatic cancer: an update. *Archives of Toxicology* vol. 90 (2016).
11. Zhao, Z. Y. & Liu, W. Pancreatic Cancer: A Review of Risk Factors, Diagnosis, and Treatment. *Technol. Cancer Res. Treat.* **19**, (2020).
12. Gordon-Dseagu, V. L., Devesa, S. S., Goggins, M. & Stolzenberg-Solomon, R. Pancreatic cancer incidence trends: Evidence from the Surveillance, Epidemiology and End Results (SEER) population-based data. *Int. J. Epidemiol.* **47**, 427–439 (2018).
13. Sadr-Azodi, O., Konings, P. & Brusselaers, N. Menopausal hormone therapy and pancreatic cancer risk in women: a population-based matched cohort study. *United Eur. Gastroenterol. J.* **5**, (2017).
14. Hidalgo, M. Pancreatic Cancer. *N. Engl. J. Med.* **362**, 1605–1617 (2010).
15. Vincent, A., Herman, J., Schulick, R., Hruban, R. H. & Goggins, M. Pancreatic cancer. in *The Lancet* vol. 378 607–620 (Elsevier B.V., 2011).
16. Prieto-García, E. *et al.* Tumor–stromal interactions in a co-culture model of human

pancreatic adenocarcinoma cells and fibroblasts and their connection with tumor spread. *Biomedicines* **9**, (2021).

17. Murakami, T. *et al.* Role of the tumor microenvironment in pancreatic cancer. *Annals of Gastroenterological Surgery* vol. 3 (2019).
18. Chang, J. H., Jiang, Y. & Pillarisetty, V. G. Role of immune cells in pancreatic cancer from bench to clinical application An updated review. *Medicine (United States)* vol. 95 (2016).
19. Amedei, A., Niccolai, E. & Prisco, D. Pancreatic cancer: Role of the immune system in cancer progression and vaccine-based immunotherapy. *Human Vaccines and Immunotherapeutics* vol. 10 (2014).
20. Pancreatic Cancer: Introduction | Cancer.Net. <https://www.cancer.net/cancer-types/pancreatic-cancer/introduction>.
21. Pereira, S. P. *et al.* Early detection of pancreatic cancer. *The Lancet Gastroenterology and Hepatology* vol. 5 698–710 (2020).
22. Dhanda, J. S. & Dhanda, J. S. *Structural Levels of Nucleic Acids and Sequencing*. (2008).
23. Boyle, J. Lehninger principles of biochemistry (4th ed.): Nelson, D., and Cox, M. *Biochem. Mol. Biol. Educ.* **33**, (2005).
24. Yadav, S. P. The wholeness in suffix -omics, -omes, and the word om. *Journal of Biomolecular Techniques* vol. 18 (2007).
25. Orfao, A. *et al.* *Informes Anticipando Ciencias ómicas*. [www.institutoroche.es](http://www.institutoroche.es).
26. Manzoni, C. *et al.* Genome, transcriptome and proteome: The rise of omics data and their integration in biomedical sciences. *Brief. Bioinform.* **19**, (2018).
27. Dong, Z. C. & Chen, Y. Transcriptomics: Advances and approaches. *Science China Life Sciences* vol. 56 (2013).
28. Doménech-Sáncheza, A. & Vilab, J. Fundamento, tipos y aplicaciones de los arrays de ADN en la microbiología médica. *Enferm. Infecc. Microbiol. Clin.* **22**, (2004).
29. Wang, Z., Gerstein, M. & Snyder, M. RNA-Seq: A revolutionary tool for transcriptomics. *Nature Reviews Genetics* vol. 10 (2009).
30. García García, F. Métodos de análisis de enriquecimiento funcional en estudios genómicos. (2016).
31. Grant, G. R., Manduchi, E. & Stoeckert, C. J. Analysis and management of microarray gene expression data. *Current protocols in molecular biology / edited by Frederick M. Ausubel ... [et al.]* vol. Chapter 19 (2007).
32. Haidich, A. B. (2010). Meta-analysis in medical research. *Hippokratia* **14**, (2010).
33. Normand, S. T. Meta-analysis: formulating, evaluating, combining, and reporting. *Stat. Med.* **18**, (1999).
34. Liberati, A. *et al.* The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and

- elaboration. *PLoS Medicine* vol. 6 (2009).
35. McDermaid, A., Monier, B., Zhao, J., Liu, B. & Ma, Q. Interpretation of differential gene expression results of RNA-seq data: Review and integration. *Briefings in Bioinformatics* vol. 20 (2019).
  36. Ashburner, M. *et al.* Gene ontology: Tool for the unification of biology. *Nature Genetics* vol. 25 (2000).
  37. Kanehisa, M. & Goto, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research* vol. 28 (2000).
  38. Subramanian, A. *et al.* Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* **102**, (2005).
  39. R Core Team. R: A language and environment for statistical computing. *R Foundation for Statistical Computing* (2019).
  40. Barrett, T. *et al.* NCBI GEO: Archive for functional genomics data sets - Update. *Nucleic Acids Res.* **41**, (2013).
  41. Kolesnikov, N. *et al.* ArrayExpress update-simplifying data submissions. *Nucleic Acids Res.* **43**, (2015).
  42. Sean, D. & Meltzer, P. S. GEOquery: A bridge between the Gene Expression Omnibus (GEO) and BioConductor. *Bioinformatics* **23**, (2007).
  43. Maglott, D., Ostell, J., Pruitt, K. D. & Tatusova, T. Entrez gene: Gene-centered information at NCBI. *Nucleic Acids Res.* **39**, (2011).
  44. Ritchie, M. E. *et al.* Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, (2015).
  45. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B* **57**, (1995).
  46. Viechtbauer, W. Conducting Meta-analysis in R with the metafor package. *J. Stat. Softw.* **36**, (2010).
  47. DerSimonian, R. & Laird, N. Meta-analysis in clinical trials revisited. *Contemp. Clin. Trials* **45**, (2015).
  48. Mi, H. *et al.* PANTHER version 16: A revised family classification, tree-based classification tool, enhancer regions and extensive API. *Nucleic Acids Res.* **49**, (2021).
  49. Montaner, D. & Dopazo, J. Multidimensional gene set analysis of genomic data. *PLoS One* **5**, (2010).
  50. Li, L. *et al.* Genetic variations associated with gemcitabine treatment outcome in pancreatic cancer. *Pharmacogenet. Genomics* **26**, 527–537 (2016).
  51. Ellsworth, K. A. *et al.* Contribution of FKBP5 Genetic Variation to Gemcitabine Treatment and Survival in Pancreatic Adenocarcinoma. *PLoS One* **8**, (2013).
  52. Pei, H. *et al.* FKBP51 Affects Cancer Cell Response to Chemotherapy by

