



**UNIVERSIDAD DE JAÉN**  
*Facultad de Ciencias Experimentales*

Trabajo Fin de Grado

# **Análisis de datos de expresión génica en reproducción humana**

**Alumno: María Angustias Martín Estepa**

**Junio, 2018**



**UNIVERSIDAD DE JAÉN**  
*Facultad de Ciencias Experimentales*

Trabajo Fin de Grado

# **ANÁLISIS DE DATOS DE EXPRESIÓN GÉNICA EN REPRODUCCIÓN HUMANA**

**Alumno: María Angustias Martín Estepa**



**Junio, 2018**

## ÍNDICE

ABSTRACT .....	4
RESUMEN.....	4
1. INTRODUCCIÓN .....	5
1.1. Bioinformática .....	5
1.2. Síndrome del Ovario Poliquístico (PCOS).....	6
1.3. Microarrays .....	8
2. OBJETIVOS .....	10
3. MATERIAL Y MÉTODOS .....	11
3.1. Obtención de datos .....	11
3.2. Preprocesamiento .....	12
3.3. Análisis exploratorio .....	12
3.4. Análisis estadístico.....	13
3.4.1. Análisis paramétrico .....	13
3.4.2. Análisis no paramétrico .....	14
3.5. Análisis funcional.....	14
3.6. Interacciones entre genes diferencialmente expresados .....	14
4. RESULTADOS .....	15
4.1. Análisis exploratorio .....	15
4.2. Análisis paramétrico .....	16
4.3. Análisis no paramétrico .....	17
4.4. Análisis funcional.....	19
4.5. Interacción entre genes diferencialmente expresados .....	20
5. DISCUSIÓN .....	22
6. CONCLUSIONES.....	25
7. BIBLIOGRAFÍA .....	26

## **ABSTRACT**

Polycystic Ovary Syndrome (PCOS) is an endocrine disease which causes hormonal imbalance producing menstrual problems and infertility due to the lack of ovulation. This work has been carried out with the aim to compare the gene expression profile of skeletal muscle samples from healthy women and women with insulin resistance and PCOS, obtained by microarray technology, to identify the genes that may be involved in the pathogenesis, as well as inquire in the knowledge and use of computational Systems Biology approaches.

## **RESUMEN**

El Síndrome del Ovario Poliquístico (PCOS) es una endocrinopatía que causa un desequilibrio hormonal dando lugar a problemas menstruales e infertilidad por anovulación. En este Trabajo de Fin de Grado se han comparado los perfiles de expresión génica obtenidos mediante tecnología de microarrays de muestras de músculo esquelético de mujeres sanas y mujeres con PCOS y resistencia a la insulina, con el objetivo de identificar genes que pueden estar implicados en la patogénesis, además de indagar en el conocimiento y la utilización de herramientas computacionales propias de la Biología de Sistemas.

# 1. INTRODUCCIÓN

## 1.1. Bioinformática

La bioinformática es una disciplina que utiliza el conocimiento y tecnologías computacionales para el estudio de datos experimentales, además de otras muchas aplicaciones (Esteban et al., 2008). Esta rama de la Biología permitió la administración de grandes cantidades de datos biológicos producidos por microarrays, espectrómetros de masas cada vez más eficientes y máquinas de secuenciación de ADN de elevado rendimiento (Primig, 2012).

A día de hoy los investigadores tienen a su disposición una gran cantidad de información de acceso público o mediante pago, repositorios de datos y diversas herramientas que no requieren de gran habilidad en informática (Brazas et al., 2011; Galperin y Cochrane, 2011). La genómica ha permitido resolver cuestiones importantes sobre la evolución y el desarrollo del *Homo sapiens* (Collins, 2011; Venter et al., 2015) y numerosos estudios genómicos han ayudado a una mejor comprensión de los procesos moleculares relacionados con la gametogénesis humana (Ostermeier et al., 2002; Kocabas et al., 2006; Chalmel et al., 2007; Ellis et al., 2007; Feig et al., 2007).

Actualmente, la genómica personalizada se está introduciendo en la búsqueda de una mejor comprensión de la composición genética de cada individuo para ayudar en el diagnóstico y tratamiento de enfermedades (Wheeler et al., 2008). Este nuevo enfoque ayudará a aumentar la capacidad de los investigadores para indagar en los orígenes genéticos de determinadas formas de infertilidad humana (Matzuk y Lamb, 2008; Handel y Schimenti, 2010; Hwang et al., 2010). Sin embargo, aún quedan muchos objetivos por cumplir antes de conseguir que la biología del genoma marque la diferencia en la clínica (Chan y Ginsburg, 2011; Fernald et al., 2011).

