



**Universidad de Jaén**  
*Centro de Estudios de Postgrado*

Trabajo Fin de Máster

**Análisis e integración de datos de  
expresión génica para el estudio de la  
infertilidad masculina**

**Alumno/a: Ortega Velasco, Juan Carlos**

Tutor/a: Prof. D. Francisco José Esteban Ruiz

Eva Vargas Liébanas

Dpto: Biología Experimental

*Septiembre, 2019*



**Universidad de Jaén**  
*Centro de Estudios de Postgrado*

Trabajo Fin de Máster

**Análisis e integración de datos de  
expresión génica para el estudio de la  
infertilidad masculina**

**Alumno/a: Ortega Velasco, Juan Carlos**

Tutores: Francisco José Esteban Ruiz  
Eva Vargas Liébanas  
Dpto: Biología Experimental

*Septiembre, 2019*

# ÍNDICE DE CONTENIDOS

ABSTRACT .....	4
RESUMEN .....	4
1. INTRODUCCIÓN .....	5
1.1. <i>La espermatogénesis</i> .....	5
1.2. <i>Tipos de esterilidad e infertilidad</i> .....	7
1.3. <i>Genómica</i> .....	11
1.4. <i>Otras disciplinas “-ómicas”</i> .....	12
2. OBJETIVOS.....	14
3. MATERIAL Y MÉTODOS.....	15
3.1. <i>Obtención de datos</i> .....	15
3.2. <i>Construcción y análisis de redes</i> .....	16
3.3. <i>Tratamiento de la información con RStudio</i> .....	17
3.4. <i>Análisis de enriquecimiento funcional</i> .....	18
3.5. <i>Validación de resultados en GEO DataSets</i> .....	18
4. RESULTADOS.....	20
4.1. <i>Recogida y clasificación de datos</i> .....	20
4.2. <i>Análisis de clúster</i> .....	21
4.3. <i>Selección de genes y construcción de redes</i> .....	23
4.4. <i>Identificación de procesos biológicos</i> .....	25
4.5. <i>Validación de resultados</i> .....	26
4.5.1 <i>Datasets analizados</i> .....	27
4.5.1.1. <i>Azoospermia</i> .....	27
4.5.1.2. <i>Criptorquidismo/ Criptorquidia</i> .....	28
4.5.1.3. <i>Induración plástica del pene/ Enfermedad de La Peyronie</i> .....	28
4.5.1.4. <i>Hiperplasia prostática</i> .....	29
4.5.1.5. <i>Neoplasma prostático</i> .....	30
4.5.1.6. <i>Neoplasma testicular</i> .....	30
4.5.1.7. <i>Neoplasma peneal</i> .....	30
5. DISCUSIÓN .....	34
6. CONCLUSIONES .....	40
7. BIBLIOGRAFÍA .....	41
ANEXOS.....	CD

## ABSTRACT

Infertility is one of the major health problems affecting humans globally. The aim of this Master's Thesis is to detect biological processes and biomarkers that can give rise to male infertility using bioinformatic approaches. The results obtained show that a set of genes seem to have a differential expression that could be related to male infertility. We could also detect the main biological processes in which these genes are involved: immune response, presentation and processing of antigens, regulation of gene expression and transcription, cell proliferation, oxide-reduction processes and response to drugs or xenobiotic compounds.

Taken together, bioinformatic analyses are presented as useful tools that support wet lab experiments and help to discover new clues in the clinical field.

**Keywords:** *Infertility-Bioinformatics-Spermatogenesis-Genomics-Proteomics-Metabolomics-Transcriptomics-Epigenetics*

## RESUMEN

La infertilidad es uno de los principales problemas de salud que afecta a los seres humanos de forma global. En este Trabajo de Fin de Máster se han tratado de encontrar, a través del uso de procedimientos bioinformáticos, procesos biológicos y potenciales biomarcadores que puedan explicar por qué se produce la infertilidad masculina. Los resultados obtenidos muestran un conjunto de genes que parecen tener una expresión diferencial relacionada con la infertilidad masculina. A través de los análisis realizados también se han podido conocer los procesos biológicos en los que se encuentran implicados dichos genes: respuesta inmune, presentación y procesamiento de antígenos, regulación de la expresión génica y la transcripción, proliferación celular, procesos de óxido-reducción y respuesta a fármacos o compuestos xenobióticos.

Como conclusión, se puede decir que los análisis bioinformáticos constituyen herramientas interesantes que dan soporte a los experimentos de laboratorio y ayudan a descubrir nuevos indicios en el campo de la clínica.

**Palabras clave:** *Infertilidad-Bioinformática-Espermatogénesis-Genómica-Proteómica-Metabolómica-Transcriptómica-Epigenética*

# 1. INTRODUCCIÓN

La infertilidad es una enfermedad que se podría denominar como relativamente común, ya que afecta a un 15% de las parejas (Agarwal *et al.*, 2015). Se entiende como infertilidad la imposibilidad de llevar a término un embarazo después de un período de 12 meses, o más, manteniendo relaciones sexuales sin protección de forma regular (Lindsay & Vitrikas, 2015). La incidencia de esta enfermedad varía en función del sexo: entre el 20-30% de los casos tiene su origen en el hombre, mientras que en la mujer puede llegar a ocupar un 50% de los mismos; asimismo, puede ocurrir que tenga un carácter compartido, hecho que sucede en un 20-30% de las ocasiones (Agarwal *et al.*, 2015; Sharlip *et al.*, 2002).

Un paso muy importante que puede ayudar a tener una mejor comprensión de las enfermedades que causan este tipo de desórdenes es el conocimiento de los factores genéticos que los propician. Sin embargo, en la infertilidad masculina se puede ver que en una gran cantidad de casos se consideran las causas idiopáticas, lo que refleja el poco conocimiento existente relacionado con la formación del espermatozoides (Ferlin *et al.*, 2007). Por otra parte, a la hora de realizar la diagnosis se llevan a cabo análisis de la calidad del espermatozoides basados en concentración, motilidad y morfología. Sin embargo, estos parámetros pueden verse alterados a su vez por otras causas subyacentes, cuyos defectos moleculares permanecen aún desconocidos (Krausz *et al.*, 2015). Los análisis que se realizan actualmente se basan en la forma y función de los espermatozoides y para ello es conveniente estudiar la biología de las células espermáticas.

## 1.1. La espermatogénesis

La espermatogénesis tiene lugar en los testículos, concretamente en los túbulos seminíferos, y ocurre en tres fases (Figura 1) (Amann, 2008):

1. En la primera fase se parte de una célula diploide indiferenciada que da lugar a dos espermatogonias: A y B. Las espermatogonias A se replican constantemente para mantener la población de espermatogonias. Las B, por su parte, se dividen a través de meiosis dando lugar a espermatozoides.
2. Los espermatozoides pasan a ser espermatocitos, que a través de dos divisiones meióticas consecutivas dan lugar a espermátidas haploides.
3. En el último paso se produce la espermiogénesis: cambios en las espermátidas a nivel celular que dan lugar a los espermatozoides maduros (Neto *et al.*, 2016).

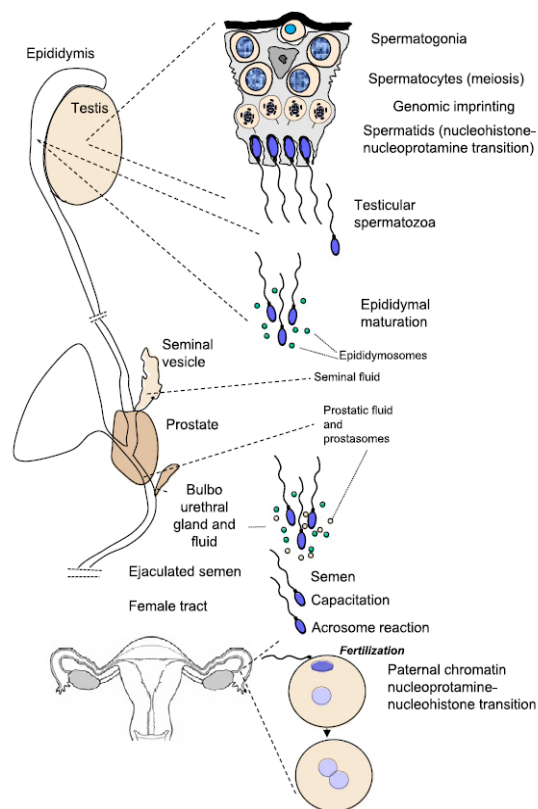


Figura 1. Fases de maduración de los espermatozoides. El proceso de maduración consta de las siguientes etapas: comienzo de la espermatogénesis en el testículo a partir de las espermatogonias, maduración epididimal, combinación con el fluido seminal y finalización de la maduración en el tracto femenino (Jodar *et al.*, 2017).

Algunos de los cambios remarcables que se producen durante la espermiogénesis son la formación del acrosoma (con las enzimas necesarias para que se produzca la fecundación) y la cola. En lo que al exceso de citoplasma respecta, una parte se elimina en forma de cuerpos residuales y otra parte se almacena en el cuello para ser eliminado posteriormente durante la fase final de la maduración (Cooper, 2005). Las espermatogonias resultantes de estos procesos de diferenciación son todavía unas células germinales inmaduras, incapaces de fertilizar un oocito por sí mismas. Para que adquieran esa capacidad (adquiriendo por lo tanto movilidad y potencial fertilizante), es necesario que transcurran una o dos semanas avanzando en el tránsito por el epidídimo hasta llegar al tracto femenino (Figura 1) (Jodar *et al.*, 2017).

Las células epiteliales del epidídimo secretan epididismos, vesículas que interactúan con el espermia originando cambios en su superficie y en la adquisición de nuevos componentes celulares (proteínas y ARNs) (Sullivan *et al.*, 2005). Estos cambios son los que contribuyen, entre otros factores nombrados anteriormente, a la adquisición por parte de los espermatozoides de motilidad y potencial de fecundación, así como la capacidad de transferir

información epigenética translacional (Atala, 2015; Sharma *et al.*, 2016). En el epidídimo, además de llevarse a cabo estas etapas de maduración, también se provee a las células espermáticas de un microambiente que permite que dichas células permanezcan en estado quiescente durante su almacenamiento en el epidídimo, evitando así una maduración temprana (Acott *et al.*, 1984).

Las últimas etapas de la maduración del esperma ocurren en el interior del tracto femenino. Justo después de la eyaculación, los espermatozoides se exponen a distintas secreciones vaginales que inician el proceso de capacitación. Durante la capacitación, el esperma vuelve a sufrir modificaciones en su superficie, perdiendo de esta forma la membrana específica proteica (Travis *et al.*, 2002). Es conveniente señalar que alteraciones del esperma en algunas de las fases de maduración citadas pueden deberse no solo a algún factor intrínseco al organismo del hombre, sino que también pueden estar provocadas por sustancias exógenas como toxinas, alcohol o tabaco (Lindsay & Vitrikas, 2015).

## **1.2. Tipos de esterilidad e infertilidad**

Tanto en hombres como en mujeres podemos encontrar dos tipos de esterilidad: primaria y secundaria. La primaria es aquella que afecta a la estructura de las células germinales o su fisiología, lo que provoca la parada del desarrollo de dichas células y en última instancia, su muerte. La secundaria, en cambio, surge a partir de defectos genéticos, incluidos los que se refieren al desarrollo, al sistema endocrino o a defectos metabólicos (Zorrilla & Yatsenko, 2013). De la misma manera, se pueden observar dos tipos de infertilidad siguiendo el mismo criterio: primaria y secundaria. La primaria se produce cuando no se ha podido lograr el embarazo y la secundaria cuando sí que ha sido posible que este ocurra en el pasado al menos una vez, pero no en la actualidad (Poongothai *et al.*, 2009).

Como se puede observar, hasta ahora la evaluación de la infertilidad masculina está limitada a analizar los parámetros del esperma antes mencionados (concentración, morfología, movilidad y vitalidad, entre otros). Aunque dichos parámetros pueden ayudar con el tratamiento, en algunos casos tienen un valor predictivo muy bajo en lo que al embarazo se refiere. Este hecho indica que conviene estudiar el desarrollo de nuevos biomarcadores que mejoren el valor predictivo de los ya existentes. Algunas de las técnicas que nos pueden ayudar a ello son las denominadas tecnologías “-ómicas” de alta resolución como la proteómica, la genómica y la metabolómica, entre otras (Jodar *et al.*, 2017). Dichas disciplinas aplican técnicas como los SNP (*Single Nucleotide Polymorphism*) arrays y las de NGS (*Next Generation Sequencing*), que aportan una gran cantidad de datos sobre distintas variantes

(Krausz *et al.*, 2015). Además, estas técnicas no solo ayudan a mejorar el valor predictivo a la hora de que se produzca el embarazo (de cara a saber cómo influye la enfermedad o enfermedades que parecen imposibilitar dicho embarazo), sino que también lo hacen en el momento de realizar un diagnóstico, ya que, como se ha comentado antes, en una buena parte de los casos las causas permanecen desconocidas (infertilidad idiopática) (Figura 2) (Poongothai *et al.*, 2009).

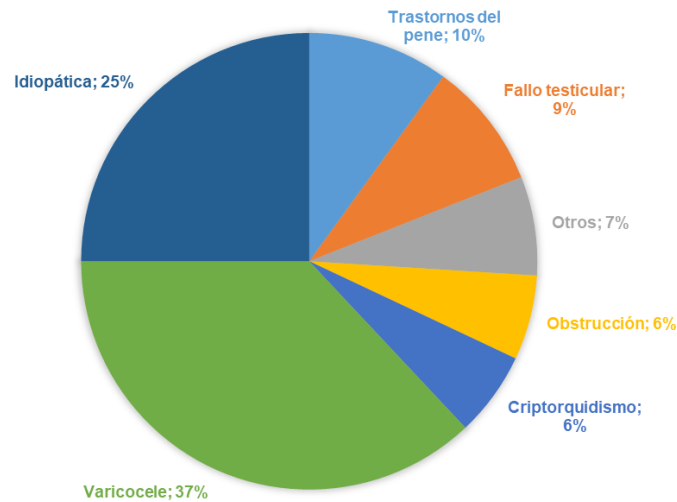


Figura 2. Distribución de las causas conocidas de infertilidad masculina: el 37% se atribuye a varicocele, el 25% a causas idiopáticas, el 10% a enfermedades que afectan al pene el 9% a fallos testiculares, tanto las obstrucciones de los conductos como el criptorquidismo tienen un 6%, y el resto de causas con un porcentaje menor copan el 7% (adaptado de Poongothai *et al.*, 2009).

Pese a las presentes limitaciones, las causas de infertilidad conocidas hasta la fecha son numerosas y pueden ser agrupadas en varias categorías, asociadas con factores genéticos y no genéticos (Poongothai *et al.*, 2009) (Figura 3).