- Negatively Regulating Akt. *Cancer Cell* **16**, 259–266 (2009).
53. DiPardo, B. J., Winograd, P., Court, C. M. & Tomlinson, J. S. Pancreatic cancer circulating tumor cells: applications for personalized oncology. *Expert Review of Molecular Diagnostics* vol. 18 (2018).
  54. Irigoyen, A. *et al.* Integrative multi-platform meta-analysis of gene expression profiles in pancreatic ductal adenocarcinoma patients for identifying novel diagnostic biomarkers. *PLoS One* **13**, (2018).
  55. Caba, O. *et al.* Transcriptional Profiling of Peripheral Blood in Pancreatic Adenocarcinoma Patients Identifies Diagnostic Biomarkers. *Dig. Dis. Sci.* **59**, (2014).
  56. Abdollahi, A. *et al.* Transcriptional network governing the angiogenic switch in human pancreatic cancer. *Proc. Natl. Acad. Sci. U. S. A.* **104**, (2007).
  57. Lunardi, S. *et al.* IP-10/CXCL10 induction in human pancreatic cancer stroma influences lymphocytes recruitment and correlates with poor survival. *Oncotarget* **5**, (2014).
  58. Voskuhl, R. R., Sawalha, A. H. & Itoh, Y. Sex chromosome contributions to sex differences in multiple sclerosis susceptibility and progression. *Mult. Scler.* **24**, 22–31 (2018).
  59. Wilkinson, M. D. *et al.* Comment: The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* **3**, (2016).
  60. Goonesekere, N. C. W., Wang, X., Ludwig, L. & Guda, C. A meta analysis of pancreatic microarray datasets yields new targets as cancer genes and biomarkers. *PLoS One* **9**, (2014).
  61. Saloman, J. L. *et al.* Ablation of sensory neurons in a genetic model of pancreatic ductal adenocarcinoma slows initiation and progression of cancer. *Proc. Natl. Acad. Sci. U. S. A.* **113**, (2016).
  62. Zang, X. P., Lerner, M., Brackett, D. & Pento, J. T. Influence of KGF on the progression of pancreatic cancer. *Anticancer Res.* **29**, (2009).
  63. Ishiwata, T., Friess, H., Büchler, M. W., Lopez, M. E. & Korc, M. Characterization of keratinocyte growth factor and receptor expression in human pancreatic cancer. *Am. J. Pathol.* **153**, (1998).
  64. Grapa, C. M. *et al.* Epidermal growth factor receptor and its role in pancreatic cancer treatment mediated by nanoparticles. *International Journal of Nanomedicine* vol. 14 (2019).
  65. Banh, R. S. *et al.* Neurons Release Serine to Support mRNA Translation in Pancreatic Cancer. *Cell* **183**, (2020).
  66. Zhu, X. *et al.* MyD88 signalling is critical in the development of pancreatic cancer cachexia. *J. Cachexia. Sarcopenia Muscle* **10**, (2019).
  67. Schneider, G. & Schmid, R. M. Genetic alterations in pancreatic carcinoma. *Molecular Cancer* vol. 2 (2003).

68. Perkhofer, L. *et al.* DNA damage repair as a target in pancreatic cancer: State-of-the-art and future perspectives. *Gut* vol. 70 (2021).
69. Ochoa, D. *et al.* Open Targets Platform: Supporting systematic drug-target identification and prioritisation. *Nucleic Acids Res.* **49**, (2021).
70. Wang, X., Xiao, L. & Yu, H. Expression levels of long non-coding RNA HOXA distal transcript antisense RNA and metabotropic glutamate receptor 1 in pancreatic carcinoma, and their prognostic values. *Oncol. Lett.* **15**, (2018).
71. Liu, C. *et al.* Overexpression of B2M and loss of ALK7 expression are associated with invasion, metastasis, and poor-prognosis of the pancreatic ductal adenocarcinoma. *Cancer Biomarkers* **15**, (2015).
72. Mori-Iwamoto, S. *et al.* Proteomics finding heat shock protein 27 as a biomarker for resistance of pancreatic cancer cells to gemcitabine. *Int. J. Oncol.* **31**, (2007).



# Anexo I - Figuras

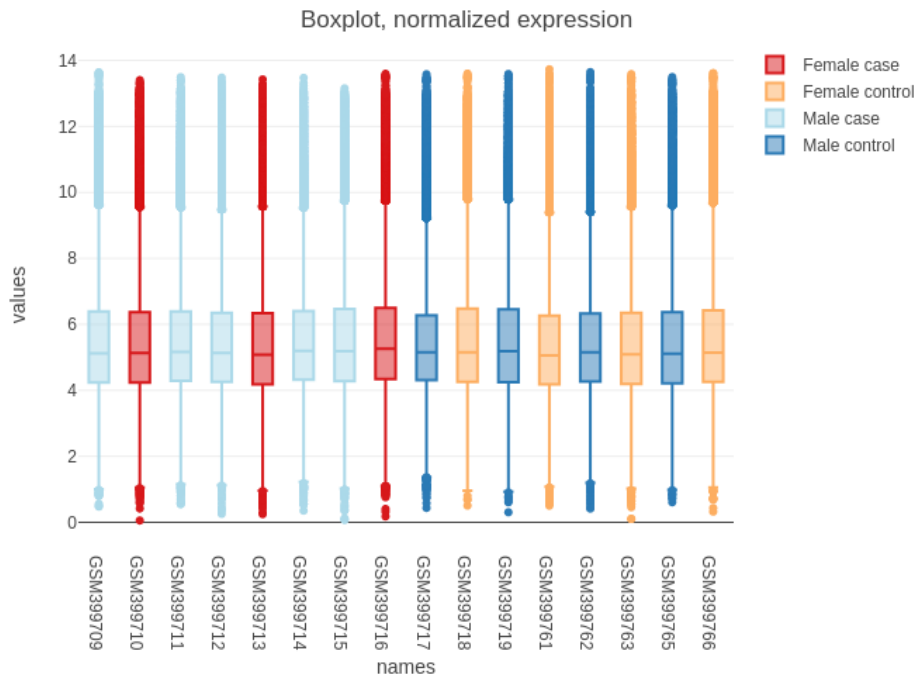


Figura A 1: Diagrama de cajas del estudio GSE15932.