Por otro lado tenemos la Biología de Sistemas, que surge como una nueva disciplina que se centra en las interacciones entre los diferentes elementos biológicos, su integración y su análisis mediante diversas herramientas computacionales (Díaz-Beltrán et al., 2013). Esta rama emergente de la biología, además de facilitar la investigación biológica básica, proporciona innovadores medios que permiten comprender las enfermedades humanas, identificar biomarcadores de diagnóstico y

desarrollar tratamientos para diversas enfermedades (Wang et al., 2015). En las últimas décadas se han desarrollado una gran variedad de tecnologías de alto rendimiento entre las que se encuentran los microarrays (Brown y Botstein, 1999; Mardis, 2008). Estas tecnologías impulsan a la Biología de Sistemas a transformarse en un valioso enfoque para investigar la Biología Molecular en enfermedades humanas complejas (Sobie et al., 2011).

La Biología de Sistemas es una herramienta muy útil, pues se basa en nuevas técnicas de Biología Molecular y Bioinformática que hacen posible el estudio de diversos genes, proteínas o moléculas en un solo procedimiento (Shafiee et al., 2013), lo cual nos permite una mejor comprensión y búsqueda de los mecanismos responsables del trastorno en el que se va a centrar este Trabajo de Fin de Grado, que es el Síndrome del Ovario Poliquístico.

## **1.2. Síndrome del Ovario Poliquístico (PCOS)**

El Síndrome del Ovario Poliquístico (PCOS, siglas en inglés de *Polycystic Ovary Syndrome*) constituye uno de los trastornos del sistema endocrino más comunes entre las mujeres en edad fértil, encontrándose aproximadamente en el 5% de estas y siendo, además, la principal causa de infertilidad por cese de la ovulación o anovulación (McCartney y Marshall, 2016). Este trastorno se caracteriza fenotípicamente por la presencia de hiperandrogenismo, presencia de múltiples quistes en el ovario y disfunción ovulatoria (The Rotterdam ESHRE/ASRM-Sponsored PCOS Consensus Workshop Group, 2004). En ocasiones puede asociarse con desórdenes metabólicos tales como la resistencia a la insulina, diabetes mellitus tipo 2 (Skov et al., 2007), la hiperinsulinemia y la obesidad. Un 20% de las mujeres con problemas de fertilidad son diagnosticadas con PCOS (Diamanti-Kandarakis et al., 1998). Aunque no se conoce la causa de su aparición, existen evidencias que muestran que diversos factores ambientales y genéticos están implicados en su etiología (Franks y McCarthy, 2004).

La principal causa de infertilidad en mujeres con este trastorno es la secreción excesiva de andrógenos y de hormona luteinizante (LH), que da lugar a anovulación, ciclos menstruales irregulares y menor probabilidad de fecundación (Bellver et al., 2018). Las menstruaciones irregulares constituyen el primer síntoma y resulta fácil

de detectar. Normalmente, las mujeres afectadas tendrán un mayor número de folículos antrales y altas concentraciones de hormona antimülleriana (Bachelot, 2016). Se ha demostrado que se produce la detención del crecimiento del folículo antral, lo que refleja una actividad endocrina anormal (Franks et al., 2000). Las células de la granulosa presentes en el folículo parecen adquirir la capacidad de responder a la presencia de LH cuando tienen un tamaño mucho menor del normal (en este caso tendrían un tamaño aproximado de 3 – 4 mm de diámetro, mientras que el normal sería de 10 mm), produciendo además cantidades anormales de estradiol (y progesterona) dado su pequeño tamaño (Willis et al., 1998). Esta maduración inadecuada del folículo podría deberse en parte a la hipersecreción de LH o también es muy probable que la hiperinsulinemia, muy típica en mujeres con PCOS, desempeñe un papel importante (Dunaif et al., 1987; Conway et al., 1989; Robinson et al., 1993).