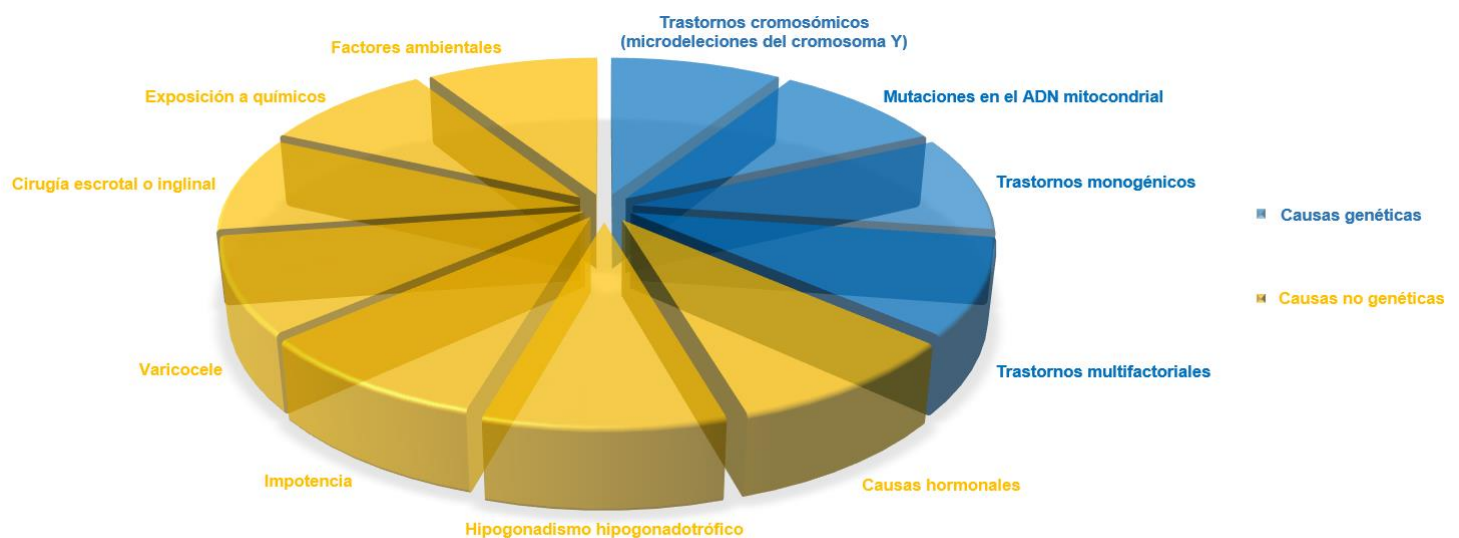


Figura 3. Algunas causas conocidas asociadas a infertilidad masculina. Fundamentalmente se consideran dos tipos de causas: genéticas (en azul) y no genéticas (se muestran en amarillo). Dentro de las genéticas

encontramos: trastornos cromosómicos, mutaciones en el ADN mitocondrial, trastornos monogénicos y trastornos multifactoriales; para las causas no genéticas tenemos las siguientes: factores ambientales, exposición a químicos, cirugía, varicocele, impotencia, hipogonadismo hipogonadotrófico y procesos hormonales (adaptado de Poongothai *et al.*, 2009).

Teniendo en cuenta la cantidad de factores que intervienen en la infertilidad (genéticos, epigenéticos y ambientales) es fácil llegar a la conclusión de que un enfoque reduccionista no parece la clave para llegar a la(s) causa(s) última(s) que provoca(n) la enfermedad. Es ahí donde entran en juego las “-ómicas”, disciplinas antes mencionadas, con un enfoque más global, y que parece que pueden ayudar a avanzar más rápida y eficazmente en la búsqueda de respuestas.

El comienzo de la aplicación de estas disciplinas al estudio de la infertilidad se remonta a la finalización del proyecto “Genoma Humano”, gracias al cual se incrementó el número de trabajos que identificaban genes que se expresaban en los testículos y que, por tanto, estaban implicados en la espermatogénesis, así como de las variantes de dichos genes asociadas a la infertilidad (Figura 4) (Carrell *et al.*, 2016).

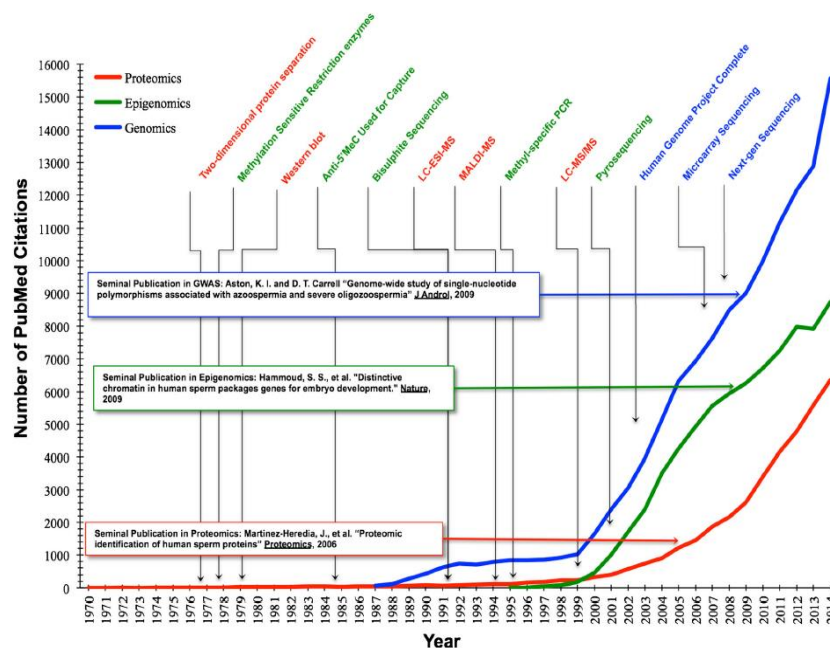


Figura 4. Incremento del número de publicaciones “-ómicas” durante los últimos 50 años. Se muestra el número de citas registradas en PubMed, así como los principales hitos en las diferentes disciplinas: proteómica (rojo), epigenómica (verde) y genómica (azul) (Carrell *et al.*, 2016).

Por ejemplo, en 2007 se llevaron a cabo análisis de SNPs (*Single Nucleotide Polymorphisms*) en personas globozoopérmicas mediante el uso de arrays (Dam *et al.*, 2007)

y en 2009 se llevó a cabo el primer estudio de asociación génica (GWAS, siglas en inglés de *Genome-Wide Association Studies*) de pacientes con espermatogénesis deficiente (Aston & Carrell, 2009).

Hasta hace relativamente poco dichas técnicas tenían costes prohibitivos, solo hay que observar las dimensiones que abarcó el proyecto “Genoma Humano”, llevado a cabo por un consorcio público de ámbito nacional e iniciativa privada de forma paralela con un coste millonario. Sin embargo, esta tendencia está cambiando en los últimos años, y en la actualidad se está consiguiendo abaratar los costes y el acceso a estas técnicas, lo que tiene como resultado en última instancia un aumento del número de publicaciones que utilizan las “-ómicas” como herramienta (Figuras 4 y 5) (Carrell *et al.*, 2016).

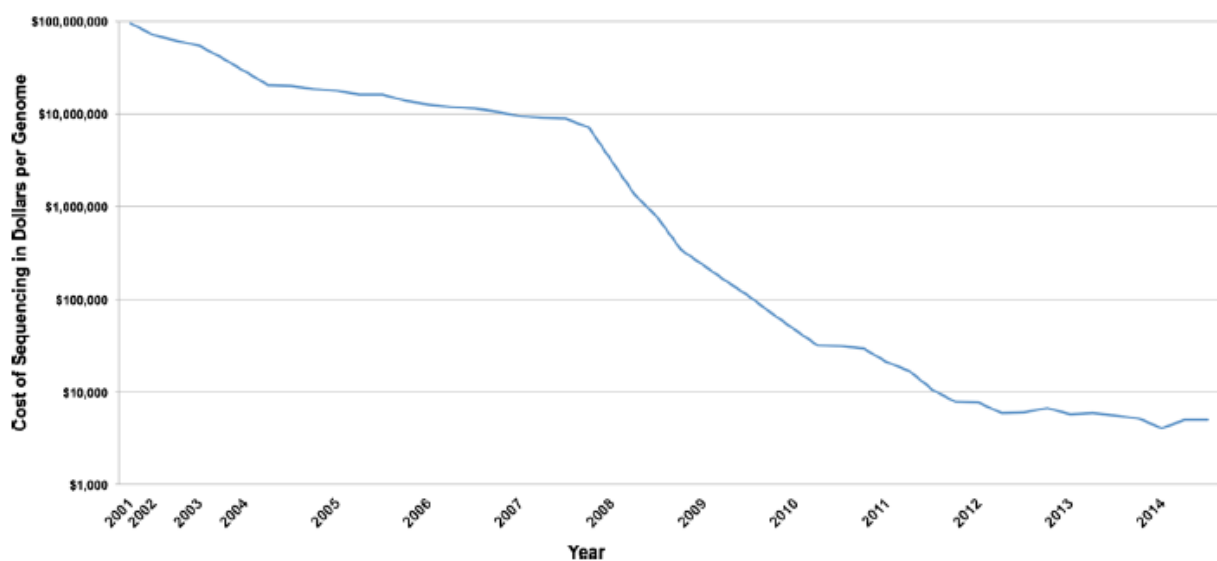


Figura 4. Evolución del coste de las técnicas de secuenciación desde el año 2000. Tal y como se observa, se ha producido una mejora de la accesibilidad a la secuenciación del genoma a lo largo de los últimos 15 años (Carrell *et al.*, 2016).

Durante muchos años, el estudio de la infertilidad masculina ha residido en el estudio de secuencias de genes individuales en cohortes de un tamaño modesto compuestas por hombres infértiles y controles fértiles o normozoospermicos. Estos estudios se basaban en el análisis de numerosos SNPs potencialmente asociados con la enfermedad. Sin embargo, estas secuencias estudiadas daban fuertes indicios de la necesidad de una caracterización más completa del genoma (Carrell *et al.*, 2016).

El incremento del número de publicaciones parece ser uno de los mejores indicadores que hay para confiar en que la maduración de las tecnologías “-ómicas” puede ayudar a una mejor caracterización genómica, epigenómica y proteómica de la arquitectura de la infertilidad en hombres.

### 1.3. Genómica

El éxito de la aplicación de herramientas de análisis masivo del genoma ha permitido caracterizar mejor la etiología de la infertilidad masculina. Sin embargo, se necesitan estudios del genoma completo basados en la población y diseñados cuidadosamente para identificar la mayoría de las variantes responsables de la infertilidad masculina (Carrell *et al.*, 2016). El mayor desafío al que se enfrenta la Biología de la Reproducción es el desarrollo de herramientas analíticas efectivas para priorizar y detectar con precisión variantes genómicas con bajas incidencias en una población.

La acumulación de datos de la secuencia del genoma de hombres infértiles cuidadosamente fenotipados indudablemente arrojará en los próximos años nueva luz sobre las complejidades de la infertilidad masculina y, con suerte, allanará el camino para mejorar las estrategias de diagnóstico y tratamiento (Carrell *et al.*, 2016).

Las herramientas analíticas genómicas se pueden clasificar fundamentalmente en dos grupos: enfoques basados en matriz (array) y secuenciación (NGS) (Aston & Conrad, 2013). Estos dos enfoques son los que han permitido el descubrimiento de variantes genéticas y asociaciones genómicas en numerosas enfermedades complejas sin un conocimiento específico de las regiones genómicas que participan. Esto se debe a su capacidad para interrogar el genoma completo a varios niveles de resolución (Carrell *et al.*, 2016). Los enfoques basados en microarrays, que son una de las herramientas utilizadas para la obtención de los datos con los que se va a trabajar en este proyecto, están diseñados para evaluar el genoma en su totalidad (o un subconjunto del mismo), con una resolución y cobertura que dependen del diseño y del número de sondas de la matriz. Estos soportes se han utilizado en el estudio de enfermedades complejas y han demostrado ser eficaces para identificar características genómicas asociadas con diversas enfermedades (Carrell *et al.*, 2016). Inevitable e indudablemente, como todas las técnicas, tienen ciertas limitaciones, ya que solo interrogan una pequeña fracción del genoma y, por lo general, están diseñados para identificar variables con una frecuencia relativamente alta dentro de la población (>1%), aunque hay algunos fabricantes que intentan solucionar esto en sus soportes (Carrell *et al.*, 2016).

Los enfoques basados en NGS superan algunas de las limitaciones de los microarrays y aumentan la cantidad de datos generados, lo que a su vez aumenta la complejidad del análisis y el coste del mismo. Además, el valor de los datos de la secuenciación completa del genoma está limitado, ya que desconocemos la función de una parte significativa de este.

Debido a estas razones, la interpretación de las asociaciones identificadas en regiones del genoma no conocidas está limitada significativamente. A medida que las tecnologías genómicas avanzan en su desarrollo, disminuyen los costes asociados a su uso y se vuelven más accesibles para el estudio de enfermedades complejas, incluida la infertilidad masculina.

Estos enfoques han producido éxitos en sus etapas iniciales, pero dada la complejidad genética de la espermatogénesis y la heterogeneidad de los fenotipos de infertilidad masculina, la caracterización exhaustiva de la base genética de esta enfermedad continuará siendo un área activa de investigación durante muchos años. Desde que estos primeros estudios fueron publicados, los estudios genómicos que se han realizado con relación a la infertilidad en hombres se han expandido rápidamente. Se han utilizado GWAS para identificar polimorfismos asociados con el síndrome de la disgénesis testicular (Dalgaard *et al.*, 2012) y el tamaño de la familia (Kosova *et al.*, 2012), así como para profundizar en el estudio de la espermatogénesis deficiente. En la mayoría de estos estudios se identificaron pequeñas variantes comunes, sin embargo, cuando se intentó replicar en otras situaciones se produjo un fracaso absoluto (Chihara *et al.*, 2015). Esto puede deberse a que no solo entran en juego los SNPs, sino que encontramos también otros fenómenos como las variaciones en el número de copias (CNVs, del inglés *Copy Number Variants*) cuya implicación puede ser desvelada gracias al análisis a través de GWAS (Krausz *et al.*, 2012).

#### 1.4. Otras disciplinas “-ómicas”

En este Trabajo de Fin de Máster se ha utilizado principalmente la información procedente del genoma y transcriptoma (entendido este último como el conjunto de los transcritos de un tejido, célula u órgano), si bien también existen otras disciplinas “-ómicas” como las que se mencionan a continuación:

- **Transcriptómica:** se trata del estudio de los transcritos de un tejido, célula u órgano (Vargas *et al.*, 2018).
- **Proteómica:** la proteómica, término acuñado por analogía con la genómica, estudia los cambios en todas las proteínas expresadas y traducidas a partir de un único genoma (Bhagwat *et al.*, 2013). El término proteoma, una fusión entre proteína y genoma, se refiere a todas las proteínas producidas por las células y liberadas en el fluido biológico circundante (Vargas *et al.*, 2018).
- **Metabolómica:** es la ciencia que explora las interacciones dinámicas entre múltiples moléculas de bajo peso molecular (Huang *et al.*, 2009b). Los metabolitos pueden (potencialmente) contener más información sobre un sistema biológico

que los ARNms o las proteínas, debido a que representan los productos finales de los procesos regulados en las células (Vargas *et al.*, 2018).

- **Epigenética:** esta rama del conocimiento explica las variaciones en la expresión y regulación génica sin que se produzcan variaciones en el ADN. La fertilización implica numerosos procesos fisiológicos, incluyendo el movimiento de las células germinales masculinas a lo largo de todo el sistema reproductivo femenino; el apego y la penetración de la zona pelúcida; y la unión, fusión y penetración en el ovocito (Vargas *et al.*, 2018).

## 2. OBJETIVOS

El objetivo de este Trabajo de Fin de Máster es caracterizar las posibles similitudes entre distintas enfermedades que afectan al tracto reproductivo masculino y que dan lugar a infertilidad masculina, a través de la Biología de Sistemas. Para llevar a cabo dicho objetivo general se deben completar los siguientes objetivos específicos:

- Se revisará la información existente relacionada con la infertilidad masculina.
- Se estudiarán los genes y procesos biológicos involucrados en las distintas alteraciones de la infertilidad masculina.
- Por último, se buscará un patrón entre todas las enfermedades estudiadas que pueda explicar las causas moleculares subyacentes a la infertilidad.

### 3. MATERIAL Y MÉTODOS

En la figura 6 se puede observar el esquema de trabajo general que se ha seguido en este trabajo, y que se detalla posteriormente a lo largo de este apartado.