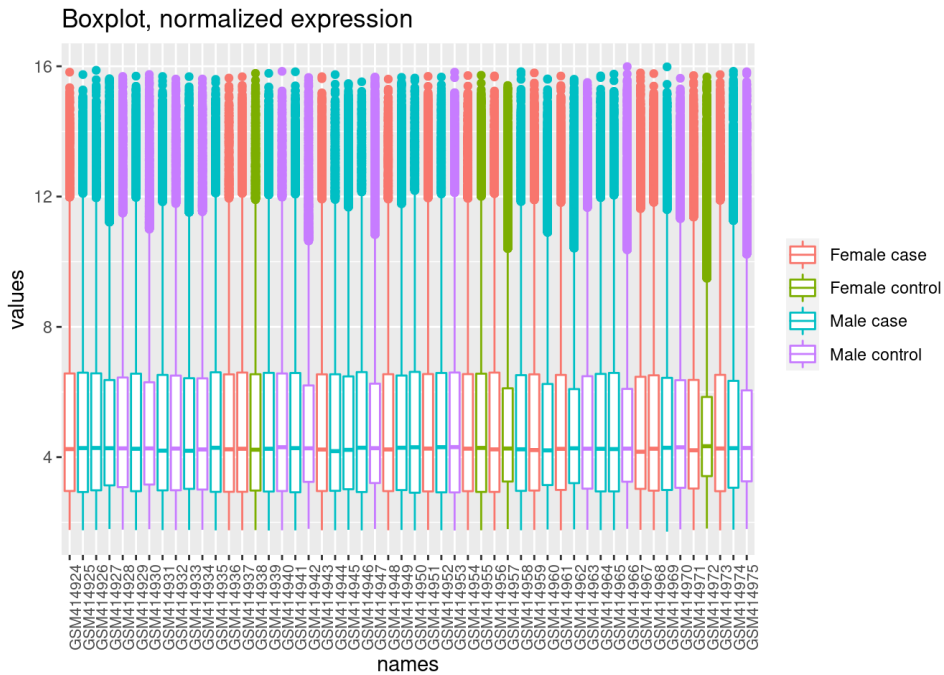


Figura A 2: Diagrama de cajas del estudio GSE16515.

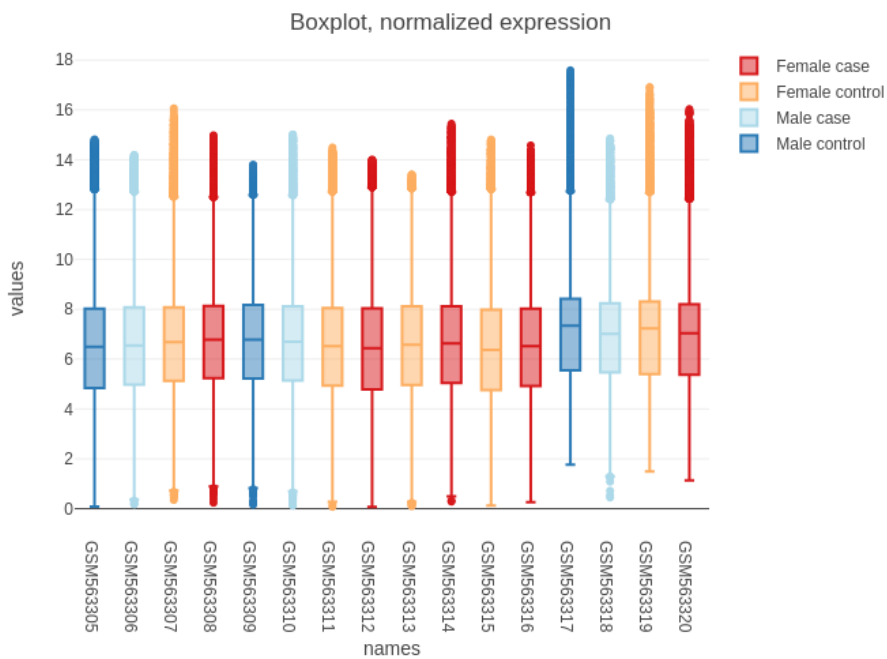


Figura A 3: Diagrama de cajas del estudio GSE22780.

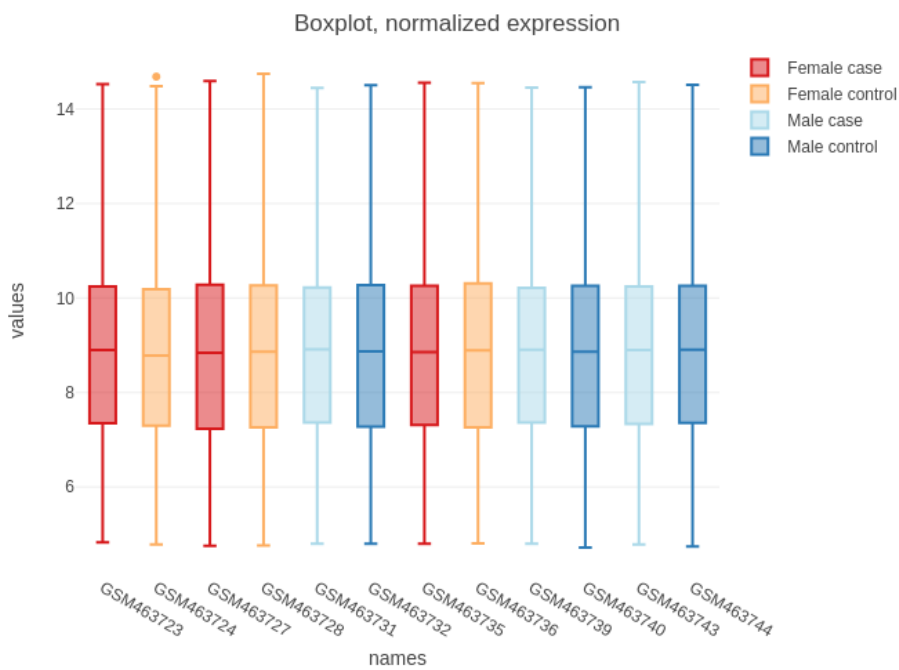


Figura A 4: Diagrama de cajas del estudio GSE18670.