La resistencia a la insulina en mujeres afectadas por este trastorno proporciona un mayor riesgo de desarrollar diabetes tipo 2 y problemas cardiovasculares (Dunaif, 1997; Legro, 2003; Ehrmann, 2005; Orio et al., 2006). Además, el aumento en la aparición de PCOS va acorde con el aumento de la obesidad, lo que la convierte en un problema importante en los países desarrollados de cara al futuro (Ehrmann, 2005).

El principal sitio de eliminación de la glucosa estimulada por la insulina es el músculo esquelético (Shulman et al., 1990); por lo tanto, la resistencia a la insulina en este tejido representa un factor de riesgo esencial para la diabetes tipo 2 en mujeres afectadas por PCOS, lo que se refleja en el deterioro de la eliminación oxidativa y no oxidativa de la glucosa total estimulada por la insulina en PCOS (Glintborg et al., 2006), siendo esto muy parecido a los defectos que se observan en la diabetes tipo 2 (Hojlund y Beck-Nielsen, 2006; Petersen y Shulman, 2006).

Por otro lado, las mujeres obesas con PCOS (que son hiperinsulinémicas) son más propensas a padecer alteraciones menstruales (Kiddy et al., 1990) y muestran una respuesta pobre ante la inducción de la ovulación en comparación con mujeres de peso normal que padecen PCOS (Polson et al., 1989; Hamilton-Fairley et al., 1992; White et al., 1996). La imposición de una dieta más baja en calorías en estas mujeres proporciona menores niveles de insulina y mejora el patrón menstrual y la

fertilidad (Pasquali et al., 1989; Kiddy et al., 1992; Clark et al., 1995; Hickey et al., 2002).

El diagnóstico en adolescentes resulta más complicado debido a varios factores. Existen muchas variaciones en el ciclo menstrual en los primeros años tras la menarquia. Además, hay presencia de acné, hirsutismo (presencia de vello en lugares poco comunes en mujeres) y obesidad en diversos casos, siendo todos estos síntomas del PCOS (Bellver et al., 2018). Finalmente, otro factor que dificulta el diagnóstico es que el volumen de los ovarios durante la adolescencia, en concreto durante la pubertad, es mayor que en la edad adulta (Fruzzetti et al., 2015). Hay que tener en cuenta además que los ovarios multifoliculares pueden constituir una etapa del desarrollo normal en la adolescencia (Johnstone et al., 2010).

Debido a que, como anteriormente se ha indicado, este trastorno se manifiesta con síntomas tales como hiperandrogenismo, anovulación, infertilidad o resistencia a la insulina, los objetivos del tratamiento serán mejorar estos aspectos, especialmente los problemas de fertilidad y los de resistencia a la insulina (Nandi et al., 2014).

### **1.3. Microarrays**

Un microarray es un pequeño chip que hace posible el análisis de miles de genes del genoma de forma simultánea, precisa y rápida, todo ello en un único experimento (Lee et al., 2013). Esta técnica es frecuentemente utilizada en biomedicina y se basa en el principio de complementariedad de bases del ADN, de forma que un solo microarray puede incluir cientos o miles de secuencias cortas de ADN (sondas) que hibridan con secuencias complementarias de una forma determinada, proceso llevado a cabo sobre una superficie sólida. Tienen múltiples características que hacen que sean una herramienta valiosa y esencial en lo que se conoce como era genómica (Harris et al., 2007), dado que permiten la monitorización de la expresión génica con un coste y esfuerzo relativamente bajos.

Los arrays se pueden clasificar de acuerdo a dos de sus características principales: la constitución de la sonda de ADN y los fluoróforos que se utilizan para su marcaje y detección. De este modo podemos distinguir dos grupos distintos de arrays. Tendríamos un primer grupo formado por los arrays de ADN complementario

(ADNc), los cuales están formados por fragmentos de ADN obtenidos por transcripción inversa a partir de ARN mensajero, quedando así sobre el array diversas sondas de transcritos de genes concretos (Vargas et al., 2018). En este primer grupo también se encuentran englobados los arrays de oligonucleótidos, en los que se utilizan fragmentos de ADN de cadena sencilla con una longitud de entre 15 y 25 nucleótidos (cortas) o de entre 50 y 120 nucleótidos (largas), siendo las primeras las que se utilizan habitualmente; estas sondas se unirán a regiones concretas conocidas de determinados genes (Barrero, 2005).