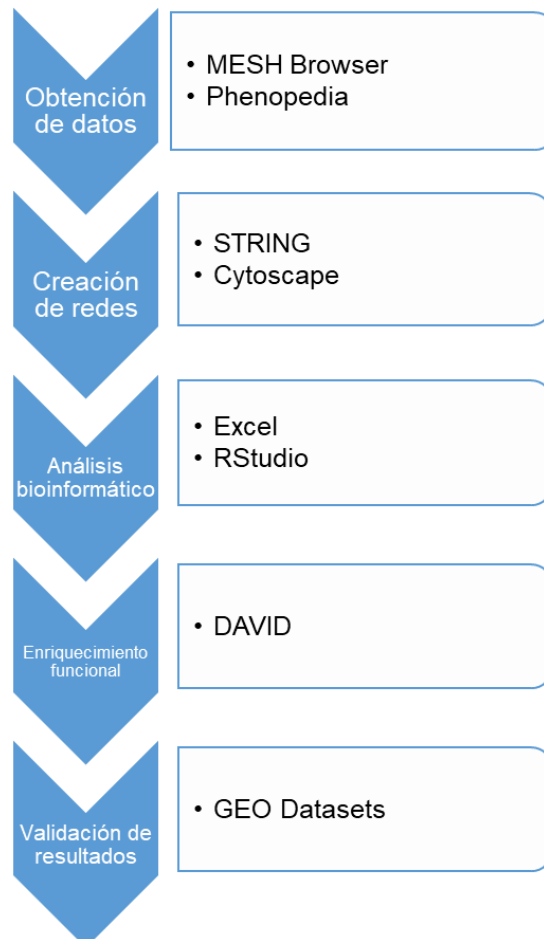


Figura 6. Etapas seguidas durante el proceso de análisis, desde la obtención de datos hasta la validación de resultados, así como software empleado en cada una de ellas.

#### 3.1. Obtención de datos

Para la caracterización de todos los genes que participaban en trastornos responsables de la aparición de infertilidad masculina se realizó en primer lugar una búsqueda de todas las enfermedades vinculadas con este trastorno registradas hasta el momento (MeSH Terms) en la web del “U.S. National Library of Medicine”, más en concreto en el buscador de términos médicos “Medical Subject Headings” (MeSH Browser) (<https://www.nlm.nih.gov/mesh/>). A continuación, se extrajeron los códigos ICD10 de las diferentes enfermedades de la web “ICD10data.com” (<https://www.icd10data.com/>).

Una vez obtenida la lista de enfermedades, se procedió a buscar los genes registrados que se han relacionado con ellas. Para esta tarea, se utilizó la base de datos Phenopedia (Yu *et al.*, 2009) (<https://phgkb.cdc.gov/PHGKB/startPagePhenoPedia.action>). Phenopedia es una aplicación web que relaciona asociaciones genéticas humanas procedentes de la literatura a través de una base de datos constantemente actualizada desde PubMed (<https://www.ncbi.nlm.nih.gov/pubmed/>) (Díaz-Beltrán *et al.*, 2017).

### 3.2. Construcción y análisis de redes

Para la construcción de redes de interacción molecular proteína-proteína se utilizó el software STRING (*Search Tool for Retrieval of Interacting Genes/Proteins*) ([https://string-db.org/cgi/input.pl?sessionId=gFa80GsaDG6p&input\\_page\\_show\\_search=on](https://string-db.org/cgi/input.pl?sessionId=gFa80GsaDG6p&input_page_show_search=on)), que permite obtener y representar las interacciones existentes entre un grupo de genes de interés. Se emplearon los parámetros por defecto. La base de datos STRING tiene como objetivo recopilar, puntuar e integrar todas las fuentes públicas de información sobre la interacción proteína-proteína, y complementarlas con predicciones computacionales. Su objetivo es lograr una red mundial amplia y objetiva, que incluya interacciones funcionales directas (físicas) e indirectas. Una de las funcionalidades de STRING es la de cargar conjuntos de datos completos de todo el genoma como entrada, lo que permite a los usuarios visualizar subconjuntos como redes de interacción y realizar análisis de enriquecimiento del conjunto de genes utilizado como entrada (*input*). Para el análisis de enriquecimiento, STRING implementa sistemas de clasificación bien conocidos como Gene Ontology y KEGG, pero también ofrece nuevos sistemas de clasificación adicionales basados en la minería de texto de alto rendimiento, así como en una agrupación jerárquica de la propia red de asociaciones (Szklarczyk *et al.*, 2019).

Para un análisis avanzado de las redes previamente generadas con STRING se utilizó el software Cytoscape en su versión 3.6.1 (<http://www.cytoscape.org/>), un software libre que integra redes de interacción molecular con datos de expresión de alta resolución y otro tipo de datos moleculares (Vargas *et al.*, 2015). Utilizando esta herramienta se analizaron dos medidas clásicas de centralidad de redes: la centralidad de grado (*degree*) (1) y la intermediación (*betweenness*) (2) (Boccaletti *et al.*, 2006).

La centralidad de grado o *degree*  $k_i$  de un nodo  $i$  es el número de vértices incidentes con el nodo y se define en términos de la matriz de adyacencia  $a$  como:

$$k_i = \sum_{j \in N} a_{ij}, \quad (1)$$

La intermediación o *betweenness*  $b_i$  de un nodo  $i$  se define como:

$$b_i = \sum_{j,k \in N, j \neq k} \frac{n_{jk(i)}}{n_{jk}}, \quad (2)$$

donde  $n_{jk}$  es el número de las rutas más cortas que conectan  $i$  y  $j$ , y  $n_{jk(i)}$  es el número de las rutas más cortas que conectan  $i$  y  $j$  y pasan a través de  $i$  (Vargas *et al.*, 2015).

### 3.3. Tratamiento de la información con RStudio

R Studio es una herramienta de código abierto para el entorno de programación R (<https://www.rstudio.com/>). A través de un script personalizado (archivo de órdenes para una tarea concreta), diseñado especialmente para el análisis de genes, se trató de observar en cuántos grupos se organizaban las enfermedades en función de la similitud obtenida entre los genes que compartían.

Para llegar a hacer la clasificación por grupos se siguieron los siguientes pasos: en primer lugar, se construyó una matriz binaria con las enfermedades y sus correspondientes genes (se le asignaba el valor de 1 cuando el gen estaba presente en la enfermedad y 0 cuando no). Después se utilizó el paquete *vegan* (Oksanen *et al.*, 2019) para construir la matriz de disimilitud correspondiente, que se calculó utilizando el índice de Jaccard (definido como el tamaño de la intersección dividido por el tamaño de la unión de los conjuntos de la muestra) como medida para comparar la similitud y diversidad de los conjuntos de la muestra. Como resultado, obtuvimos una matriz de distancia que se utilizó para realizar un análisis jerárquico de conglomerados con el fin de construir un dendrograma basado en genes para mostrar la relación entre las distintas enfermedades analizadas. Utilizamos el método de enlace promedio (*average linkage method*) para calcular la distancia entre cada clúster, que define la distancia entre los dos clústeres como la distancia promedio entre cada punto de un clúster y cada punto del otro clúster (Diaz-Beltran *et al.*, 2017).

Para comprobar la fiabilidad de los clústeres obtenidos, utilizamos *clusterboot()*, una función incluida en el paquete de R *fpc* (traducción de las iniciales de *procedimientos flexibles para la agrupación*) que evalúa la estabilidad de los clústeres mediante un “remuestreo” (Hennig, 2007). Esta función integrada realiza un muestreo de tipo *bootstrap* para determinar

la estabilidad de un clúster dado usando el coeficiente de Jaccard como medida de similitud entre conjuntos. Para lograr este propósito, este método primero agrupa los datos y luego dibuja un nuevo conjunto de datos “remuestreando” el original; después de eso, agrupa el nuevo conjunto de datos y para cada grupo en el grupo original trata de encontrar aquel que da el máximo coeficiente de Jaccard (el más similar). Este método disuelve todos los clústeres con coeficientes de Jaccard inferiores a 0,5 (lo que significa que probablemente no es un clúster real) y repite el procedimiento varias veces. Finalmente, la estabilidad de cada clúster en el clúster original será igual al valor medio de su coeficiente de Jaccard sobre todas las iteraciones del *bootstrap*. En general, se considera que los conglomerados con un valor de estabilidad inferior a 0,6 son inestables, y los valores superiores a 0,75 están asociados a una medida "real" de los patrones de los datos y podrían considerarse muy estables, ya que es probable que sean “conglomerados reales” (Hennig, 2007). Para comprobar la fiabilidad de los resultados del análisis de clústeres, aplicamos un método de *bootstrap* (B=1000 ejecuciones) para “remuestrear” con reemplazo de los datos originales con el fin de generar matrices de *bootstrap* y clústeres y usamos, de nuevo, el coeficiente de Jaccard como medida de la similitud de los árboles obtenidos (Díaz-Beltrán *et al.*, 2017).

La estabilidad y validez de las agrupaciones se calcularon sobre la base de la media de los coeficientes de Jaccard en todas las permutaciones. El número de subconjuntos para cada permutación se estableció en 13 (k=13) de acuerdo con el número total de conglomerados obtenidos en el árbol observado originalmente.

### **3.4. Análisis de enriquecimiento funcional**

Para realizar un estudio completo de las funciones que llevan a cabo los genes de interés se realizó un análisis de enriquecimiento funcional de los mismos utilizando la base de datos DAVID (*Database for Annotation, Visualization and Integrated Discovery*) en su versión 6.7 (<https://david.ncifcrf.gov/>), que permite analizar los procesos biológicos en los que se encuentran implicados un grupo dado de genes (Huang *et al.*, 2009a; 2009b).

### **3.5. Validación de resultados en GEO DataSets**

Con los anteriores pasos pudimos obtener una serie de genes y procesos biológicos comunes y altamente representados dentro de las enfermedades estudiadas. Para la búsqueda de estudios ya publicados que respaldasen los resultados obtenidos utilizamos la base de datos “GEO DataSets” (<https://www.ncbi.nlm.nih.gov/gds/?term=>). En este repositorio se incluye la herramienta “GEO2R” (<https://www.ncbi.nlm.nih.gov/geo/geo2r/>), en la que se permite realizar el análisis de los diferentes datasets para, por ejemplo, comparar dos o más

grupos de muestras para identificar genes diferencialmente expresados según los parámetros estadísticos que se establezcan. Esta información se presenta como una tabla de genes ordenados por significancia.

De esta forma, podemos saber si los genes presentes tanto en las enfermedades estudiadas como en los procesos biológicos identificados tienen relevancia real ya demostrada o, por el contrario, pueden ser un punto a tener en cuenta a la hora de avanzar en futuras investigaciones.

## 4. RESULTADOS

### 4.1. Recogida y clasificación de datos

El conjunto total de enfermedades detectadas a partir de la base de datos Phenopedia con las que se trabajó a lo largo del desarrollo de este Trabajo de Fin de Máster se representa en la Tabla 1. La contextualización desde el punto de vista clínico se realizó a través de la transformación de los códigos ICD-10 en términos MeSH, lo que facilitaría la posterior obtención de los listados de genes asociados en la base de datos Phenopedia, pues utiliza dichos términos como motor de búsqueda.

	<b>Enfermedad – MeSH terms</b>	<b>ICD-10</b>	<b>Genes</b>
Genital Neoplasms	Penile neoplasms	Q54.1	13
	Prostatic neoplasms	N42.31	1152
	Testicular neoplasms	N46.9	107
	Sertoli-Leidyg tumor	N46.9	11
Infertility	Aspermia	N46.01	2
	Asthenozoospermia	R86.9	44
	Azoospermia	N46.0	228
	Oligospermia	N46.1	171
Penile diseases	Hypospadias	Q54.0	70
	Penile induration	N48.6	6
	Penile Neoplasms	N46.9	13
	Priapism	N48.30	14
Prostatic diseases	Prostatic hyperplasia	N40	152
	Prostatic neoplasms	N46.9	1152
	Prostatitis	N41	10
Sexual dysfunction, physiological	Dyspareunia	N53.12	12
Sexually transmitted diseases (bacterial)	Chlamydia	N46.9	69
	Gonorrhoea	A54.22	9
	Syphilis genitalia, primary	A51.0	12
Testicular diseases	Cryptorchidism	A07.2	84
Varicocele	Varicocele	I86.1	13
Herpes genitalis	Herpes simplex, genital	A60.0	8

Tabla 1. Relación de códigos ICD10, términos MeSH y número de genes implicados en cada una de las enfermedades analizadas.

Una vez obtenido el listado de enfermedades de partida y sus códigos, se accedió al conjunto de genes asociado a cada una de ellas a través de Phenopedia. Se analizaron un

total de 1507 genes expresados en 21 enfermedades. Los listados completos de genes para cada enfermedad pueden consultarse en el material suplementario (Anexo 1).

Los listados de genes para cada enfermedad fueron analizados con STRING con el objetivo de obtener las redes de interacción proteína-proteína para cada enfermedad. El único gen del que no se pudo obtener información en STRING y que por tanto no aparecerá en la posterior construcción de redes es el gen *GSTT1*.

Posteriormente, los datos obtenidos fueron sometidos a un análisis avanzado en Cytoscape que nos permitió realizar un estudio más en profundidad y detectar cuáles eran más representativos dentro de cada enfermedad. Para determinar la importancia de los genes (nodos) dentro de la estructura de la red, se escogieron dos medidas clásicas de centralidad: el *degree* y el *betweenness*, de manera que se consideraron como genes altamente conectados aquellos que poseían un *betweenness* igual a 1 y un *degree* superior a la media más dos veces la desviación estándar de los *degrees* de la lista de genes a la que pertenece. Siguiendo este criterio, solo se encontró un gen que respondía a dichos filtros, el gen *HLA-C*, que aparecía como un gen altamente conectado en la red de interacción de la sífilis (con un *degree*=4 (en este caso la media más 2 veces la desviación estándar es de 3,28) y un *betweenness*=1).

#### **4.2 Análisis de clúster**

A continuación, se procedió a estudiar qué genes aparecían más veces representados dentro del conjunto total de enfermedades. Para ello se construyó un dendrograma con R Studio (Figuras 7 y 8). La estabilidad del clúster se evaluó con el coeficiente de Jaccard (utilizado para medir el grado de similitud entre conjuntos) y de todos los obtenidos el clúster más estable fue el de 13 grupos ( $k=13$ ) con una media aritmética de 0,67 (Figura 8).