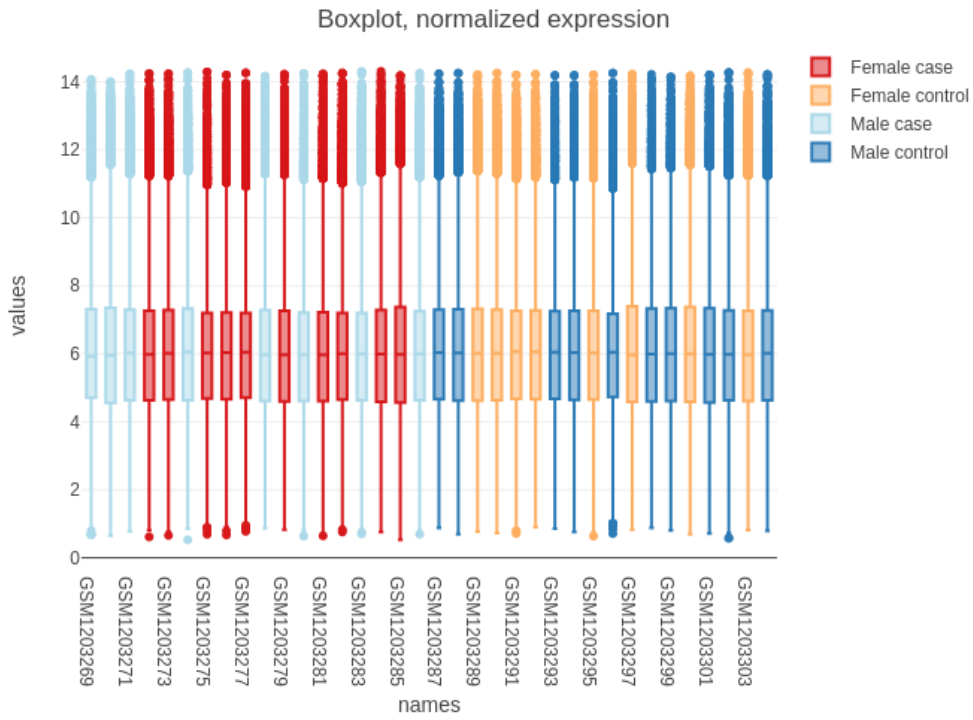


Figura A 5: Diagrama de cajas del estudio GSE49641.

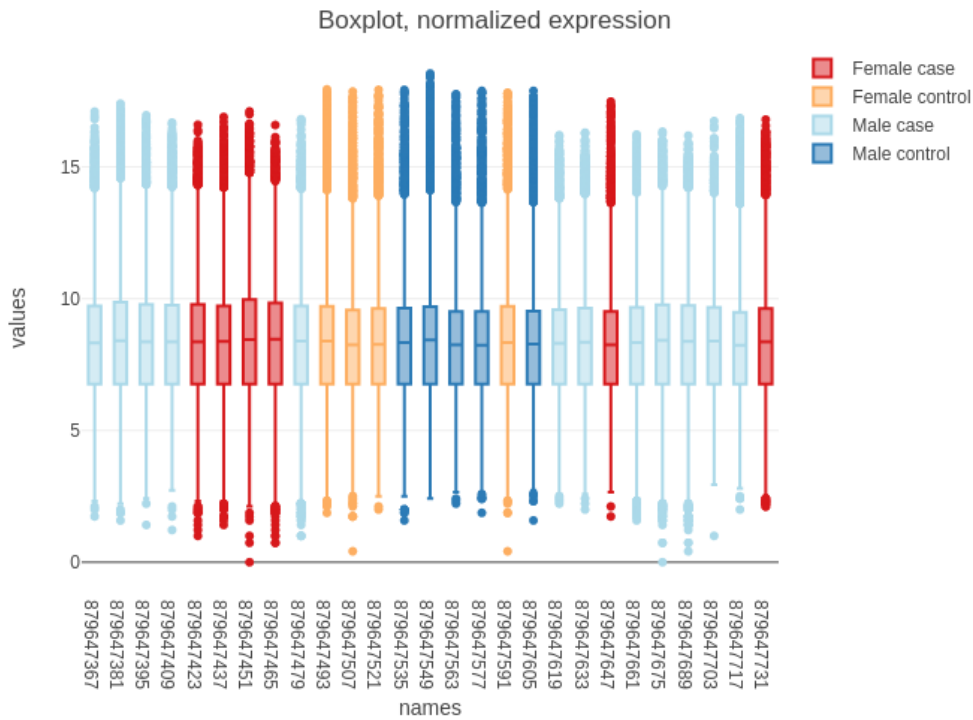


Figura A 6: Diagrama de cajas del estudio E-EMBL-6.

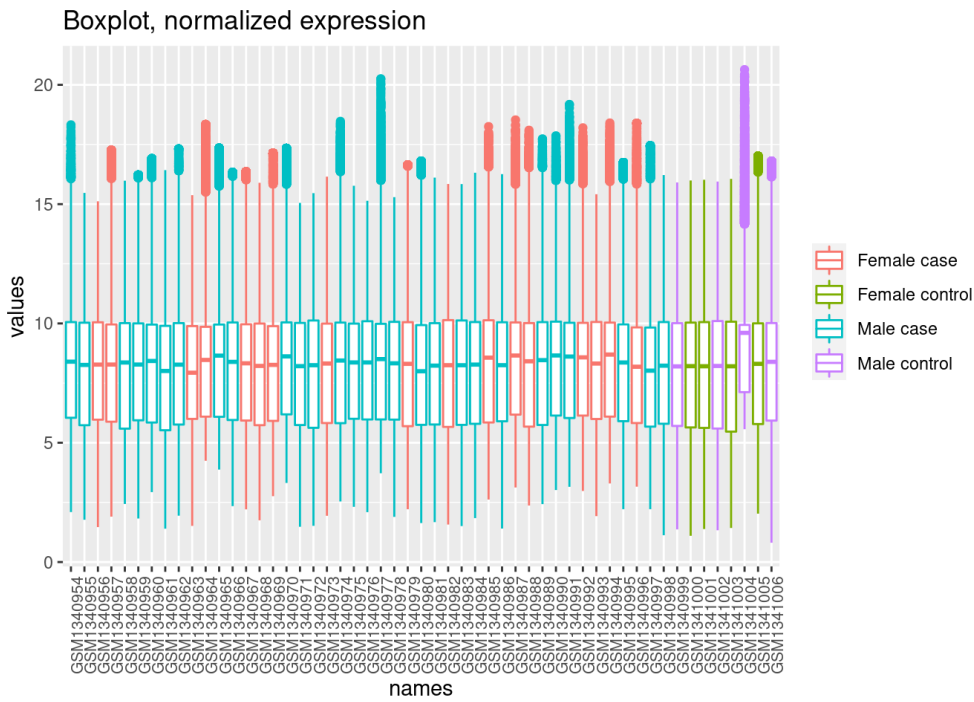


Figura A 7: Diagrama de cajas del estudio GSE55643.