Por otro lado, existe un segundo grupo constituido por arrays de uno o dos colores. Los arrays de un color utilizan un solo fluoróforo y su utilidad radica en la estimación de valores absolutos de expresión génica; un ejemplo de estos arrays son los de tipo Affymetrix (Vargas et al., 2018). Sin embargo, los arrays de dos colores utilizan dos fluoróforos distintos que marcarán dos muestras determinadas que serán analizadas, de manera que estas dos muestras se hibridarán sobre un array y emitirán una señal que será observada mediante un escáner, tras lo cual se conocerán las intensidades de cada fluoróforo, lo que proporcionará información para la identificación de diferentes genes (Oberthuer et al., 2010). En este trabajo se analizarán datos procedentes de la aplicación de arrays de un color, tipo Affymetrix.

Una vez se lleva a cabo la hibridación, se utilizan diversos métodos para la detección y cuantificación de la señal producida por esta. Esta información generada es reunida en forma de matrices de expresión génica (Vargas et al., 2018).

## **2. OBJETIVOS**

El objetivo general que ha llevado al desarrollo de este Trabajo de Fin de Grado es el uso de diversas herramientas computacionales propias de la Biología de Sistemas para el análisis de datos procedentes de la aplicación de la tecnología de microarrays, con el fin de indagar en los aspectos moleculares que llevan al fracaso reproductivo en el caso de las mujeres afectadas por el Síndrome del Ovario Poliquístico.

Existen además una serie de objetivos específicos:

1. Aprender a utilizar las herramientas computacionales correspondientes a la Biología de Sistemas.
2. Determinar el perfil de expresión génica de mujeres sanas y con ovario poliquístico.
3. Conocer los procesos biológicos en los que intervienen los genes cuya expresión se encuentra diferencialmente modificada mediante diversos análisis bioinformáticos.
4. Establecer si existen relaciones entre los genes diferencialmente expresados mediante la construcción de redes de interacción.

### 3. MATERIAL Y MÉTODOS

#### 3.1. Obtención de datos

Llevamos a cabo el análisis de una serie de datos obtenidos de un repositorio de acceso público, llamado *Gene Expression Omnibus* (GEO), que almacena y distribuye libremente diversos conjuntos de datos que pueden reutilizarse como base para nuevos estudios biológicos (Rung y Brazma, 2013). En concreto, nuestros datos pertenecen al experimento GSE6798, basados en un array de tipo *Affymetrix Human Genome*, es decir, un array de un color, el cual permite el estudio de la expresión de un elevado número de transcritos, en este caso se dispone de 54675 sondas (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE6798>).

El conjunto de datos analizados consta de 29 muestras, de las cuales 13 son muestras control procedentes de sujetos sanos y 16 son muestras de pacientes con PCOS, resistentes a la insulina, en edad fértil. La información relativa a la codificación utilizada para denominar a cada muestra se encuentra en la Tabla 1.

NOMBRE DE MUESTRA	TIPO DE MUESTRA
GSM155631.cel	Control 1
GSM155643.cel	Control 2
GSM155644.cel	Control 3
GSM155729.cel	Control 4
GSM156170.cel	Control 5
GSM156171.cel	Control 6
GSM156176.cel	Control 7
GSM156177.cel	Control 8
GSM156178.cel	Control 9
GSM156179.cel	Control 10
GSM156180.cel	Control 11
GSM156181.cel	Control 12
GSM156184.cel	Control 13
GSM156186.cel	PCOS 1
GSM156187.cel	PCOS 2
GSM156510.cel	PCOS 3
GSM156511.cel	PCOS 4
GSM156512.cel	PCOS 5
GSM156749.cel	PCOS 6
GSM156750.cel	PCOS 7
GSM156751.cel	PCOS 8
GSM156752.cel	PCOS 9
GSM156753.cel	PCOS 10
GSM156763.cel	PCOS 11
GSM156946.cel	PCOS 12
GSM156948.cel	PCOS 13
GSM156949.cel	PCOS 14

GSM156950.cel	PCOS 15
GSM156951.cel	PCOS 16

**Tabla 1.** Tabla de equivalencias que muestra el tipo de muestras con su nombre correspondiente.

### 3.2. Preprocesamiento

En primer lugar, se importaron los datos para obtener una matriz de expresión sobre la cual se realizaron análisis posteriores. Esta matriz de expresión refleja, para distintos tipos de sondas, los valores de expresión para cada muestra.