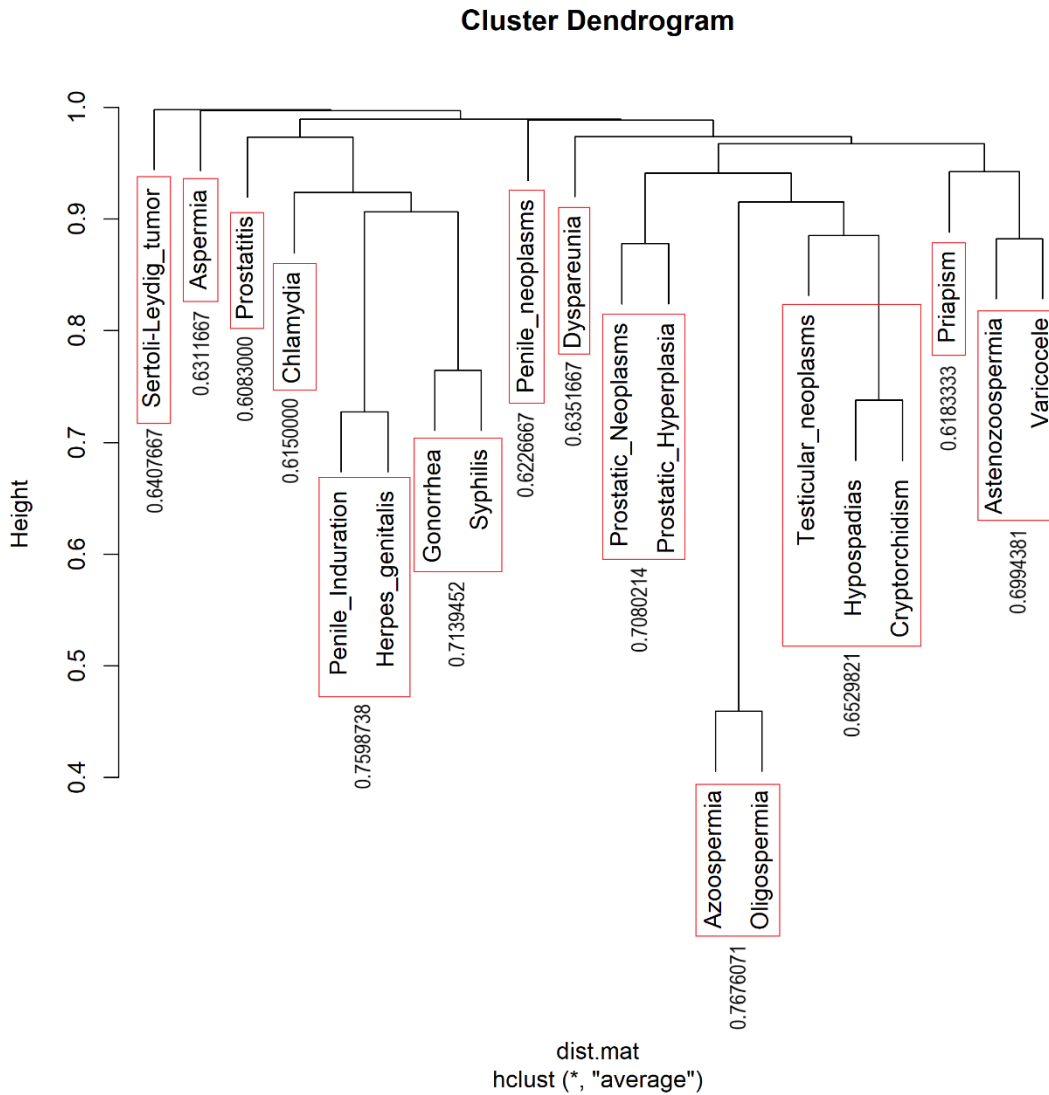


Figura 7. Clúster dendrograma que muestra la distribución de las enfermedades estudiadas en base a la similitud entre los genes implicados en las mismas. Hay un total de 13 subconjuntos, cada uno de los cuales lleva asignado su correspondiente coeficiente de estabilidad.

partition	Prostatic_Neoplasms	Penile_neoplasms	Testicular_neoplasms	Sertoli-Leydig_tumor
	1	2	3	4
	Aspermia	Astenozoospermia	Azoospermia	Oligospermia
	5	6	7	7
	Hypospadias	Penile_Induration	Priapism	Prostatic_Hyperplasia
	3	8	9	1
	Prostatitis	Dyspareunia	Chlamydia	Gonorrhea
	10	11	12	13
	Syphilis	Cryptorchidism	Varicocele	Herpes_genitalis
	13	3	6	8

Figura 8. Grupos establecidos en R en función del coeficiente de Jaccard.

### 4.3. Selección de genes y construcción de redes

Así pues, procedimos a hacer lo siguiente: seleccionamos el total de los genes y los agrupamos en función del número de enfermedades en las que se encontraban presentes (Tabla 2). En el material suplementario se puede consultar la tabla completa en la que se recopilan todos los genes y la frecuencia con la que aparecen en cada una de las enfermedades (Anexo 2).

Frecuencia acumulada			
n enfermedades	Genes	n enfermedades	Genes
x=1	1116	x≥1	1507
x=2	245	x≥2	391
x=3	90	x≥3	146
x=4	15	x≥4	56
x=5	19	x≥5	41
x=6	6	x≥6	22
x=7	10	x≥7	16
x=8	4	x≥8	6
x=9	2	x≥9	2

Tabla 2. Clasificación del número de genes que aparecen en un determinado número de enfermedades. En rojo se destaca el grupo seleccionado para continuar con el análisis.

Decidimos trabajar con los 41 genes que estaban presentes en 5 o más enfermedades (Tablas 2 y 3), ya que utilizar aquellos que estaban representados en un número de enfermedades mayor añadía más ruido al análisis y trabajar con un número menor nos hacía perder genes relevantes. Con estos datos se construyó una nueva red en Cytoscape compuesta por los genes seleccionados.

Genes presentes en 5 o más enfermedades			
AR	9	TGFB1	6
GSTM1	9	ACE	5
ESR1	8	AHR	5
GSTT1	8	COMT	5
HLA-DRB1	8	CYP1A2	5
NOS3	8	DICER1	5
CYP17A1	7	ERCC1	5
CYP19A1	7	FSHR	5
CYP1A1	7	HLA-A	5
CYP3A4	7	HLA-B	5
ESR2	7	IL10	5
GSTP1	7	IL1B	5
HLA-C	7	LTA	5
MTHFR	7	MTRR	5
SRD5A2	7	SHBG	5
TNF	7	TGFBR3	5
CFTR	6	TLR1	5
CYP1B1	6	TLR2	5
HFE	6	TLR6	5
HLA-DQB1	6	TP53	5
IL6	6		

Tabla 3. En las columnas 1 y 3 se representa el nombre de los genes presentes en 5 o más enfermedades. En las columnas 2 y 4, número de enfermedades en las que están presentes.

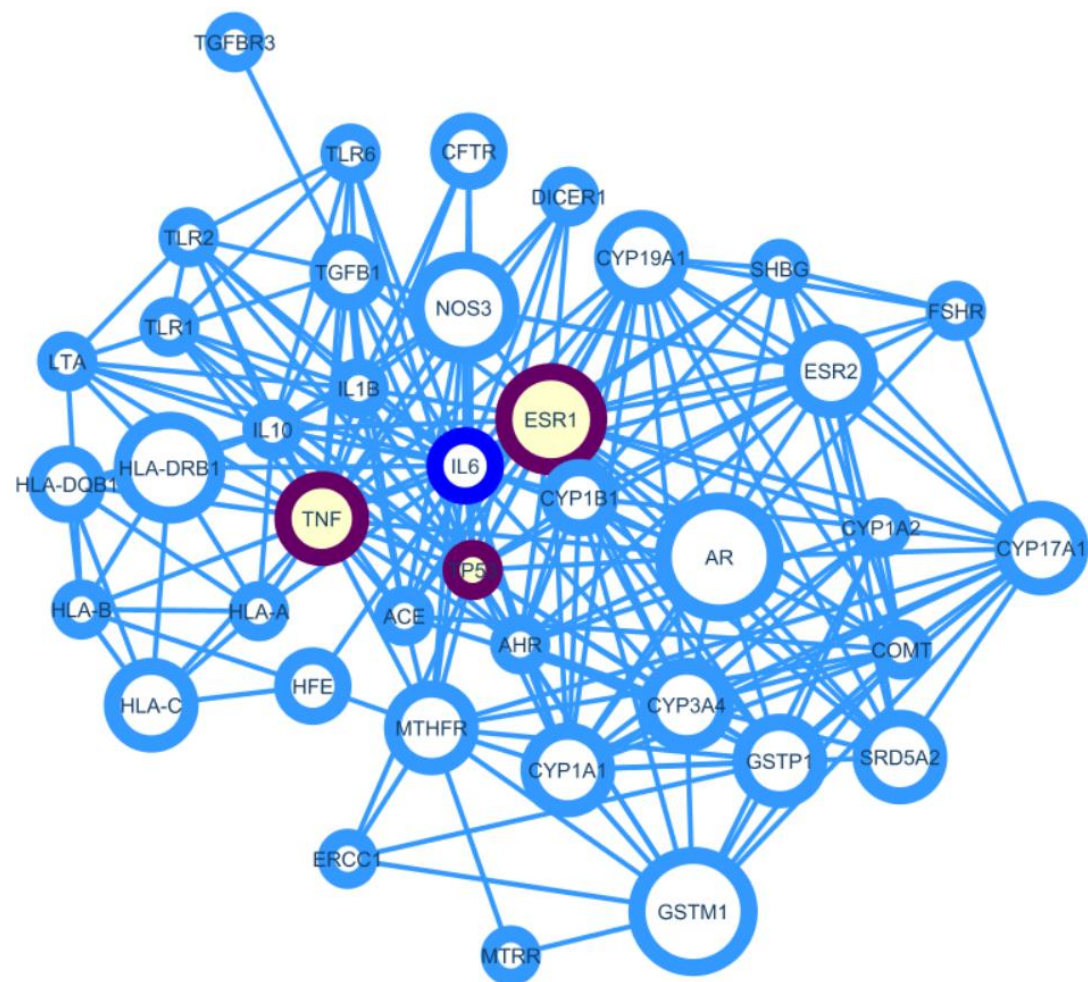


Figura 9. Red de interacción correspondiente a los genes presentes en 5 o más enfermedades. Los nodos (genes) con tamaño mayor son aquellos que aparecen en un mayor número de enfermedades. Aquellos con un color diferente al de la matriz son los que poseen un valor superior de centralidad (*degree*), según los parámetros antes mencionados (media más dos veces la desviación estándar); encontramos: *IL6* ( $k=26$ ), *ESR1* ( $k=23$ ), *TP53* ( $k=23$ ) y *TNF* ( $k=23$ ).

#### 4.4. Identificación de procesos biológicos

El análisis continuó con la búsqueda en DAVID de los procesos biológicos (BP) en los que intervenían dichos genes y la selección de aquellos procesos significativos con un FDR (*False Discovery Rate*) inferior a 0,05 (Tabla 4). Se detectaron 17 procesos biológicos significativos (FDR<0,05). En el material suplementario – Anexo 3 – puede consultarse la tabla completa.

Procesos Biológicos	Genes	PValor	FDR
Respuesta inmune	HLA-DQB1, IL6, TNF, HLA-DRB1, TLR1, TLR2, HLA-A, HLA-C, HLA-B, TLR6, IL10, IL1B, TGFB1, LTA	5,78E-12	9,08E-09
Procesamiento y presentación antigénica de antígenos peptídicos vía MHC clase I	HLA-DQB1, ACE, HLA-A, HFE, HLA-C, HLA-B	7,07E-09	1,11E-05
Procesamiento y presentación antigénica	HLA-DQB1, HLA-DRB1, HLA-A, HFE, HLA-C, HLA-B	1,66E-07	2,60E-04
Vía de señalización mediada por lipopolisacáridos	TNF, TLR2, IL1B, NOS3, TGFB1	8,52E-07	0,00133916
Regulación positiva de la expresión génica	AR, IL6, TNF, TP53, TLR2, HFE, IL1B, TGFB1	2,09E-06	0,00327736
Regulación positiva de la transcripción, ADN-templado	AR, IL6, TNF, TP53, ESR1, IL1B, ESR2, TGFB1, IL10, AHR	2,09E-06	0,00328063
Regulación negativa de la proliferación celular	AR, IL6, CYP1B1, TP53, TLR2, IL1B, NOS3, TGFB1, IL10	2,89E-06	0,00454149
Regulación positiva de la actividad del factor de transcripción de unión de una secuencia de ADN específica	IL6, TNF, ESR1, IL1B, ESR2, IL10	4,22E-06	0,00663427
Respuesta a fármacos	IL6, MTHFR, CYP1A1, SRD5A2, COMT, TGFB1, IL10, LTA	5,57E-06	0,0087446
Proceso metabólico andrógeno	CYP3A4, ESR1, SRD5A2, CYP19A1	6,35E-06	0,00997961
Procesos de óxido reducción	CYP3A4, CYP17A1, MTHFR, CYP1B1, CYP1A1, NOS3, SRD5A2, CYP1A2, CYP19A1, MTRR	6,51E-06	0,01022444
Respuesta celular a compuestos orgánicos cíclicos	TNF, CYP1B1, CYP1A1, IL1B, TGFB1	1,03E-05	0,01620153
Regulación positiva de la actividad del factor de transcripción NF-κβ	AR, IL6, TNF, TLR2, IL1B, TGFB1	1,34E-05	0,02106044
Procesamiento y presentación antigénicos de péptidos endógenos vía MHC clase I vía ER, TAP-independiente	HLA-A, HLA-C, HLA-B	1,57E-05	0,02473493
Ruta de señalización mediada por interferón-gamma	HLA-DQB1, HLA-DRB1, HLA-A, HLA-C, HLA-B	2,16E-05	0,03390027
Respuesta inflamatoria	IL6, TNF, TLR1, TLR2, IL1B, TLR6, TGFB1, IL10	2,33E-05	0,03657862
Procesos xenobióticos metabólicos	CYP3A4, CYP1B1, CYP1A2, AHR, GSTP1	3,13E-05	0,04918706

Tabla 4. Resultado del análisis de enriquecimiento funcional del grupo de 41 genes presentes en 5 o más enfermedades. Se muestran todos los procesos estadísticamente significativos (FDR<0,05), así como los genes implicados y los correspondientes p-valores asociados. FDR: *False Discovery Rate*.

El único gen que a través de DAVID no se encontró como involucrado en ningún proceso fue, al igual que con STRING, *GSTT1*.

#### 4.5. Validación de resultados

Se accedió a la base de datos GEO DataSets del NCBI para la búsqueda de datasets que validaran los resultados obtenidos para los genes más representativos. Para ello se utilizaron los siguientes filtros:

- *Study type: Expression profiling by array.*
- *Top organisms: Homo sapiens*
- *Entry type: datasets* (en este caso solo se habilitó para aquellos resultados de búsqueda en los que aparecía disponible dicha opción).
- Palabras clave:
  - *Male infertility*: estas palabras clave generaron resultados demasiado genéricos.
  - Nombre de las enfermedades estudiadas:
    - *Azoospermia*: se obtuvieron 4 estudios relacionados.
    - *Cryptorchidism*: se obtuvieron 2 estudios relacionados.
    - *Penile induration*: se obtuvo 1 estudio relacionado.
    - *Penile neoplasms*: se obtuvo 1 estudio relacionado.
    - *Prostatic hyperplasia*: se obtuvieron 3 estudios relacionados.
    - *Prostatic neoplasms*: se obtuvo 1 estudio relacionado.
    - *Testicular neoplasms*: se obtuvo 1 estudio relacionado.
- A la hora de seleccionar estudios se priorizaron aquellos en los que se obtenían biopsias por encima de los cultivos celulares.

A la hora de realizar el análisis con GEO2R la diferenciación en grupos de estudio se llevó a cabo de la misma forma que en el estudio seleccionado en cada caso y las opciones de análisis se mantuvieron en las ya seleccionadas por defecto. En cuanto a la selección de genes diferencialmente expresados, se escogieron aquellos que tenían un p-valor ajustado inferior a 0,05.

## 4.5.1 Datasets analizados

### 4.5.1.1. Azoospermia

En el primer estudio relacionado con la azoospermia se analizan perfiles de biopsias de testículos humanos con el objetivo de identificar genes que pudieran ser potenciales biomarcadores para el tratamiento de la azoospermia. Se analizaron 6 biopsias testiculares con microarrays de Affymetrix (*Affymetrix Human Gene 1.0 ST microarrays*), 3 de ellas pertenecientes a respondedores a la terapia hormonal y 3 no respondedores (número de referencia del dataset: [GSE51111](#)). En este estudio se realizó una selección de genes y exones. Solamente fue seleccionado el análisis genético y se pudo comprobar que algunos de los 41 genes incluidos en nuestro grupo de estudio coincidían con sus resultados, concretamente fueron: *HLA-DRB1*, *CYP17A1*, *HLA-DQB1* y *TLR6*.