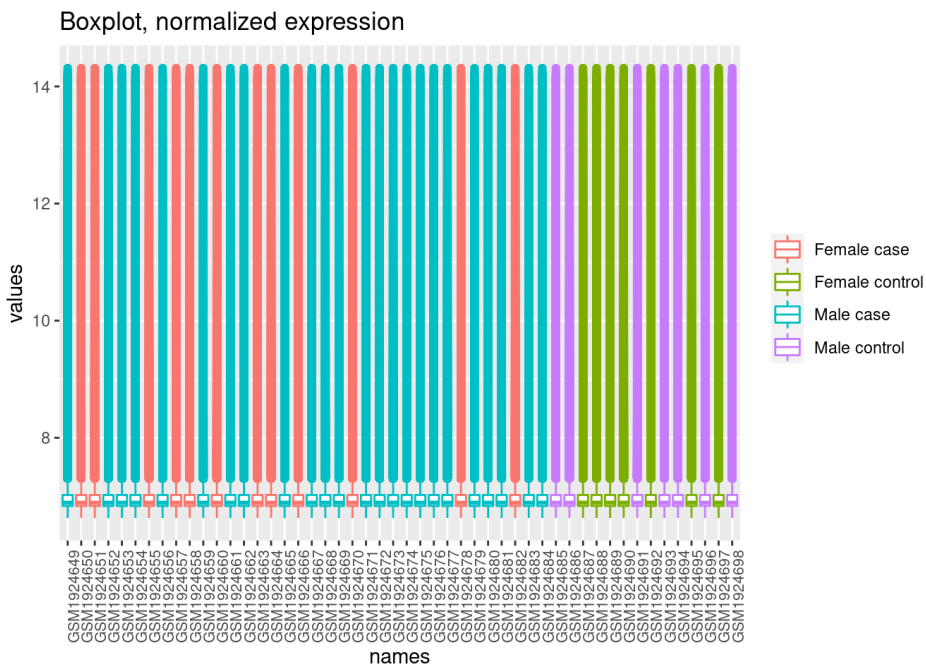


Figura A 8: Diagrama de cajas del estudio GSE74629.

## Anexo II - Tablas

| ID         | lor         | pval       | padj       | GO   |
|------------|-------------|------------|------------|--|
| GO:0050907 | 1,43356354  | 2,2601E-09 | 2,9376E-06 | detection of chemical stimulus involved in sensory perception          |
| GO:0050911 | 1,47250346  | 5,148E-09  | 5,5761E-06 | detection of chemical stimulus involved in sensory perception of smell |
| GO:0007608 | 1,38639219  | 2,4561E-08 | 2,2803E-05 | sensory perception of smell  |
| GO:0031424 | 1,45789069  | 7,9769E-08 | 5,7602E-05 | keratinization   |
| GO:0007586 | 1,34202321  | 2,8088E-05 | 0,00325968 | digestion  |
| GO:0009913 | 1,02054229  | 5,4594E-05 | 0,0054517  | epidermal cell differentiation   |
| GO:0030216 | 1,07248925  | 7,1889E-05 | 0,00614747 | keratinocyte differentiation   |
| GO:0007210 | 1,71426707  | 0,00015648 | 0,01168928 | serotonin receptor signaling pathway                                   |
| GO:0015816 | 1,95325166  | 0,00023272 | 0,01559194 | glycine transport  |
| GO:0006805 | 1,2853615   | 0,00036822 | 0,02155927 | xenobiotic metabolic process   |
| GO:0071466 | 1,26731278  | 0,00038613 | 0,02220779 | cellular response to xenobiotic stimulus                               |
| GO:0007218 | 1,29698407  | 0,00043433 | 0,02412599 | neuropeptide signaling pathway   |
| GO:0009410 | 1,20013681  | 0,0008412  | 0,03904966 | response to xenobiotic stimulus  |
| GO:0015808 | 1,94786853  | 0,00100976 | 0,04404315 | L-alanine transport  |
| GO:0008544 | 0,78360744  | 0,00106972 | 0,04573777 | epidermis development  |
| GO:0043588 | 0,81416806  | 0,00118284 | 0,04927754 | skin development   |
| GO:0009266 | -1,29694018 | 0,00119546 | 0,04948606 | response to temperature stimulus                                       |
| GO:0010332 | -2,17401912 | 0,00117914 | 0,04927754 | response to gamma radiation  |
| GO:0006296 | -2,47355293 | 0,00108687 | 0,0458672  | nucleotide-excision repair, DNA incision, 5'-to lesion                 |
| GO:0016078 | -3,15647773 | 0,00108107 | 0,0458672  | tRNA catabolic process   |
| GO:1903706 | -0,93509291 | 0,00104629 | 0,04503218 | regulation of hemopoiesis  |
| GO:0002440 | -1,33515409 | 0,00102018 | 0,04420088 | production of molecular mediator of immune response                    |
| GO:0051052 | -1,0822099  | 0,00100233 | 0,04401465 | regulation of DNA metabolic process                                    |