A continuación, se aplicó una normalización de tipo RMA, una función que se encuentra incluida en el paquete *affy* de Bioconductor (<https://www.bioconductor.org/packages/release/bioc/html/affy.html>). Bioconductor es un software de código y desarrollo abierto que posibilita la comprensión así como el análisis de datos genómicos (Gentleman et al., 2005) y utiliza el lenguaje de programación estadística R, el cual dispone de una amplia variedad de herramientas destinadas al estudio estadístico, así como a la representación gráfica de los datos (Okoniewski y Miller, 2008). La normalización de tipo RMA es un procedimiento que se aplica a los datos procedentes de arrays de tipo Affymetrix, como es el caso. Esta función RMA sustrae el fondo, con lo que elimina los factores de confusión posibles; tras esto, transforma logarítmicamente los datos uniformando la escala en las tasas de expresión y realiza una normalización de tipo cuantiles, de modo que se elimina toda la variabilidad no biológica (Esteban et al., 2008).

### 3.3. Análisis exploratorio

El análisis de datos de microarrays continuó con un análisis exploratorio utilizando el programa MeV – TM4 (*Multi Experiment Viewer – TM4 Software Suite*; <http://mev.tm4.org>), que consta de múltiples herramientas para la gestión y análisis de datos así como la visualización de los mismos (Saeed et al., 2006). En concreto, llevamos a cabo un agrupamiento o *clustering* jerárquico y un análisis de componentes principales (PCA), lo que nos permitió hacer un análisis preliminar para ver si existía un patrón de expresión de las muestras previamente establecido.

El agrupamiento o *clustering* jerárquico permite dividir un determinado conjunto de datos en diferentes grupos según características concretas. De este modo los datos

reunidos en un grupo serán similares entre sí y diferentes de los datos de otros grupos, siendo representados gráficamente en forma de dendrograma (Jiang et al., 2004). En nuestro caso los análisis se realizaron tomando como medida de distancia la euclídea, y como método de agrupamiento, el de tipo completo (*complete linkage*).

Por otra parte, el análisis de componentes principales (PCA, siglas en inglés para *Principal Component Analysis*) es una herramienta de análisis de datos multivariante que permite, conservando la mayor cantidad de información posible, reducir dicha información de un gran número de variables a un menor número de estas. El análisis estadístico, así como la visualización de estas nuevas variables obtenidas, los componentes principales, pueden ser de gran ayuda para estudiar similitudes y diferencias entre las diversas muestras (Groth et al., 2013). Para la realización del PCA en este trabajo se utilizaron los parámetros por defecto en TM4.

### **3.4. Análisis estadístico**

Tras realizar el análisis exploratorio llevamos a cabo un análisis estadístico, con el fin de detectar si existían o no diferencias significativas entre ambos grupos experimentales. Para ello, aplicamos dos tests estadísticos, uno paramétrico (test de Welch) y otro no paramétrico (*Rank Product*), ambos realizados mediante el programa MeV – TM4, anteriormente mencionado.

#### *3.4.1. Análisis paramétrico*

Para llevar a cabo el análisis paramétrico, se utilizó como prueba estadística el test de Welch. Esta prueba nos permite detectar las diferencias que puedan existir entre las muestras, dado que asume que las varianzas de las muestras son distintas (Markowski y Markowski, 1993). Aquellos genes cuyo p-valor corregido fuese menor a 0,05 y cuyo valor absoluto de la tasa de cambio fuese superior a 1,5 se consideraron como diferencialmente expresados.

### 3.4.2. Análisis no paramétrico

En cuanto al análisis no paramétrico, se optó por el *Rank Product*, que clasifica los genes en cuanto a la intensidad de expresión que presentan dentro de cada microarray, pudiendo identificar si existen genes diferencialmente expresados (Alekseyenko, 2016). Así, se consideraron como genes diferencialmente expresados aquellos cuyo p-valor corregido fuera inferior a 0,05 y cuyo valor absoluto de tasa de cambio fuese igual o superior a 1,5.

### 3.5. Análisis funcional

Realizamos un análisis de enriquecimiento funcional con el fin de conocer los procesos biológicos en los que intervienen los genes diferencialmente expresados que habíamos obtenido. Para ello utilizamos la plataforma DAVID (*Database for Annotation, Visualization and Integrated Discovery*, <https://david.ncicrf.gov/>) a la que se puede acceder libremente y que nos proporciona las herramientas necesarias para llevar a cabo una interpretación funcional de una determinada lista de genes (Jiao et al., 2012).