En el segundo estudio relacionado con la azoospermia se analizaron biopsias de testículos de pacientes que padecían azoospermia idiopática no obstructiva (NOA), también con el objetivo de encontrar posibles biomarcadores que permitieran identificar etapas particulares en caso de una espermatogénesis deficiente. Se analizaron 31 muestras de biopsias con microarrays de Affymetrix (*Affymetrix Human Gene 1.0 ST microarrays*), 27 de las cuales procedían de pacientes con NOA y 4 de sujetos con una espermatogénesis normal (número de referencia del dataset: [GSE45885](#)) (Malcher *et al.*, 2013). De este estudio se obtuvieron las siguientes coincidencias con el listado seleccionado: *ACE*, *DICER1*, *ERCC1* y *SHBG*.

En el tercer estudio en el que se analiza la azoospermia se busca la evidencia del papel regulador que pueden desempeñar las citoquinas en el desarrollo gonadal masculino en casos de patologías. Para ello se estudian los niveles de expresión de las familias *IL-1*, *IL-6*, *IL-10*, *TNF*, *SCF* y *c-kit* en pacientes con NOA comparándolos con sujetos con espermatogénesis normal. Se analizaron biopsias de 16 pacientes con NOA y 4 con espermatogénesis normal con microarrays de Affymetrix (*Affymetrix Human Gene 1.0 ST microarrays*) (número de referencia del dataset: [GSE45887](#)) (Malcher *et al.*, 2013). En este caso, los genes compartidos con los resultados de dicho estudio son: *GSTM1*, *ESR1*, *ACE*, *DICER1*, *ERCC1*, *SHBG*, *TLR1* y *TLR6*. Cabe decir que sí aparecieron algunos de los genes pertenecientes a las familias que buscaban los investigadores, pero al no encontrarse como tal en la lista que se ha generado en este trabajo no se han considerado como genes compartidos.

En el cuarto y último estudio relacionado con la azoospermia se compararon biopsias procedentes de 4 pacientes con NOA y 4 pacientes con espermatogénesis normal. Se pudo ver que 414 genes se expresaban diferencialmente en el caso de los hombres infértiles. Las biopsias se analizaron con microarrays de Agilent (*Agilent-012097 Human 1A Microarray (V2) G4110B*) (número de referencia del dataset: [GSE9194](#)) (Dubé *et al.*, 2008).

En este caso no se encontraron genes con un p-valor que se ajustara a los parámetros establecidos.

#### **4.5.1.2. Criptorquidismo/ Criptorquidia**

En el primer estudio sobre el criptorquidismo se analiza la probabilidad de desarrollar azoospermia después de intervenciones quirúrgicas para tratar dicho desorden. Se obtuvieron 23 biopsias de 22 niños: 19 de testículos de 18 niños con criptorquidismo, y 4 de testículos contralaterales descendidos procedentes de pacientes con agenesia testicular. Se identificaron 483 genes que o no estaban o estaban expresados a la baja en el grupo de azoospermia en comparación con el grupo control, para ello se utilizaron microarrays de Affymetrix (*Affymetrix Human Genome U133 Plus 2.0 Array*) (número de referencia del dataset: [GSE25518](#)) (Hadziselimovic *et al.*, 2011). Los genes identificados compartidos con este estudio son: *GSTP1*, *COMT*, *DICER1*, *MTRR* y *TP53*.

En el segundo estudio se aisló ARN testicular procedente de biopsias de niños con criptorquidismo, diferenciándose en tres grupos: criptorquidismo unilateral, bilateral y control; los perfiles de expresión génica fueron analizados con microarrays de Affymetrix (*Affymetrix Human Genome U133 Plus 2.0 Array*) (número de referencia del dataset: [GSE16191](#)) (Hadziselimovic *et al.*, 2009). Los genes compartidos con la lista establecida son: *GSTP1*, *TGFB1*, *ACE*, *COMT*, *CY1A2*, *DICER1*, *MTRR* y *TP53*.

#### **4.5.1.3. Induración plástica del pene/ Enfermedad de La Peyronie**

La enfermedad de La Peyronie es una alteración consistente en la aparición de una placa de tejido fibroso cicatricial en la túnica albugínea del pene (Hauck & Weidner, 2001).

En el estudio analizado, se cultivaron células procedentes de tejido normal y de células con enfermedad de La Peyronie y se extrajo el ARN con el objetivo de investigar diferencias entre la placa “anormal” así como establecer una línea de expresión base en los tejidos normales para realizar comparaciones. Se cultivaron 6 placas en presencia de *TGFβ* y se

utilizaron microarrays de Affymetrix para llevar a cabo el análisis de expresión génica (*Affymetrix Human Gene 1.0 ST Array [transcript (gene) version]*) (número de referencia del dataset: [GSE97301](#)) (Mulhall *et al.*, 2002). Los genes comunes que se encontraron fueron: *GSTP1*, *IL6*, *ACE* y *HLA-A*.

#### 4.5.1.4. Hiperplasia prostática

En el primer estudio relacionado con la hiperplasia prostática se investiga la expresión diferencial de genes en el cáncer de próstata (PCa) y poder discriminar estos casos y los de hiperplasia prostática benigna (BPH). Para ello se extrajeron muestras de ARN procedentes de biopsias de 4 muestras de PCa y 4 de BPH y se analizaron con microarrays de Affymetrix (*Affymetrix Human Genome U133 Plus 2.0 Array*) (número de referencia del dataset: [GSE104749](#)) (Shan *et al.*, 2017). En este estudio no se obtuvo ningún gen compartido entre la lista generada en este trabajo y el estudio.

En el segundo estudio en el que se trata la hiperplasia prostática se analiza la expresión de genes procedentes de pacientes con BPH con y sin metilación del gen *SRD5A2*, ya que según la hipótesis que barajaban estos investigadores este gen interviene en la conversión de testosterona a dihidrotestosterona, lo que puede indicar una ruta alternativa para el crecimiento del tejido prostático. Se analizaron 22 muestras (12 de *SRD5A2*-metilado y 10 de *SRD5A2*-no metilado) con microarrays de Illumina (*Illumina HumanHT-12 V4.0 expression beadchip*) (número de referencia del dataset: [GSE101486](#)) (Wang *et al.*, 2017). En este caso tampoco se obtuvo ningún gen diferencialmente expresado compartido.

En el tercer caso de hiperplasia prostática se emplearon perfiles de expresión de cDNA completos, tanto de BPH como de PCa (en sujetos en los que se pensaba que estaban afectados por estas enfermedades debido a los valores obtenidos en las analíticas), y se analizaron con microarrays de la casa comercial Agilent (*Agilent-014850 Whole Human Genome Microarray 4x44K*). Las biopsias extraídas de los pacientes se dividieron en muestras de próstata normales y muestras de PCa (número de referencia del dataset: [GSE30994](#)) (Kim *et al.*, 2014). En este caso sí encontramos varios genes compartidos entre la lista de partida y los generados por el estudio: *ESR1*, *GSTT1*, *CYP3A4*, *MTHFR*, *SRD5A2*, *TNF*, *CFTR*, *CYP1B1*, *HFE*, *AHR*, *COMT*, *HLA-A*, *TGFBR3*, *TLR1*, *TLR2*, *TLR6* y *TP53*.

#### **4.5.1.5. Neoplasma prostático**

En el estudio relacionado con el neoplasma prostático se estudiaron alteraciones proteómicas y transcriptómicas, lo que reveló que existía un 48-64% de concordancia entre los niveles de proteína y transcritos. Es importante destacar que las alteraciones proteómicas diferenciales entre el cáncer de próstata metastásico y el cáncer de próstata localizado clínicamente que se mapearon de manera concordante con la transcripción de genes sirvieron como predictores del resultado clínico en el cáncer de próstata, así como en otros tumores sólidos. Se analizaron 13 muestras individuales de cáncer de próstata benigno, primario y metastásico y 6 muestras combinadas de tejidos de cáncer de próstata benigno, primario o metastásico con microarrays de Affymetrix (*Affymetrix Human Genome U133 Plus 2.0 Array*) (número de referencia del dataset: [GSE3325](#)) (Varambally *et al.*, 2005). Se obtuvieron coincidencias en la expresión diferencial de los siguientes genes de interés: *GSTM1*, *ESR1*, *GSTT1*, *HLA-DRB1*, *CYP1A1*, *ESR2*, *GSTP1*, *HLA-C*, *MTHFR*, *SRD52A*, *CFTR*, *CYP1B1*, *HFE*, *HLA-DQB1*, *IL6*, *COMT*, *CYP1A2*, *DICER1*, *ERCC1*, *HLA-B*, *IL1B*, *LTA*, *TLR2* y *TP53*.

#### **4.5.1.6. Neoplasma testicular**

En el caso del neoplasma testicular, el objetivo era determinar el patrón diferencial de expresión génica a lo largo de las regiones del epidídimo humano, para lo cual se llevó a cabo el análisis de secuencias de bases como tal en lugar de estudiar los genes que formaban dichas secuencias, por lo que no se pudo comparar con la lista de genes que ha servido de modelo para las comparaciones en los otros estudios. El análisis se llevó a cabo con microarrays de Agilent (*Agilent-012097 Human 1A Microarray (V2) G4110B*) (número de referencia del dataset: [GSE23812](#)) (Thimon *et al.*, 2007).

#### **4.5.1.7. Neoplasma peneal**

En el caso del neoplasma peneal, las herramientas de análisis del perfil de expresión génica se utilizaron para evaluar genes alterados relacionados con el estado de infección por HPV (virus del papiloma humano) con el fin de identificar marcadores moleculares potenciales en el cáncer de pene (PC). El ARN fue extraído de 39 muestras de PC frescas congeladas e hibridadas sobre microarrays de expresión génica. El conjunto perteneciente a muestras de pene normales consistía en el ARN total de cinco glándulas procedentes de una autopsia. Las alteraciones específicas de la expresión génica fueron investigadas por RT-qPCR. También se extrajo ADN de muestras de pene y se sometió al genotipado del HPV. Para el análisis de

los genes se utilizaron microarrays de Agilent (*Agilent-014850 Whole Human Genome Microarray 4x44K*) (número de referencia del dataset: [GSE57955](#)) (Kuasne *et al.*, 2015).

Se encontraron los siguientes genes de interés: *GSTP1*, *IL6*, *ACE*, *HLA-A*.

De los 41 genes que se intuía que podían intervenir como causantes de infertilidad masculina a través de otras enfermedades, se han podido validar 36 con los resultados de otras investigaciones. Los únicos genes sobre los que no se ha encontrado información han sido: *AR*, *CYP19A1*, *FSHR*, *IL10* y *NOS3*.

En la tabla 5 se puede observar un resumen de la información relativa a los estudios mencionados en los datasets estudiados. Por otro lado, en la tabla 6 se puede observar, también, otro resumen en referencia a los resultados obtenidos en comparación con los datasets analizados.

Enfermedad	Autores	Plataforma de array	Control	Muestras	Criterios de selección de los investigadores	Genes diferencialmente expresados más significativos	Umbral de selección seguido en el trabajo
Azoospermia*	<a href="#">Malcher et al. 2015</a>	Affymetrix Human Gene 1.0 ST microarrays	Control=3	n=3	N	N	pvalor ajustado < 0,05 FDR < 0,05
Azoospermia	<a href="#">Malcher et al. 2013</a>	Affymetrix Human Gene 1.0 ST microarrays	Control=4	n=27	p < 0,05 FDR < 0,05	4946	pvalor ajustado < 0,05 FDR < 0,05
Azoospermia	<a href="#">Malcher et al. 2013</a>	Affymetrix Human Gene 1.0 ST microarrays	Control=4	n=16	p < 0,05	3	pvalor ajustado < 0,05 FDR < 0,05
Azoospermia	<a href="#">Dubé et al., 2008</a>	Agilent-012097 Human 1A Microarray (V2) G4110B	Control=4	n=4	ANOVA de una cola p < 0,05	414	pvalor ajustado < 0,05 FDR < 0,05
Criptorquidia	<a href="#">Hadziselimovic et al., 2011</a>	Affymetrix Human Genome U133 Plus 2.0 Array	Control=4	n=18	p valor ajustado < 0,05 log2 fold change > 0,585	483/370**	pvalor ajustado < 0,05 FDR < 0,05
Criptorquidia	<a href="#">Hadziselimovic et al., 2009</a>	Affymetrix Human Genome U133 Plus 2.0 Array	Control=4	n=16	p valor ajustado < 0,05 log2 fold change > 0,585	432/578**	pvalor ajustado < 0,05 FDR < 0,05
Induración plástica del pene	<a href="#">Mulhall et al., 2002</a>	Affymetrix Human Gene 1.0 ST Array	Control=3	n=3	p < 0,05	N	pvalor ajustado < 0,05 FDR < 0,05
Neoplasma peneal	<a href="#">Kuasne et al., 2015</a>	Agilent-014850 Whole Human Genome Microarray 4x44K	Control=10	n=29	FDR < 0,05 pvalor ajustado < 0,05	1119	pvalor ajustado < 0,05 FDR < 0,05
Hiperplasia prostática	<a href="#">Shan et al., 2017</a>	Affymetrix Human Genome U133 Plus 2.0 Array	Control=4	n=4	p < 0,05 FDR < 0,05	255	pvalor ajustado < 0,05 FDR < 0,05
Hiperplasia prostática	<a href="#">Wang et al., 2017</a>	Illumina HumanHT-12 V4.0 expression beadchip	Control=12	n=10	p < 0,05 p < 0,01	N	pvalor ajustado < 0,05 FDR < 0,05
Hiperplasia prostática	<a href="#">Kim et al., 2014</a>	Agilent-014850 Whole Human Genome Microarray 4x44K	Control=3	n=3	ANOVA de dos colas p < 0,05 Student T-Test	N	pvalor ajustado < 0,05 FDR < 0,05
Neoplasma prostático	<a href="#">Varambally et al., 2005</a>	Affymetrix Human Genome U133 Plus 2.0 Array	Control=6	n=13	p < 0,05 - p < 0,001***	64	pvalor ajustado < 0,05 FDR < 0,05
Neoplasma testicular*	<a href="#">Thimon et al., 2007</a>	Agilent-012097 Human 1A Microarray (V2) G4110B	Control=4	n=4	N	N	pvalor ajustado < 0,05 FDR < 0,05
			ntotal= 225	n=65			
			n=160				
<p><i>N: Los autores no hacen referencia al número total de genes diferencialmente expresados pese a la utilización de microarrays y análisis estadísticos. El objetivo del estudio se centra en unos pocos genes y solo nombran aquellos de interés, sin embargo se puede extraer más información del dataset publicado.</i></p> <p><i>*: En el caso del primer dataset no fue utilizado para la publicación de un artículo por sí mismo, aunque está publicado y se puede acceder a él a través de la base de datos del NCBI. Para el último dataset no se han podido extraer datos debido a que el objeto de estudio eran exones y no genes.</i></p> <p><i>** : Utilizan varias comparaciones entre las situaciones control y problema y en función de las situaciones analizadas se halla un número de genes diferencialmente expresados distinto.</i></p> <p><i>***: En este estudio se utilizaron distintos p-valores que iban desde p &lt; 0,05 a p &lt; 0,001 en función de la comparación entre las distintas muestras problemas y el control.</i></p>							

Tabla 5. Resumen de la información obtenida de los datasets utilizados para realizar la comparación entre los genes obtenidos en nuestro trabajo y los resultados encontrados por otros investigadores. En la última columna se encuentran los parámetros estadísticos establecidos en el trabajo original para hacer la selección de genes.