|            |             |            |            |   |
|------------|-------------|------------|------------|---|
| GO:0044843 | -1,2249149  | 0,00097171 | 0,04296029 | cell cycle G1/S phase transition  |
| GO:0050863 | -1,13358958 | 0,00095427 | 0,04247827 | regulation of T cell activation   |
| GO:0002183 | -2,52806481 | 0,00094305 | 0,04226804 | cytoplasmic translational initiation  |
| GO:0010212 | -1,59866223 | 0,00093037 | 0,04226804 | response to ionizing radiation  |
| GO:0032729 | -2,12460416 | 0,00093674 | 0,04226804 | positive regulation of interferon-gamma production                            |
| GO:0045930 | -1,15414145 | 0,00094289 | 0,04226804 | negative regulation of mitotic cell cycle                                     |
| GO:0002706 | -1,56859468 | 0,00089473 | 0,04124012 | regulation of lymphocyte mediated immunity                                    |
| GO:0002381 | -2,17117114 | 0,00081398 | 0,0380978  | immunoglobulin production involved in immunoglobulin mediated immune response |
| GO:0009451 | -1,5222987  | 0,00081483 | 0,0380978  | RNA modification  |
| GO:0034341 | -1,42918084 | 0,0007662  | 0,03634699 | response to interferon-gamma  |
| GO:0031343 | -2,18563814 | 0,00073409 | 0,03507985 | positive regulation of cell killing   |
| GO:1901988 | -1,3163639  | 0,00072552 | 0,034927   | negative regulation of cell cycle phase transition                            |
| GO:0034154 | -3,24911534 | 0,00070353 | 0,03412138 | toll-like receptor 7 signaling pathway  |
| GO:0045876 | -3,28654691 | 0,00070084 | 0,03412138 | positive regulation of sister chromatid cohesion                              |
| GO:0006911 | -2,16871666 | 0,0006883  | 0,03395407 | phagocytosis, engulfment  |
| GO:1903037 | -1,18206343 | 0,00068963 | 0,03395407 | regulation of leukocyte cell-cell adhesion                                    |
| GO:0051169 | -1,1284334  | 0,00066669 | 0,03332942 | nuclear transport   |
| GO:0009615 | -1,14621981 | 0,00065995 | 0,03324813 | response to virus   |
| GO:0006406 | -1,79499081 | 0,00065341 | 0,03317604 | mRNA export from nucleus  |
| GO:0071427 | -1,79499081 | 0,00065341 | 0,03317604 | mRNA-containing ribonucleoprotein complex export from nucleus                 |
| GO:0006913 | -1,13744326 | 0,00063116 | 0,03255478 | nucleocytoplasmic transport   |
| GO:1901991 | -1,37345986 | 0,00061606 | 0,03203006 | negative regulation of mitotic cell cycle phase transition                    |

|            |             |            |            |  |
|------------|-------------|------------|------------|--|
| GO:0050870 | -1,42110607 | 0,00055454 | 0,02906433 | positive regulation of T cell activation   |
| GO:0050657 | -1,46677126 | 0,00054637 | 0,02894011 | nucleic acid transport   |
| GO:0050658 | -1,46677126 | 0,00054637 | 0,02894011 | RNA transport  |
| GO:0051236 | -1,45741228 | 0,00054772 | 0,02894011 | establishment of RNA localization  |
| GO:0002824 | -1,87631307 | 0,0005187  | 0,02809195 | positive regulation of adaptive immune response based on somatic recombination of immune receptors built from immunoglobulin superfamily domains |
| GO:0000070 | -1,59656288 | 0,0004939  | 0,02697359 | mitotic sister chromatid segregation   |
| GO:0033260 | -2,25049037 | 0,00044307 | 0,0244028  | nuclear DNA replication  |
| GO:0002717 | -2,72491358 | 0,00041731 | 0,02375894 | positive regulation of natural killer cell mediated immunity   |
| GO:0002753 | -2,16768071 | 0,00042247 | 0,02375894 | cytoplasmic pattern recognition receptor signaling pathway   |
| GO:0042254 | -1,23461972 | 0,00042407 | 0,02375894 | ribosome biogenesis  |
| GO:0042110 | -1,02280956 | 0,00037638 | 0,02184018 | T cell activation  |
| GO:0045954 | -2,86595759 | 0,00034632 | 0,02046101 | positive regulation of natural killer cell mediated cytotoxicity   |
| GO:0000082 | -1,36163145 | 0,00034027 | 0,02028806 | G1/S transition of mitotic cell cycle  |
| GO:0000731 | -2,36994362 | 0,00032996 | 0,01985559 | DNA synthesis involved in DNA repair   |
| GO:0031124 | -1,93567111 | 0,00032288 | 0,01966613 | mRNA 3'-end processing   |
| GO:0032201 | -2,81379073 | 0,00032378 | 0,01966613 | telomere maintenance via semi-conservative replication   |
| GO:0045010 | -2,37412702 | 0,00031537 | 0,01952008 | actin nucleation   |
| GO:0140053 | -1,62162537 | 0,00031301 | 0,01952008 | mitochondrial gene expression  |
| GO:0042769 | -2,58904574 | 0,0003054  | 0,0192697  | DNA damage response, detection of DNA damage   |
| GO:0002697 | -1,11162298 | 0,00029097 | 0,01865443 | regulation of immune effector process  |
| GO:0009408 | -1,63319435 | 0,00029278 | 0,01865443 | response to heat   |
| GO:0006301 | -2,40202118 | 0,00027665 | 0,01797972 | postreplication repair   |
| GO:0098813 | -1,35425776 | 0,00026258 | 0,01723714 | nuclear chromosome segregation   |

|            |             |            |            |  |
|------------|-------------|------------|------------|--|
| GO:1900034 | -2,11611926 | 0,00024418 | 0,01619315 | regulation of cellular response to heat                |
| GO:0043488 | -1,58644115 | 0,00021633 | 0,01464488 | regulation of mRNA stability                           |
| GO:0070987 | -2,99358675 | 0,00021228 | 0,01452198 | error-free translesion synthesis                       |
| GO:0006405 | -1,75580412 | 0,00020785 | 0,01437039 | RNA export from nucleus                                |
| GO:0071346 | -1,61857623 | 0,00020775 | 0,01437039 | cellular response to interferon-gamma                  |
| GO:0043487 | -1,56473524 | 0,00019841 | 0,01402038 | regulation of RNA stability                            |
| GO:0044786 | -2,27909928 | 0,00019847 | 0,01402038 | cell cycle DNA replication                             |
| GO:0006403 | -1,46509576 | 0,00017926 | 0,01294476 | RNA localization                                       |
| GO:0051168 | -1,5549241  | 0,00017176 | 0,01268053 | nuclear export   |
| GO:0071103 | -1,25010079 | 0,00017365 | 0,01268053 | DNA conformation change                                |
| GO:0000819 | -1,60947506 | 0,00012575 | 0,00957919 | sister chromatid segregation                           |
| GO:0061013 | -1,56243539 | 0,00012676 | 0,00957919 | regulation of mRNA catabolic process                   |
| GO:0002708 | -2,00331401 | 0,00012092 | 0,00935556 | positive regulation of lymphocyte mediated immunity    |
| GO:0002755 | -2,68800828 | 0,00011618 | 0,00909714 | MyD88-dependent toll-like receptor signaling pathway   |
| GO:0019985 | -2,63753523 | 0,00010477 | 0,0083797  | translesion synthesis                                  |
| GO:0071166 | -1,87177665 | 0,00010573 | 0,0083797  | ribonucleoprotein complex localization                 |
| GO:0006446 | -2,211819   | 9,8785E-05 | 0,00802506 | regulation of translational initiation                 |
| GO:0071426 | -1,88472067 | 9,6365E-05 | 0,00792757 | ribonucleoprotein complex export from nucleus          |
| GO:0006417 | -1,26252483 | 8,8569E-05 | 0,00737959 | regulation of translation                              |
| GO:0001912 | -2,50236884 | 8,728E-05  | 0,00736663 | positive regulation of leukocyte mediated cytotoxicity |
| GO:0006260 | -1,44726871 | 7,0682E-05 | 0,00612487 | DNA replication  |
| GO:0006910 | -2,99985931 | 7,0506E-05 | 0,00612487 | phagocytosis, recognition                              |
| GO:0010389 | -1,65191431 | 6,4409E-05 | 0,00573412 | regulation of G2/M transition of mitotic cell cycle    |
| GO:0000086 | -1,50387799 | 6,2358E-05 | 0,00564255 | G2/M transition of mitotic cell cycle                  |
| GO:0006611 | -1,70122183 | 6,1149E-05 | 0,00564255 | protein export from nucleus                            |