### 3.6. Interacciones entre genes diferencialmente expresados

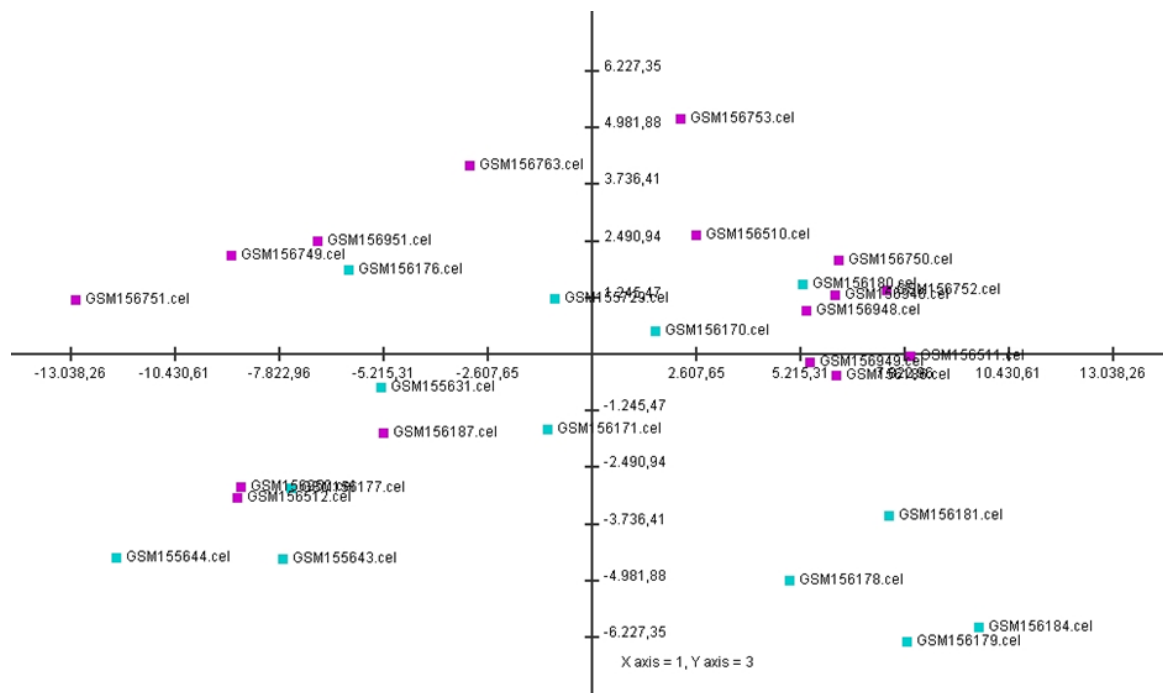
Para poder observar si existían interacciones entre los distintos genes diferencialmente expresados llevamos a cabo la construcción de una red mediante el uso de la plataforma STRING, en la cual se introduce un conjunto de genes o proteínas que este programa identifica y a partir de estos genera una red con las relaciones existentes entre los distintos componentes, si las hubiera (<https://string-db.org/>; Szklarczyk et al., 2015).