Azoospermia				
Referencia	<a href="#">Malcher et al., 2015 GSE51111</a>	<a href="#">Malcher et al., 2013 GSE45885</a>	<a href="#">Malcher et al., 2013 GSE45887</a>	<a href="#">Dubé et al., 2008 GSE9194</a>
Genes compartidos	<i>HLA-DRB1, CYP17A1, HLA-DQB1, TLR6</i>	<i>ACE, DICER1, ERCC1, SHBG</i>	<i>GSTM1, ESR1, ACE, DICER1, ERCC1, SHBG, TLR1, TLR6</i>	N
Cryptorchidism				
Referencia	<a href="#">Hadziselimovic et al., 2011 GSE25518</a>	<a href="#">Hadziselimovic et al., 2009 GSE16191</a>		
Genes compartidos	<i>GSTP1, COMT, DICER1, MTRR, TP53</i>	<i>GSTP1, TGFB1, ACE, COMT, CY1A2, DICER1, MTRR, TP53</i>		
Penile induration				
Referencia	<a href="#">Hadziselimovic et al., 2009 GSE97301</a>			
Genes compartidos	<i>GSTP1, IL6, ACE, HLA-A</i>			
Penile neoplasms				
Referencia	<a href="#">Kuasne et al., 2015 GSE57955</a>			
Genes compartidos	<i>GSTP1, IL6, ACE, HLA-A</i>			
Prostatic hyperplasia				
Referencia	<a href="#">Hadziselimovic et al., 2009 GSE104749</a>	<a href="#">Wang et al., 2017 GSE101486</a>	<a href="#">Kim et al., 2014 GSE30994</a>	
Genes compartidos	N	N	<i>ESR1, GSTT1, CYP3A4, MTHFR, SRD5A2, TNF, CFTR, CYP1B1, HFE, AHR, COMT, HLA-A, TGFB3, TLR1, TLR2, TLR6, TP53</i>	
Prostatic neoplasms				
Referencia	<a href="#">Varambally et al., 2005 GSE3325</a>			
Genes compartidos	<i>GSTM1, ESR1, GSTT1, HLA-DRB1, CYP1A1, ESR2, GSTP1, HLA-C, MTHFR, SRD52A, CFTR, CYP1B1, HFE, HLA-DQB1, IL6, COMT, CYP1A2, DICER1, ERCC1, HLA-B, IL1B, LTA, TLR2, TP53</i>			
Testicular neoplasms				
Referencia	<a href="#">Hadziselimovic et al., 2009 GSE23812</a>			
Genes compartidos	N			

Tabla 6. Resultados de los estudios correspondientes a los datasets analizados y comparación con la lista de potenciales biomarcadores establecida en nuestro trabajo.

## 5. DISCUSIÓN

Como ya se ha visto anteriormente a lo largo de este Trabajo de Fin de Máster, los eventos por los que se llega a padecer infertilidad en los hombres se atribuyen a causas idiopáticas, debido en parte a la falta de información relacionada con las enfermedades que dan lugar a la infertilidad. Con el análisis de datos realizado en este Trabajo de Fin de Máster se pretende demostrar la utilidad de las tecnologías “-ómicas” para hallar nuevos posibles marcadores o procesos involucrados a nivel genómico en la infertilidad masculina.

Con el objetivo de encontrar uno o varios elementos comunes entre toda la información registrada en bases de datos públicas de las enfermedades relacionadas con la infertilidad masculina, se ha llevado a cabo el análisis de la información recabada a través de la construcción de redes complejas, mediante el uso de STRING y Cytoscape, y el tratamiento con RStudio.

El uso de estas herramientas nos ha permitido detectar algunos genes y procesos biológicos comunes entre las enfermedades estudiadas. Para encontrar genes relevantes a través del estudio de redes (en este caso de forma individual para cada enfermedad) se han analizado dos medidas de centralidad: la centralidad de grado y la intermediación (*degree centrality* y *betweenness*, respectivamente). Según los parámetros establecidos en nuestros análisis, hemos encontrado un gen que podemos considerar como altamente conectado por ajustarse a ellos: se trata del gen *HLA-C* (*Major Histocompatibility Complex, Class I, C*) en la red de interacción de la sífilis, que se encuentra involucrado en la presentación de antígenos exógenos al sistema inmune. Algunas rutas o procesos en los que se encuentra presente son: rechazo de injertos, presentación cruzada en el procesamiento de antígenos, infección en humanos por citomegalovirus, se encuentra presente de forma innata en el sistema inmune... Parece ser que pacientes con sífilis presentan alteraciones en la expresión del gen *HLA-C* junto con genes de la familia *KIR* (*Killer cell Immunoglobulin-like receptors*), ambos interaccionan e intervienen en la regulación de las células NK (*Natural Killers*) y algunas células T, aunque algunos autores recomiendan realizar más estudios para ver cuál de los dos complejos (*KIR* o *HLA-C*) tiene un papel dominante, en caso de que pueda establecerse (Zhuang *et al.*, 2012).

Los análisis conjuntos de todas las enfermedades en busca de genes comunes sí han arrojado más resultados. Por un lado, se estudió la aparición de un mismo gen en varias enfermedades y los grupos que formaban dichas enfermedades en base a los genes

compartidos a través de RStudio. Algunos de los que aparecen en un mayor número de enfermedades son:

- *AR (Androgen Receptor)*, factor de transcripción activado por ligando que regula la expresión génica eucariota y afecta a la proliferación y diferenciación celular en los tejidos diana. Algunas investigaciones demuestran que mutaciones y polimorfismos asociados al gen *AR* están significativamente asociados a una espermatogénesis deficiente y a la infertilidad masculina idiopática (Yong *et al.*, 2003). Hasta la fecha, se han documentado más de 300 mutaciones. Las mutaciones en *AR* que afectan severamente la cantidad, estructura o función de los receptores androgénicos causan el conocido síndrome completo A/S (síndrome de feminización testicular) (Yong *et al.*, 2003).
- *GSTM1 (Glutathione S-Transferase Mu 1)*, que interviene en los procesos de óxido-reducción exógenos y endógenos. Parece ser que hay consenso acerca de la relación de los genes de la familia *GST* (*GSTM1*, *GSTT1* (este lo encontramos también altamente representado en nuestras redes, pero no hemos encontrado información relacionada en *STRING* ni *DAVID*) y *GSTP1*) (Paracchini *et al.*, 2006; Safarinejad *et al.*, 2010; Salehi *et al.*, 2012; Vani *et al.*, 2010). Se ha observado que el aumento de la frecuencia de los genotipos *GSTM1* nulo y *GSTT1* nulo en pacientes infértiles en comparación con los controles indica que los polimorfismos *GSTM1* y *GSTT1* están asociados con un aumento del riesgo de infertilidad (Safarinejad *et al.*, 2010). Por otro lado, Finotti *et al.* encontraron que polimorfismos individuales en los genotipos asociados a *GSTM1* y *GSTP1* son susceptibles de disminuir la calidad del semen y causar infertilidad (Finotti *et al.*, 2009). Otro estudio que muestra que un genotipo *GSTM1* nulo afecta a la calidad del esperma, a la movilidad concretamente, es el de Vani *et al.*, 2010. Además de a la calidad del semen y al estrés oxidativo se ha establecido relación entre *GSTM1* y *MTHFR*, lo que podría afectar a las rutas metabólicas encargadas de la metilación, afectando a la fertilidad y a la salud cardiovascular e interviniendo en procesos tumorales, entre otros (Paracchini *et al.*, 2006).

De igual forma, se construyeron redes de interacción y, como se ha visto antes, gracias al estudio estadístico con RStudio y el análisis de redes esta vez de forma conjunta entre todas las enfermedades) se pudo elaborar una lista de 41 genes. Centrándonos en los parámetros con los que se estableció el filtro para el estudio de redes (*degree* y *betweenness*), algunos de los genes más relevantes en la red de interacción han sido:

- *IL6 (Interleukin 6)*, que realiza una gran cantidad de funciones a nivel biológico: es un potente inductor de la respuesta en fase aguda, desempeña un papel esencial en la diferenciación final de las células B en células secretoras de inmunoglobulinas implicadas en la diferenciación de linfocitos y monocitos, actúa sobre las células B, las células T, los hepatocitos, las células progenitoras hematopoyéticas y las células del SNC, es necesaria para la generación de células T(H)17, también actúa como una mioquina descargándose en el torrente sanguíneo después de la contracción muscular y actúa para aumentar la descomposición de las grasas y mejorar la resistencia a la insulina; por último, induce el crecimiento del mieloma y el plasmacitoma y la diferenciación de las células nerviosas. Dado que las citoquinas pueden actuar sobre varias células diferentes, pueden mostrar una superposición considerable en sus efectos biológicos y pueden inducir la síntesis de otras citoquinas, los resultados presentados por (Nallella *et al.*, 2004) pueden reflejar las intrincadas relaciones entre las diferentes citoquinas y su respuesta al estado inmunológico y endocrino genital subyacente, en lugar de ser la causa de la gravedad de la disfunción testicular.

En el caso de los pacientes infértiles con varicocele se puede apreciar que tienen mayores niveles de citoquinas y estrés oxidativo que los pacientes sanos, como lo indican los niveles elevados de IL-6 y ROS y la disminución de los niveles de capacidad antioxidante total. La citoquina pro-inflamatoria IL-6 y el estrés oxidativo pueden contribuir a la fisiopatología de la infertilidad en hombres con varicocele (Nallella *et al.*, 2004).

Por otro lado, se ha visto que las interleucinas (*IL-1* e *IL-6*) y *TNF- $\alpha$*  están íntimamente relacionadas con la acción del factor de crecimiento espermatogonial en las células de Sertoli y germinales (Lamb, 1993).

- *ESR1 (Estrogen Receptor 1)*, participa en la regulación de la expresión génica eucariótica y afecta a la proliferación y diferenciación celular en los tejidos diana. Con el paso del tiempo está quedando más en evidencia que los estrógenos juegan un papel vital en la capacidad reproductora humana (Giwercman, 2011). Otros meta-análisis sugieren que polimorfismos en *ERs (ESR1 y ESR2)* pueden tener diferentes papeles en la predisposición a la infertilidad masculina que varían en diferentes etnias (Ge *et al.*, 2014).
- *TP53 (Tumor Protein 53)*, actúa como un supresor tumoral en muchos tipos de tumores; induce la detención del crecimiento o apoptosis dependiendo de las circunstancias fisiológicas y el tipo de célula. La presencia de *p53* en los espermatozoides sugiere que desempeña un papel en la profase de la meiosis. *p53* se expresa en los testículos tanto en los espermatozoides como en las

espermatogonias. Esto sugiere que el gen *TP53* es importante para la regulación de la apoptosis durante la espermatogénesis y puede estar asociado con la infertilidad masculina (de Morais *et al.*, 2016).

Según otro estudio, los cambios en la expresión de *TP53* están asociados con la fase principal que regula la progresión meiótica, con un pico en el estadio de paquitene (Schwartz *et al.*, 1993).

- *TNF* (*Tumor Necrosis Factor*) es secretada principalmente por macrófagos y puede inducir la muerte celular de ciertas líneas celulares tumorales. Como ya se ha comentado antes, parece ser que *TNF* y algunos genes de la familia de las interleukinas están íntimamente relacionados, al menos en el caso de la infertilidad masculina. En un estudio reciente, se ha analizado si la viscosidad del líquido seminal estaba asociada con la producción de ROS, la concentración seminal de leucocitos y los niveles seminales de citoquinas pro-inflamatorias (*TNF-a* e *IL-6*) y una citoquina antiinflamatoria (*IL-10*), así como la influencia de las infecciones/inflamaciones (inducidas por la producción de ROS) en las glándulas accesorias masculinas, todo ello con resultados positivos (Castiglione *et al.*, 2014).

Como se puede observar, estos genes están directamente relacionados con la infertilidad masculina y no son solo una muestra de relación meramente estadística o sin ausencia total de pruebas de laboratorio. Prueba de ello es que las referencias descritas hacen alusión a investigaciones en las que se estudia el efecto de dichos genes en la infertilidad.

Por otro lado, no solo se identificaron genes cuya expresión está relacionada con la infertilidad sino los procesos biológicos a los que dan lugar la interacción entre dichos genes. Algunos de los procesos más significativos derivados de la red de interacción son: respuesta inmune, presentación y procesamiento de antígenos, regulación de la expresión génica y la transcripción, proliferación celular, procesos de óxido-reducción y respuesta a fármacos o compuestos xenobióticos.

También hemos de tener en cuenta que varios de los procesos y de los genes resultantes están íntimamente relacionados con procesos tumorales, algo bastante lógico ya que algunas de las enfermedades estudiadas son neoplasias y/o son producidas por alteraciones del sistema inmune. Esta relación también se ha establecido con anterioridad en otros estudios (Shan *et al.*, 2017; Thimon *et al.*, 2007; Varambally *et al.*, 2005).

El estrés oxidativo está cada vez cobrando más importancia en el proceso de espermatogénesis (Safarinejad *et al.*, 2010), lo que se ve reflejado en estudios que apoyan la hipótesis de que una espermatogénesis deficiente está asociada con un incremento del estrés oxidativo en el plasma seminal (Agarwal & Said, 2005; Pasqualotto *et al.*, 2000). Además, bajo

condiciones patológicas la producción de ROS excede la capacidad antioxidante del plasma seminal y aumenta el estrés oxidativo (Naughton *et al.*, 2001).

Los factores ambientales que afectan a la fertilidad masculina copan un amplio abanico de causas, desde la luz ultravioleta hasta la polución en el ambiente (Vargas *et al.*, 2018). Además de estas causas, hay otro tipo de exposiciones: exposiciones ocupacionales (asociadas al trabajo), tratamientos médicos, hormonas anabólicas o fármacos supresores del sistema inmune, que afectan a la fertilidad masculina (Sharpe, 2010). Estos factores causan alteraciones en los mecanismos de regulación genéticos y epigenéticos (Vargas *et al.*, 2018), dando lugar a distintas consecuencias entre las que se incluyen: morfología anormal y descenso del número de espermatozoides y volumen (Carlsen *et al.*, 1992; Irvine *et al.*, 1996; Swan *et al.*, 2014). Desde esta perspectiva parece crucial entender los mecanismos genéticos y epigenéticos en los que intervienen estos factores medioambientales y sustancias exógenas. En esta misma línea, recientemente, estudios *in vivo* han mostrado que algunas sustancias tóxicas podrían cambiar la expresión de algunos ARNm que juegan un papel importante en la producción de espermatozoides, en lo que a la regulación epigenética se refiere (Zhang & Pan, 2009).

Otros de los efectos epigenéticos que se pueden dar son cambios en la metilación del ADN, con el agravante de que este cambio puede ser heredado por los descendientes (Anway *et al.*, 2006; Wilhelm-Benartzi *et al.*, 2012).

Con el objetivo de confirmar si algunas de las conclusiones que se podían extraer de las funciones de los genes y de la relación establecida por los procesos biológicos eran ciertas se realizó, además de la revisión de otros estudios, la validación en datasets. Si bien con los datasets encontrados se pudo refutar la presencia de una gran mayoría de genes, variando los filtros y realizando una búsqueda aún más exhaustiva en lo que a la búsqueda de datasets se refiere, puede ser posible encontrar la totalidad de los genes de la lista.

Este estudio presenta algunas limitaciones inherentes al proceso mediante el cual se lleva a cabo. Por un lado, debido a la variabilidad derivada de trabajar con multitud de enfermedades simultáneamente no se han podido tener en cuenta parámetros, además de los ya utilizados (FDR y p-valor ajustado), como la tasa de cambio ( $F$ ).

Como hemos visto, la masa de datos generados por estas tecnologías de alto rendimiento está claramente subexplorada, lo que representa un gran reto para la investigación biomédica, ya que toda esta información necesita ser procesada y analizada cuidadosamente utilizando potentes aplicaciones bioinformáticas. El hecho es que hemos llegado a un cuello de botella en la gestión de datos: tenemos muchos más datos de los que somos capaces de analizar e interpretar plenamente. De hecho, millones de conjuntos de

datos de expresión génica están disponibles públicamente en repositorios (por ejemplo, Gene Expression Omnibus, GEO y ArrayExpress), de donde se extraen datos para compararlos con las investigaciones que se estén realizando de forma paralela a nivel de laboratorio (Baker, 2012).