|            |             |            |            |   |
|------------|-------------|------------|------------|---|
| GO:0045088 | -1,40370235 | 6,0025E-05 | 0,00564255 | regulation of innate immune response  |
| GO:0050871 | -2,18405421 | 6,112E-05  | 0,00564255 | positive regulation of B cell activation  |
| GO:1902749 | -1,60191476 | 6,2512E-05 | 0,00564255 | regulation of cell cycle G2/M phase transition  |
| GO:0007059 | -1,35272542 | 5,7343E-05 | 0,00556225 | chromosome segregation  |
| GO:0006302 | -1,49489402 | 5,5364E-05 | 0,0054517  | double-strand break repair  |
| GO:0050867 | -1,31895894 | 5,3563E-05 | 0,00543918 | positive regulation of cell activation  |
| GO:0044839 | -1,47688255 | 4,9269E-05 | 0,00508257 | cell cycle G2/M phase transition  |
| GO:0034605 | -1,96823436 | 4,757E-05  | 0,00498644 | cellular response to heat   |
| GO:0002495 | -2,1280108  | 3,5965E-05 | 0,00383175 | antigen processing and presentation of peptide antigen via MHC class II                   |
| GO:0000377 | -1,33715841 | 3,3457E-05 | 0,00362397 | RNA splicing, via transesterification reactions with bulged adenosine as nucleophile      |
| GO:0000398 | -1,33715841 | 3,3457E-05 | 0,00362397 | mRNA splicing, via spliceosome  |
| GO:0090305 | -1,43260388 | 3,2764E-05 | 0,00362397 | nucleic acid phosphodiester bond hydrolysis   |
| GO:0002504 | -2,13205299 | 3,1367E-05 | 0,00357634 | antigen processing and presentation of peptide or polysaccharide antigen via MHC class II |
| GO:0031123 | -1,912174   | 2,6345E-05 | 0,003113   | RNA 3'-end processing   |
| GO:0045089 | -1,68238284 | 2,6192E-05 | 0,003113   | positive regulation of innate immune response   |
| GO:0070424 | -3,54438712 | 2,6086E-05 | 0,003113   | regulation of nucleotide-binding oligomerization domain containing signaling pathway      |
| GO:0000375 | -1,35255033 | 2,4527E-05 | 0,00306544 | RNA splicing, via transesterification reactions   |
| GO:0001510 | -2,28459409 | 2,272E-05  | 0,0029163  | RNA methylation   |
| GO:0002431 | -2,22843579 | 2,1473E-05 | 0,0029163  | Fc receptor mediated stimulatory signaling pathway  |
| GO:0002696 | -1,39949536 | 2,2376E-05 | 0,0029163  | positive regulation of leukocyte activation   |
| GO:0006399 | -1,76411825 | 2,2885E-05 | 0,0029163  | tRNA metabolic process  |

|            |             |            |            |   |
|------------|-------------|------------|------------|---|
| GO:0050864 | -1,99492267 | 2,1805E-05 | 0,0029163  | regulation of B cell activation   |
| GO:0006261 | -1,91342917 | 2,0392E-05 | 0,00289892 | DNA-dependent DNA replication   |
| GO:0006310 | -1,50038163 | 1,8615E-05 | 0,00289892 | DNA recombination   |
| GO:0006413 | -1,76701888 | 1,8072E-05 | 0,00289892 | translational initiation  |
| GO:0022613 | -1,24476235 | 1,9769E-05 | 0,00289892 | ribonucleoprotein complex biogenesis  |
| GO:0038094 | -2,2551918  | 2,0442E-05 | 0,00289892 | Fc-gamma receptor signaling pathway   |
| GO:0048002 | -1,75466497 | 2,0519E-05 | 0,00289892 | antigen processing and presentation of peptide antigen                                      |
| GO:1903311 | -1,40255131 | 1,9152E-05 | 0,00289892 | regulation of mRNA metabolic process  |
| GO:0042276 | -3,22832661 | 1,5773E-05 | 0,00262841 | error-prone translesion synthesis   |
| GO:0038095 | -2,09504123 | 1,4437E-05 | 0,00246919 | Fc-epsilon receptor signaling pathway   |
| GO:0006402 | -1,37176658 | 1,3645E-05 | 0,00239679 | mRNA catabolic process  |
| GO:0034470 | -1,34414603 | 1,3358E-05 | 0,00239679 | ncRNA processing  |
| GO:0006401 | -1,325523   | 1,1576E-05 | 0,00214958 | RNA catabolic process   |
| GO:0006457 | -1,70751687 | 1,1093E-05 | 0,00212044 | protein folding   |
| GO:0032508 | -2,18768532 | 1,0252E-05 | 0,00201912 | DNA duplex unwinding  |
| GO:0002831 | -1,36031308 | 8,6423E-06 | 0,0017552  | regulation of response to biotic stimulus   |
| GO:0002433 | -2,35098813 | 7,4998E-06 | 0,00166697 | immune response-regulating cell surface receptor signaling pathway involved in phagocytosis |
| GO:0019886 | -2,26641586 | 7,9514E-06 | 0,00166697 | antigen processing and presentation of exogenous peptide antigen via MHC class II           |
| GO:0038096 | -2,35098813 | 7,4998E-06 | 0,00166697 | Fc-gamma receptor signaling pathway involved in phagocytosis                                |
| GO:0042113 | -1,61537383 | 7,8219E-06 | 0,00166697 | B cell activation   |
| GO:0002833 | -1,67588486 | 6,0419E-06 | 0,00145431 | positive regulation of response to biotic stimulus  |
| GO:0032392 | -2,18756293 | 5,0094E-06 | 0,00125217 | DNA geometric change  |
| GO:1901987 | -1,34375806 | 3,7375E-06 | 0,00097161 | regulation of cell cycle phase transition   |
| GO:1901990 | -1,42137813 | 2,1323E-06 | 0,00057742 | regulation of mitotic cell cycle phase transition   |