## 4. RESULTADOS

### 4.1. Análisis exploratorio

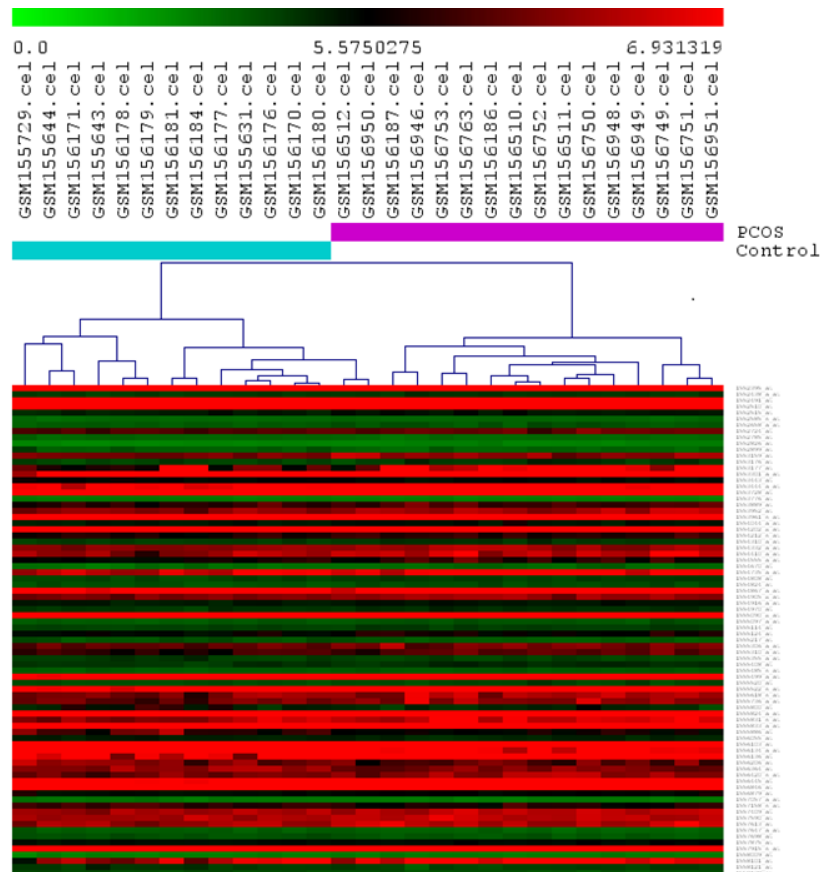
El análisis exploratorio se llevó a cabo con el fin de comprobar si existían similitudes entre las distintas muestras mediante la realización de un agrupamiento o *clustering* jerárquico y un análisis de componentes principales (PCA) utilizando el programa MeV – TM4 (*Multi Experiment Viewer – TM4 Software Suite*; <http://mev.tm4.org/>).

Teniendo en cuenta el valor de expresión de todos los genes del genoma, el análisis de componentes principales no detectó un patrón de distribución fijado entre las muestras control y las muestras de pacientes con Síndrome del Ovario Poliquístico, como puede observarse en la Figura 1.



**Figura 1.** Resultado obtenido tras llevar a cabo el análisis de componentes principales, donde se observan las muestras control en azul y las muestras de pacientes afectados por PCOS en morado. Podemos comprobar que, cuando se tienen en cuenta los valores de expresión de todos los genes del genoma, las muestras se agrupan sin seguir un determinado patrón.

En la Figura 2 se muestran parte de los resultados obtenidos en el agrupamiento jerárquico, donde podemos observar que, teniendo en cuenta los genes diferencialmente expresados (con un p-valor no corregido inferior a 0,01) sí existe un patrón de agrupamiento entre las muestras control y las muestras de pacientes con Síndrome del Ovario Poliquístico.



**Figura 2.** Resultados obtenidos tras la realización del *cluster* jerárquico, donde observamos que existe un agrupamiento diferencial previamente establecido entre las muestras, para los genes detectados como significativos por el test de Welch (p-valor no corregido inferior a 0,01). Las muestras control están representadas en azul y las muestras de pacientes con Síndrome del Ovario Poliquístico, nombradas como PCOS (*Polycystic Ovary Syndrome*), en morado.

#### 4.2. Análisis paramétrico

Tras la aplicación del test de Welch, con la correspondiente corrección del p-valor, encontramos que no existían diferencias significativas entre los grupos experimentales objeto de estudio, es decir, que no existe ningún gen diferencialmente expresado de acuerdo con los criterios fijados (p-valor corregido menor a 0,05 y valor absoluto de tasa de cambio mayor que 1,5).

### 4.3. Análisis no paramétrico

Para llevar a cabo el test de tipo no paramétrico utilizamos el método de *Rank Product*, mediante el cual detectamos cuáles eran los genes diferencialmente expresados, es decir, aquellos que poseen un p-valor corregido inferior a 0,05 y un valor absoluto de tasa de cambio por encima de 1,5 en el caso de los genes sobreexpresados, y por debajo de -1,5 en el caso de los genes expresados a la baja. En la Tabla 2 se muestran los resultados obtenidos: concretamente, se detectaron 50 genes sobreexpresados (representados en verde) y 15 genes expresados a la baja (representados en naranja).