Además, es necesaria la aplicación de experimentos en el laboratorio para corroborar algunas de las hipótesis que se pueden extraer de este tipo de estudios.

## 6. CONCLUSIONES

1. Las herramientas y análisis bioinformáticos permiten extraer información relevante de las investigaciones llevadas a cabo en el laboratorio y, además, pueden ayudar a detectar posibles indicios para validar nuevas hipótesis.

2. Los genes considerados como relevantes dentro del conjunto de enfermedades relacionadas con la infertilidad masculina muestran que los procesos biológicos más importantes que pueden dar lugar a esta enfermedad son: respuesta inmune, presentación y procesamiento de antígenos, regulación de la expresión génica y la transcripción, proliferación celular, procesos de óxido-reducción y respuesta a fármacos o compuestos xenobióticos.

3. Sería conveniente analizar de forma más profunda el gen *GSTT1* y sus interacciones utilizando otras bases de datos, ya que parece que está íntimamente relacionado con la infertilidad masculina, al igual que el resto de genes de la familia *GST*.

4. Con la validación con otro tipo de estudios complementarios podría confirmarse la expresión diferencial de dichos genes como algunas de las causas endógenas de la infertilidad masculina.

5. Sería interesante hacer hincapié en estudios epigenéticos para comprobar si la aparición de la infertilidad masculina parece ser una dolencia mayormente causada por déficits en el organismo del hombre de forma congénita en lugar de que dichos déficits puedan ser debidos a causas exógenas.

## 7. BIBLIOGRAFÍA

- Acott, S., Carr, D. W., Oregon, M. S. (1984). Inhibition of Bovine Spermatozoa by Caudal Epididymal II. Interaction of pH and a Quiescence. *Biology of Reproduction*, 30(4), 926–935. <https://doi.org/10.1095/biolreprod30.4.926>
- Agarwal, A., Mulgund, A., Hamada, A., Chyatte, M. R. (2015). A unique view on male infertility around the globe. *Reproductive Biology and Endocrinology*, 13(1), 1–9. <https://doi.org/10.1186/s12958-015-0032-1>
- Agarwal, A., Said, T. M. (2005). Oxidative stress, DNA damage and apoptosis in male infertility: a clinical approach. *BJU International*, 95, 403-507. <https://doi.org/10.1111/j.1464-410X.2005.05328.x>
- Amann, R. P. (2008). The cycle of the seminiferous epithelium in humans: A need to revisit? *Journal of Andrology*, 29(5), 469–487. <https://doi.org/10.2164/jandrol.107.004655>
- Anway, M. D., Leathers, C., Skinner, M. K. (2006). Endocrine disruptor vinclozolin induced epigenetic transgenerational adult-onset disease, *Endocrinology*, 147(12), 5515–5523. <https://doi.org/10.1210/en.2006-0640>
- Aston, K. I., Carrell, D. T. (2009). Genome-wide study of single-nucleotide polymorphisms associated with azoospermia and severe oligozoospermia. *Journal of Andrology*, 30(6), 711–725. <https://doi.org/10.2164/jandrol.109.007971>
- Aston, K. I., Conrad, D. F. (2013). Spermatogenesis: Methods and Protocols, Methods in Molecular Biology, vol. 927, [https://link.springer.com/protocol/10.1007%2F978-1-62703-038-0\\_34](https://link.springer.com/protocol/10.1007%2F978-1-62703-038-0_34)
- Atala, A. (2015). Re: Imbalanced Lipid Homeostasis in the Conditional Dicer1 Knockout Mouse Epididymis Causes Instability of the Sperm Membrane. *The Journal of Urology*, 194(4), 1162–1163. <https://doi.org/10.1016/j.juro.2015.07.012>
- Baker, M. (2012). Gene data to hit milestone. *Nature*, 487(7407), 282–283. <https://doi.org/10.1038/487282a>
- Bhagwat, S. R., Chandrashekar, D. S., Kakar, R., Davuluri, S., Bajpai, A. K., Nayak, S., Bhutada, S., Acharya, K., Sachdeva, G. (2013). Endometrial Receptivity: A Revisit to Functional Genomics Studies on Human Endometrium and Creation of HGEx-ERdb. *PLoS ONE*, 8(3), e58419 . <https://doi.org/10.1371/journal.pone.0058419>
- Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., Hwang, DU. (2006) Complex networks: structure and dynamics. *Physics Reports*, 424(4-5), 175-308.

<https://doi.org/10.1016/j.physrep.2005.10.009>

- Carlsen, E., Giwercman, A., Keiding, N., Skakkeblek, N. E., Skakkebsk, N. E. (1992). Evidence for decreasing quality of semen during past 50 years, *British Medical Journal*, 305, 609–613. <https://doi.org/10.1136/bmj.305.6854.609>
- Carrell, D. T., Aston, K. I., Oliva, R., Emery, B. R., De Jonge, C. J. (2016). The “omics” of human male infertility: integrating big data in a systems biology approach. *Cell and Tissue Research*, 363(1), 295–312. <https://doi.org/10.1007/s00441-015-2320-7>
- Castiglione, R., Salemi, M., Vicari, L. O., Vicari, E. (2014). Relationship of semen hyperviscosity with IL-6, TNF- $\alpha$ , IL-10 and ROS production in seminal plasma of infertile patients with prostatitis and prostatic-vesiculitis. *Andrologia*, 46(10), 1148–1155. <https://doi.org/10.1111/and.12207>
- Chihara, M., Yoshihara, K., Ishiguro, T., Yokota, Y., Adachi, S., Okada, H., Kashima, K., Sato, T., Tanaka, A., Tanaka, H., Enomoto, T. (2015). Susceptibility to male infertility: Replication study in Japanese men looking for an association with four GWAS-derived loci identified in European men. *Journal of Assisted Reproduction and Genetics*, 32(6), 903–908. <https://doi.org/10.1007/s10815-015-0468-4>
- Cooper, T. G. (2005). Cytoplasmic droplets: The good, the bad or just confusing? *Human Reproduction*, 20(1), 9–11. <https://doi.org/10.1093/humrep/deh555>
- Dalgaard, M. D., Weinhold, N., Edsgård, D., Silver, J. D., Pers, T. H., Nielsen, J. E., Jørgensen, N., Juul, A., Gerds, T.A., Giwercman, A., Giwercman, L.Y., Cohn-Cedermark, G., Virtanen, H. E., Toppari, J., Dugaard, G., Jensen, T.S., Brunak, S., Rajpert-De Metys, E., Skakkebak, N. E., Leffers, H., Gupta, R. (2012). A genome-wide association study of men with symptoms of testicular dysgenesis syndrome and its network biology interpretation. *Journal of Medical Genetics*, 49(1), 58–65. <https://doi.org/10.1136/jmedgenet-2011-100174>
- Dam, A. H. D. M., Kosciński, I., Kremer, J. A. M., Moutou, C., Jaeger, A.-S., Oudakker, A. R., Tournelle, H., Charlet, N., Lagier-Tourenne, C., van Bokhoven, H., Viville, S. (2007). Homozygous Mutation in SPATA16 Is Associated with Male Infertility in Human Globozoospermia. *The American Journal of Human Genetics*, 81(4), 813–820. <https://doi.org/10.1086/521314>
- Datta, S., Gill, R. (2015). Statistical Analysis of DNA Microarray Data. *Wiley StatsRef: Statistics Reference Online*, 1–7. <https://doi.org/10.1002/9781118445112.stat00177.pub2>
- Diaz-Beltran, L., Esteban, F. J., Varma, M., Ortuzk, A., David, M., Wall, D. P. (2017). Cross-

- disorder comparative analysis of comorbid conditions reveals novel autism candidate genes. *BMC Genomics*, 18(1), 1–14. <https://doi.org/10.1186/s12864-017-3667-9>
- de Moraes, M. P., Curado, R. F., e Silva, K. S. F., Moura, K. K. V. O., Arruda, J. T. (2016). Male idiopathic infertility and the TP53 polymorphism in codon 72. *Genetics and Molecular Research*, 15(4). <https://doi.org/10.4238/gmr15048882>
- Dubé, E., Hermo, L., Chan, P. T. K., Cyr, D. G. (2008). Alterations in Gene Expression in the Caput Epididymides of Nonobstructive Azoospermic Men. *Biology of Reproduction*, 78(2), 342–351. <https://doi.org/10.1095/biolreprod.107.062760>
- Ferlin, A., Raicu, F., Gatta, V., Zuccarello, D., Palka, G., Foresta, C. (2007). Male infertility: Role of genetic background. *Reproductive BioMedicine Online*, 14(6), 734–745. [https://doi.org/10.1016/S1472-6483\(10\)60677-3](https://doi.org/10.1016/S1472-6483(10)60677-3)
- Gracie, S., Pennell, C., Ekman-Ordeberg, G., Lye, S., McManaman, J., Williams, S., Palmer, L., Kelley, M, Menon, R., Gravett, M. (2011). An integrated systems biology approach to the study of preterm birth using “-omic” technology - a guideline for research. *BMC Pregnancy and Childbirth*, 11(71). <https://doi.org/10.1186/1471-2393-11-71>
- Ge, Y. Z., Xu, L. W., Jia, R. P., Xu, Z., Li, W. C., Wu, R., Sheng, L., Gao, F., Tan, S-J., Son, Q., Xin, H. (2014). Association of polymorphisms in estrogen receptors (ESR1 and ESR2) with male infertility: A meta-analysis and systematic review. *Journal of Assisted Reproduction and Genetics*, 31(5), 601–611. <https://doi.org/10.1007/s10815-014-0212-5>
- Giwerzman, A. (2011). Estrogens and phytoestrogens in male infertility. *Current Opinion in Urology*, 21(6), 519–526. <https://doi.org/10.1097/MOU.0b013e32834b7e7c>
- Hadziselimovic, F., Hadziselimovic, N. O., Demougin, P., Krey, G., Hoecht, B., Oakeley, E. J. (2009). EGR4 is a master gene responsible for fertility in cryptorchidism. *Sexual Development*, 3(5), 253–263. <https://doi.org/10.1159/000249147>
- Hadziselimovic, F., Hadziselimovic, N. O., Demougin, P., Oakeley, E. J. (2011). Testicular gene expression in cryptorchid boys at risk of azoospermia. *Sexual Development*, 5(2), 49–59. <https://doi.org/10.1159/000323955>
- Hauck, E. W., Weidner, W. (2001). François de la Peyronie and the disease named after him. *Lancet*, 357(9273), 2049–2051. [https://doi.org/10.1016/S0140-6736\(00\)05119-9](https://doi.org/10.1016/S0140-6736(00)05119-9)
- Hennig, C. (2007). Cluster-wise assessment of cluster stability. *Computational Statistics & Data Analysis* 52, 258–271. <https://doi.org/10.1016/j.csda.2006.11.025>
- Hua, J., Xu, B., Yang, Y., Ban, R., Iqbal, F., Cooke, H. J., Zhang, Y., Shi, Q. (2015). Folicle

- Online: An integrated database of follicle assembly, development and ovulation. *Database (Oxford)*, 2015, bav036. <https://doi.org/10.1093/database/bav036>
- Huang, D. W., Sherman, B. T., Lempicki, R. A. (2009a). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Research*, 37(1), 1–13. <https://doi.org/10.1093/nar/gkn923>
- Huang, D. W., Sherman, B. T., Lempicki, R. A. (2009b). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols*, 4(1), 44–57. <https://doi.org/10.1038/nprot.2008.211>
- Irvine, S., Cawood, E., Richardson, D., Macdonald, E., Aitken, J. (1996). Evidence of deteriorating semen quality in the United Kingdom: birth cohort study in 577 men in Scotland over 11 years, *British Medical Journal*, 312, 467–471. <https://doi.org/10.1136/bmj.312.7029.467>
- Jodar, M., Soler-Ventura, A., Oliva, R. (2017). Semen proteomics and male infertility. *Journal of Proteomics*, 162, 125–134. <https://doi.org/10.1016/j.jprot.2016.08.018>
- Kim, J., Mizokami, A., Shin, M., Izumi, K., Konaka, H., Kadono, Y., Kitagawa, Y., Keller, E., Zhang, J., Namiki, M. (2014). SOD3 Acts as a Tumor Suppressor in PC-3 Prostate Cancer Cells via Hydrogen Peroxide Accumulation. *Anticancer Research* 2832, 2821–2831. <http://ar.iijournals.org/content/34/6/2821.long>
- Kim, M.S., Pinto, S. M., Getnet, D., Nirujogi, R. S., Manda, S. S., Chaerkady, R., Madugundu, Kelkar, D. S., Isserlin, R., Jain, S., Thomas, J. K., Muthusamy, B., Leal-Rojas, P., Kumar, P., Sahasrabudhe, N. A., Balakrishnan, L., Advani, J., George, B., Renuse, S., Selvan, L. D., Patil, A. H., Nanjappa, V., Radhakrishnan, A., Prasad, S., Subbannayya, T., Raju, R., Kumar, M., Sreenivasamurthy, S.K., Marimuthu, A., Sathe, G. J., Chavan, S., Datta, K. K., Subbannayya, Y., Sahu, A., Yelamanchi, S. D., Jayaram, S., Rajagopalan, P., Sharma, J., Murthy, K. R., Syed, N., Goel, R., Khan, A. A., Ahmad, S., Dey, G., Mudgal, K., Chatterjee, A., Huang, T. C., Zhong, J., Wu, X., Shaw, P. G., Freed, D., Zahari, M. S., Mukherjee, K. K., Shankar, S., Mahadevan, A., Lam, H., Mitchell, C. J., Shankar, S. K., Satishchandra, P., Schroeder, J. T., Sirdeshmukh, R., Maitra, A., Leach, S. D., Drake, C. G., Halushka, M. K., Prasad, T. S., Hruban R. H., Kerr, C. L., Bader, G. D., Iacobuzio-Donahue, C. A., Gowda, H., Pandey, A. (2014). A draft map of the human proteome. *Nature*, 509(7502), 575-581. <https://doi.org/10.1038/nature13302>
- Kosova, G., Scott, N. M., Niederberger, C., Prins, G. S., Ober, C. (2012). Genome-wide association study identifies candidate genes for male fertility traits in humans. *American Journal of Human Genetics*, 90(6), 950–961. <https://doi.org/10.1016/j.ajhg.2012.04.016>

- Krausz, C., Escamilla, A. R., Chianese, C. (2015). Genetics of male infertility: From research to clinic. *Reproduction*, 150(5), R159–R174. <https://doi.org/10.1530/REP-15-0261>
- Krausz, C., Giachini, C., Lo Giacco, D., Daguin, F., Chianese, C., Ars, E., Ruiz-Castane, E., Forti, G., Rossi, E. (2012). High Resolution X Chromosome-Specific Array-CGH Detects New CNVs in Infertile Males. *PLoS ONE*, 7(10), e44887. <https://doi.org/10.1371/journal.pone.0044887>
- Kuasne, H., de Syllos Cólus, I. M., Busso, A. F., Hernandez-Vargas, H., Barros-Filho, M. C., Marchi, F. A., Scapulatempo-Neto, C., Faria, E. F., Lopes, A., Guimarães, G. C., Herceg, Z., Rogatto, S. R. (2015). Genome-wide methylation and transcriptome analysis in penile carcinoma: Uncovering new molecular markers. *Clinical Epigenetics*, 7, 46. <https://doi.org/10.1186/s13148-015-0082-4>
- Lamb, D. J. (1993). Paracrine factors and the regulation of spermatogenesis. *World Journal of Urology*, 11 (2), 120–128. <https://link.springer.com/article/10.1007/BF00182039>
- Lindsay, T. J., Vitrikas, K. R. (2015). Evaluation and treatment of infertility. *American Family Physician*, 91(5), 308–314. <https://doi.org/d11839>
- Malcher, A., Rozwadowska, N., Stokowy, T., Jedrzejczak, P., Zietkowiak, W., Kurpisz, M. (2013). The Gene Expression Analysis of Paracrine/Autocrine Factors in Patients with Spermatogenetic Failure Compared with Normal Spermatogenesis. *American Journal of Reproductive Immunology*, 70(6), 522–528. <https://doi.org/10.1111/aji.12149>
- Malcher, A., Rozwadowska, N., Stokowy, T., Kolanowski, T., Jedrzejczak, P., Zietkowiak, W., Kurpisz, M. (2013). Potential biomarkers of nonobstructive azoospermia identified in microarray gene expression analysis. *Fertility and Sterility*, 100(6), 1686-1694. <https://doi.org/10.1016/j.fertnstert.2013.07.1999>
- Nallella, K. P., Allamaneni, S. S. R., Pasqualotto, F. F., Sharma, R. K., Thomas, A. J., Agarwal, A. (2004). Relationship of interleukin-6 with semen characteristics and oxidative stress in patients with varicocele. *Urology*, 64(5), 1010–1013. <https://doi.org/10.1016/j.urology.2004.05.045>
- Nanjappa, V., Thomas, J. K., Marimuthu, A., Muthusamy, B., Radhakrishnan, A., Sharma, R., Ahmad, A., Balakrishnan, L., Sahasrabudhe, N., Kumar, S., Jhaveri, B., Sheth, K., Khatana, R., Shaw, P., Srinkath, S., Mathur, P., Shankar, S., Nagaraja, D., Christopher, R., Mathivanan, S., Raju, R., Sierdeshmukh, R., Chatterjee, A., Simpson, R., Harsha, H.C., Pandey, A., Prasad, T. S. K. (2014). Plasma Proteome Database as a resource for proteomics research: 2014 update. *Nucleic Acids Research*, 42(D1), 959–965.