|            |             |            |            |   |
|------------|-------------|------------|------------|---|
| GO:0008380 | -1,37384586 | 1,8425E-06 | 0,00052064 | RNA splicing  |
| GO:0019882 | -1,79877185 | 1,7249E-06 | 0,00050957 | antigen processing and presentation   |
| GO:0002478 | -1,99206458 | 1,3548E-06 | 0,00041926 | antigen processing and presentation of exogenous peptide antigen  |
| GO:0002455 | -2,70552739 | 1,2587E-06 | 0,00041395 | humoral immune response mediated by circulating immunoglobulin  |
| GO:0002460 | -1,69999443 | 1,2739E-06 | 0,00041395 | adaptive immune response based on somatic recombination of immune receptors built from immunoglobulin superfamily domains |
| GO:0051251 | -1,6973855  | 6,5472E-07 | 0,00023639 | positive regulation of lymphocyte activation  |
| GO:0002377 | -2,2748146  | 4,7419E-07 | 0,00018128 | immunoglobulin production   |
| GO:0006958 | -2,94789545 | 3,5112E-07 | 0,00015011 | complement activation, classical pathway  |
| GO:0050853 | -2,6755304  | 3,6957E-07 | 0,00015011 | B cell receptor signaling pathway   |
| GO:0034660 | -1,43085157 | 2,5229E-07 | 0,00011711 | ncRNA metabolic process   |
| GO:0019884 | -2,06624691 | 2,1402E-07 | 0,00010699 | antigen processing and presentation of exogenous antigen  |
| GO:0002449 | -1,82318475 | 1,6569E-07 | 8,9736E-05 | lymphocyte mediated immunity  |
| GO:0019724 | -2,29165226 | 1,3129E-07 | 7,7571E-05 | B cell mediated immunity  |
| GO:0016064 | -2,31850231 | 1,0569E-07 | 6,8691E-05 | immunoglobulin mediated immune response   |
| GO:0038093 | -2,11156688 | 6,2341E-08 | 5,0644E-05 | Fc receptor signaling pathway   |
| GO:0050852 | -2,2820499  | 7,2201E-10 | 1,1731E-06 | T cell receptor signaling pathway   |
| GO:0002429 | -2,16414602 | 5,318E-14  | 1,1521E-10 | immune response-activating cell surface receptor signaling pathway  |
| GO:0002757 | -2,16414602 | 5,318E-14  | 1,1521E-10 | immune response-activating signal transduction  |

|            |             |           |            |   |
|------------|-------------|-----------|------------|---|
| GO:0050851 | -2,52933552 | 4,072E-15 | 2,6464E-11 | antigen receptor-mediated signaling pathway |
|------------|-------------|-----------|------------|---|

Tabla A 1: Términos GO BP significativos. Se muestra el logaritmo de Odds Ratio (lor), p-valor, p-valor ajustado por BH y el término GO asociado.

| ID    | Lor         | pval       | padj       | KEGG   |
|-------|-------------|------------|------------|--|
| 04740 | 1,29839605  | 4,3776E-07 | 3,1081E-05 | Olfactory transduction                       |
| 00982 | 1,55764222  | 6,8912E-05 | 0,00133438 | Drug metabolism - cytochrome P450            |
| 00980 | 1,31108422  | 0,00398586 | 0,04042804 | Metabolism of xenobiotics by cytochrome P450 |
| 05130 | -2,00455789 | 0,00431195 | 0,04174752 | Pathogenic Escherichia coli infection        |
| 03420 | -2,23388536 | 0,0027773  | 0,02957823 | Nucleotide excision repair                   |
| 03050 | -2,25324496 | 0,00230225 | 0,02580943 | Proteasome                                   |
| 05140 | -1,97797675 | 0,00191015 | 0,0226035  | Leishmaniasis                                |
| 05145 | -1,71116534 | 0,00062605 | 0,00784405 | Toxoplasmosis                                |
| 03430 | -2,85500043 | 0,00058401 | 0,00777459 | Mismatch repair                              |
| 05416 | -2,12725694 | 0,00056234 | 0,00777459 | Viral miocarditis                            |
| 05323 | -1,98924381 | 0,00046916 | 0,00713787 | Rheumatoid arthritis                         |
| 04141 | -1,61541797 | 0,0003475  | 0,00569361 | Protein processing in endoplasmic reticulum  |
| 04940 | -2,5441478  | 0,00025434 | 0,00451457 | Type I diabetes mellitus                     |
| 05310 | -2,93389424 | 4,5108E-05 | 0,00096081 | Asthma                                       |
| 05330 | -2,83366965 | 2,756E-05  | 0,00065224 | Allograft rejection                          |
| 03013 | -1,91392603 | 1,9009E-05 | 0,00055878 | RNA transport                                |
| 03030 | -2,85738493 | 2,0987E-05 | 0,00055878 | DNA replication                              |
| 05320 | -2,66783242 | 1,1374E-05 | 0,00040379 | Autoimmune thyroid disease                   |
| 05332 | -2,94973979 | 2,1047E-06 | 8,9659E-05 | Graft-versus-host disease                    |
| 03040 | -2,20703166 | 1,1665E-06 | 6,2117E-05 | Spliceosome                                  |
| 04612 | -2,71955638 | 8,7323E-08 | 9,2999E-06 | Antigen processing and presentation          |
| 04672 | -3,20534022 | 1,8528E-09 | 3,9465E-07 | Intestinal immune network for IgA production |

Tabla A 2: Rutas Kegg significativas. Se muestra el logaritmo de Odds Ratio (lor), p-valor, p-valor ajustado por BH y la ruta Kegg asociada.