<https://doi.org/10.1093/nar/gkt1251>

- Naughton, C. K., Nangia, A. K., Agarwal, A. (2001). Varicocele and male infertility : Part II Pathophysiology of varicoceles in male infertility, *Human Reproduction Update*, 7(5), 473–481. <https://doi.org/10.1093/humupd/7.5.473>
- Neto, F. T. L., Bach, P. V., Najari, B. B., Li, P. S., Goldstein, M. (2016). Spermatogenesis in humans and its affecting factors. *Seminars in Cell and Developmental Biology*, 59, 10–26. <https://doi.org/10.1016/j.semcd.2016.04.009>
- Oksanen, A. J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., Mcglinn, D., Peter, M., O'Hara, R. B., Simpson, G., Solymos, P., Henry, M., Stevens, H., Szoecs, E., Wagner, H. (2019). Package 'vegan.' <https://github.com/vegandevs/vegan>
- Paracchini, V., Garte, S., Taioli, E. (2006). MTHFR C677T polymorphism, GSTM1 deletion and male infertility: A possible suggestion of a gene-gene interaction? *Biomarkers*, 11(1), 53–60. <https://doi.org/10.1080/13547500500442050>
- Parkinson, H., Kapushesky, M., Shojatalab, M., Abeygunawardena, N., Coulson, R., Farne, A., Holloway, E., Kolesnykov, N., Lilja, P, Lukk, M., Mani, R., Rayner, T., Sharma, A., William, E., Sarkans, U., Brazma, A. (2007). ArrayExpress - A public database of microarray experiments and gene expression profiles. *Nucleic Acids Research*, 35(Suppl. 1), 747–750. <https://doi.org/10.1093/nar/gkl995>
- Pasqualotto, F. F., Sharma, R. K., Nelson, D. R., Thomas, A. J., Agarwal, A. (2000). Relationship between oxidative stress, semen characteristics, and clinical diagnosis in men undergoing infertility investigation, *Fertility and Sterility*, 73(3), 459–464. [https://doi.org/10.1016/S0015-0282\(99\)00567-1](https://doi.org/10.1016/S0015-0282(99)00567-1)
- Mulhall, J. P., Anderson, M. S., Lubrano, T., Shankey, T. V. (2002). Peyronie's disease cell culture models: Phenotypic, genotypic and functional analyses. *International Journal of Impotence Research*, 14(5), 397–405. <https://doi.org/10.1038/sj.ijir.3900874>
- Poongothai, J., Gopenath, T. S., Manonayaki, S. (2009). Genetics of human male infertility. *Singapore Med J*, 50(4), 336–347. Retrieved from [http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list\\_uids=19421675](http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=19421675)
- Safarinejad, M. R., Shafiei, N., Safarinejad, S. (2010). The association of glutathione-S-transferase gene polymorphisms (GSTM1, GSTT1, GSTP1) with idiopathic male infertility. *Journal of Human Genetics*, 55(9), 565–570. <https://doi.org/10.1038/jhg.2010.59>

- Salehi, Z., Gholizadeh, L., Vaziri, H., Madani, A. H. (2012). Analysis of GSTM1, GSTT1, and CYP1A1 in idiopathic male infertility. *Reproductive Sciences*, 19(1), 81–85. <https://doi.org/10.1177/1933719111413302>
- Schwartz, D., Goldfinger, N., Rotter, V. (1993). Expression of p53 protein in spermatogenesis is confined to the tetraploid pachytene primary spermatocytes. *Oncogene*, 8(6), 1487–1494. <https://europepmc.org/abstract/med/8502474>
- Shan, M., Xia, Q., Yan, D., Zhu, Y., Zhang, X., Zhang, G., Guo, J., Hou, J., Chen, W., Zhu, T., Xiaoyan, Z., Xu, J., Wang, J., Ding, T., Zheng, J. (2017). Molecular analyses of prostate tumors for diagnosis of malignancy on fine-needle aspiration biopsies. *Oncotarget*, 8(62), 104761–104771. <https://doi.org/10.18632/oncotarget.22289>
- Sharlip, I. D., Jarow, J. P., Belker, A. M., Lipshultz, L. I., Sigman, M., Thomas, A. J., Schlegel, P. N., Howard, S. S., Nehra, A., Damewood, M.D., Overstreet, J. W., Sadovsky, R. (2002). Best practice policies for male infertility. *Fertility and Sterility*, 77(5), 873–882. [https://doi.org/10.1016/S0015-0282\(02\)03105-9](https://doi.org/10.1016/S0015-0282(02)03105-9)
- Sharma, U., Conine, C. C., Shea, J. M., Boskovic, A., Derr, A. G., Bing, X. Y., Belleanne, C., Kucukural, A., Serra, R. W., Sun, F., Song, L., Carone, B. R., Ricci, E. P., Li, X. Z., Lucas, F., Moore, M. J., Sullivan, R., Mello, C. C., Garber, M., Rando, O. J. (2016). Biogenesis and function of tRNA fragments during sperm maturation and fertilization in mammals. *Science*, 351(6271), 391–396. <https://doi.org/10.1126/science.aad6780>
- Sharpe, R. M. (2010). Environmental/lifestyle effects on spermatogenesis, 1697–1712. <https://doi.org/10.1098/rstb.2009.0206>
- Sullivan, R., Saez, F., Girouard, J., Frenette, G. (2005). Role of exosomes in sperm maturation during the transit along the male reproductive tract. *Blood Cells, Molecules, and Diseases*, 35(1), 1–10. <https://doi.org/10.1016/j.bcmd.2005.03.005>
- Swan, S. H., Elkin, E. P., Fenster, L., Swan, S. H., Elkin, E. P., Fenster, L. (2014). Have Sperm Densities Declined? A Reanalysis of Global Trend Data. *Environmental Health Perspective*, 105(11), 1228–1232.
- Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N. T., Morris, J. H., Bork, P., Jensen, L. J., Mering, C. V. (2019). STRING v11 : protein – protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets, *Nucleic Acids Research*, 47(D1), 607–613. <https://doi.org/10.1093/nar/gky1131>

- Thimon, V., Koukoui, O., Calvo, E., Sullivan, R. (2007). Region-specific gene expression profiling along the human epididymis. *Molecular Human Reproduction*, 13(10), 691–704. <https://doi.org/10.1093/molehr/gam051>
- Travis, A. J., Kopf, G. S. (2002). The role of cholesterol efflux in regulating the fertilization potential of mammalian spermatozoa. *Journal of Clinical Investigation*, 110(6), 731–736. <https://doi.org/10.1172/JCI0216392>
- Varambally, S., Yu, J., Laxman, B., Rhodes, D. R., Mehra, R., Tomlins, S. A., Shah, R. B., Chandran, U., Monzon, F. A., Becich, M. J., Wei, J. T., Kenneth, J. P., Ghosh, D., Rubin, M. A., Chinnaiyan, A. M. (2005). Integrative genomic and proteomic analysis of prostate cancer reveals signatures of metastatic progression. *Cancer Cell*, 8(5), 393–406. <https://doi.org/10.1016/j.ccr.2005.10.001>
- Vargas Liébanas, E. (2015). Trabajo Fin de Máster: Análisis De Datos De Receptividad Endometrial Mediante Biología De Sistemas. Universidad de Jaén.
- Vargas, E., Díaz-Beltrán, L., Esteban, F.J. (2019). Integración y análisis de datos mediante biología de sistemas. *Biología Molecular y Celular, Volumen I “Técnicas y fundamentos”*. ISBN 978-84-9159-208-2 , 43-72.
- Vargas, E., Esteban, F.J., Altmäe, S. (2018). Reproductomics: The Omics revolution and its impact on Human Reproductive Medicine. *Academic Press*, 347-383.
- Vani, G. T., Mukesh, N., Siva Prasad, B., Rama Devi, P., Hema Prasad, M., Rani, P. U., Reddy, P. P. (2010). Role of glutathione S-transferase Mu-1 (GSTM1) polymorphism in oligospermic infertile males. *Andrologia*, 42(4), 213–217. <https://doi.org/10.1111/j.1439-0272.2009.00971.x>
- Wagner, J. R., Busche, S., Ge, B., Kwan, T., Pastinen, T., Blanchette, M. (2014). The relationship between DNA methylation, genetic and expression inter. *Genome Biology*, 15(2), R37. <https://doi.org/10.1186/gb-2014-15-2-r37>
- Wang, Z., Hu, L., Salari, K., Bechis, S., Ge, R., Wu, S., Rassoulilian, C., Pham, J., Chin-lee, W., Tabatabaei, S., Strand, D. W., Olumi, A. F. (2017). Androgenic to oestrogenic switch in the human adult prostate gland is regulated by epigenetic silencing of steroid 5 $\alpha$ -reductase 2. *Journal of Pathology*, 243(4), 457–467. <https://doi.org/10.1002/path.4985>
- Warde-Farley, D., Donaldson, S. L., Comes, O., Zuberi, K., Badrawi, R., Chao, P., Franz, M., Grouios, C., Kazi, F., Tannus-Lopez, C., Maitland, A., Mostafavi, S., Montojo, J., Shao, Q., Wright, G., Bader, G. D., Morris, Q. (2010). The GeneMANIA prediction server:

- Biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Research*, 38(Suppl. 2), 214–220. <https://doi.org/10.1093/nar/gkq537>
- Wilhelm-Benartzi, C. S., Houseman, E. A., Maccani, M. A., Poage, G. M., Koestler, D. C., Langevin, S. M., Gagne, L. A., Banister, C.E., Marsit, C.J., Padbury, J. F. (2012). In Utero Exposures, Infant Growth and DNA Methylation of Repetitive Elements and Developmentally Related Genes in Human Placenta, *Environmental Health Perspectives*, 120(2), 296–302. <https://doi.org/10.1289/ehp.1103927>
- Yong, E. L., Loy, C. J., Sim, K. S. (2003). Androgen receptor gene and male infertility. *Human Reproduction Update*, 9(1), 1–7. <https://doi.org/10.1093/humupd/dmg003>
- Yu, W., Clyne, M., Khoury, M. J., Gwinn, M. (2009). Phenopedia and genopedia: Disease-centered and gene-centered views of the evolving knowledge of human genetic associations. *Bioinformatics*, 26(1), 145–146. <https://doi.org/10.1093/bioinformatics/btp618>
- Zhang, B., Pan, X. (2009). RDX Induces Aberrant Expression of MicroRNAs in Mouse Brain and Liver, *Environmental Health Perspectives*, 117(2), 231–240. <https://doi.org/10.1289/ehp.11841>
- Zorrilla, M., Yatsenko, A. N. (2013). The Genetics of Infertility: Current Status of the Field. *Current Genetic Medicine Reports*, 1(4), 247–260. <https://doi.org/10.1007/s40142-013-0027-1>
- Zhuang, Y. L., Ren, G. J., Tian, K. L., Li, X. Y., Zhu, Y. B., Liu, J. L., Si, G. L., Peng, L., Zhang, Y., Wang, L., Zhang, W. J., Wang, D. J., Zhu, C. F. (2012). Human leukocyte antigen-C and killer cell immunoglobulin-like receptor gene polymorphisms among patients with syphilis in a Chinese Han population. *Apmis*, 120(10), 828–835. <https://doi.org/10.1111/j.1600-0463.2012.02911.x>



## JUAN CARLOS ORTEGA VELASCO

GRADUADO EN BIOLOGÍA

### CARTA DE PRESENTACIÓN

Como demuestra mi formación académica siempre me he sentido atraído por las ciencias experimentales y su aplicación en el campo de la clínica. A mi "tradicional" atracción tuve que sumarle la curiosidad que despertaron más adelante las ciencias tecnológicas y que en su conjunto dieron como resultado el interés profesional que poseo en estos instantes.

Me considero una persona curiosa, versátil y con ganas de aprender antes que cualquier otro rasgo a destacar. También considero que destaco trabajando en equipo, en actividades relacionadas con la oratoria y liderando grupos de trabajo, aunque sí he de decir que a veces soy algo desorganizado si realizo tareas que me incumben solo a mí.

Entre mis intereses caben la literatura, el deporte (ámbito en el que también tengo formación profesional) los juegos de mesa, los videojuegos, el cine y las series.

Actualmente soy beneficiario del plan nacional de "Garantía juvenil".

### INFORMACIÓN PERSONAL

Fecha de nacimiento: 27/11/1993  
Teléfono: 601154376  
Email: [juan93ka@gmail.com](mailto:juan93ka@gmail.com)  
LinkedIn: <https://goo.gl/kV9XIC>  
Permiso de conducir: Clase B1

### FORMACIÓN ACADÉMICA

#### UNIVERSIDAD DE JAÉN

*Graduado en Biología (2017)*

- TFG: "Compuestos Metálicos con Actividad Antitumoral"

*Máster en Biotecnología y Biomedicina (2019)*

- TFM: Análisis y expresión de datos de expresión génica en enfermedades que causan infertilidad masculina.

#### UNIVERSITAT OBERTA DE CATALUNYA (UOC)

*Cursando el Máster en Bionformática y Bioestadística (Actualmente)*

### EMPRENDIMIENTO

*Participación en el programa "Explorer, Jóvenes con ideas" (Banco Santander)*

*Participación en el programa "Go to work. Espacios Coworking" (Escuela de Organización Industrial)*

### CURSOS Y OTRA FORMACIÓN

- Curso nutrición, genes y salud (Universidad de Navarra)
- Participación en la 4ª jornada de ciencias (Universidad de Jaén)
- Participación en el certamen de energías renovables en el Parque de las Ciencias de Granada

### EXPERIENCIA LABORAL

*Prácticas de empresa como Técnico de Medio Ambiente en la Agencia de Medio Ambiente y Agua de Andalucía (Cazorla). Julio 2015-October 2015*

- Funciones: Seguimiento de especies en peligro de extinción e inventariado del Jardín Botánico "Torre del Vinagre"

### IDIOMAS Y SOFTWARE

- Inglés: Acreditación del nivel B2 por la Universidad de Cambridge
- Francés: 2º en la Escuela Oficial de Idiomas
- Microsoft Office: Nivel intermedio
- Lenguaje R: iniciado