GENES	DESCRIPCIÓN DEL GEN	FC_PCOS/C	MEDIA CONTROL	MEDIA PCOS
RAPH1	<i>Ras association (RalGDS/AF-6) and pleckstrin homology domains 1</i>	1,635	6,375	7,085
EIF4E2	<i>Eukaryotic translation initiation factor 4E family member 2</i>	1,551	6,717	7,350
N/A	N/A	2,188	5,377	6,507
SNHG14	<i>Small nucleolar RNA host gene 14</i>	1,626	7,016	7,718
NANOS1	<i>Nanos C2HC-type zinc finger 1</i>	2,039	7,374	8,401
RBM3	<i>RNA binding motif (RNP1, RRM) protein 3</i>	1,943	6,647	7,606
FRMD6	<i>FERM domain containing 6</i>	1,741	6,383	7,184
RAPH1	<i>Ras association (RalGDS/AF-6) and pleckstrin homology domains 1</i>	1,524	5,079	5,687
RASSF3	<i>Ras association domain family member 3</i>	1,895	7,430	8,352
MSTN	<i>Myostatin</i>	1,914	5,224	6,160
SH2D1B	<i>SH2 domain containing 1B</i>	1,833	6,061	6,935
IRX3	<i>Iroquois homeobox 3</i>	1,693	5,664	6,424
SHISA2	<i>Shisa family member 2</i>	1,839	7,068	7,947
FRZB	<i>Frizzled-related protein</i>	2,413	7,456	8,727
TMEM182	<i>Transmembrane protein 182</i>	1,511	7,817	8,413
MLF1	<i>Myeloid leukemia factor 1</i>	1,643	8,129	8,846
KCNQ5	<i>Potassium voltage-gated channel subfamily Q member 5</i>	1,513	7,656	8,253
FRZB	<i>Frizzled-related protein</i>	2,397	7,062	8,323
MYLK4	<i>Myosin light chain kinase family member 4</i>	1,588	7,112	7,780
LINC01854	<i>Long intergenic non-protein coding RNA 1854</i>	1,548	7,958	8,588
HBG1	<i>Hemoglobin subunit gamma 1</i>	1,563	7,631	8,275
GRB14	<i>Growth factor receptor bound protein 14</i>	1,551	5,427	6,060
AMFR	<i>Autocrine motility factor receptor</i>	1,524	7,191	7,799

<b>LINC01091</b>	<i>Long intergenic non-protein coding RNA 1091</i>	<b>1,582</b>	4,391	5,052
<b>PVALB</b>	<i>Parvalbumin</i>	<b>1,587</b>	5,948	6,614
<b>HBG2</b>	<i>Hemoglobin subunit gamma 2</i>	<b>1,585</b>	7,130	7,794
<b>PPP3R1</b>	<i>Protein phosphatase 3 regulatory subunit B, alpha</i>	<b>1,510</b>	6,844	7,439
<b>MICAL2</b>	<i>Microtubule associated monooxygenase, calponin and LIM domain containing 2</i>	<b>1,560</b>	6,593	7,235
<b>SORL1</b>	<i>Sortilin related receptor 1</i>	<b>1,598</b>	5,420	6,097
<b>HS6ST2</b>	<i>Heparan sulfate 6-O-sulfotransferase 2</i>	<b>1,575</b>	5,405	6,061
<b>HBB</b>	<i>Hemoglobin subunit beta</i>	<b>1,620</b>	12,391	13,087
<b>MIR6787</b>	<i>MicroRNA 6787</i>	<b>1,549</b>	6,006	6,638
<b>CALML6</b>	<i>Calmodulin like 6</i>	<b>1,735</b>	6,626	7,421
<b>NMRK2</b>	<i>Nicotinamide riboside kinase 2</i>	<b>1,690</b>	8,425	9,183
<b>N/A</b>	<i>N/A</i>	<b>1,599</b>	6,998	7,676
<b>MYH1</b>	<i>Myosin heavy chain 1</i>	<b>2,041</b>	12,059	13,088
<b>PWAR6</b>	<i>Prader Willi/Angelman region RNA 6</i>	<b>1,581</b>	5,636	6,297
<b>HBB</b>	<i>Hemoglobin subunit beta</i>	<b>1,614</b>	11,919	12,609
<b>PRR16</b>	<i>Proline rich 16</i>	<b>1,538</b>	6,769	7,390
<b>HS6ST2</b>	<i>Heparan sulfate 6-O-sulfotransferase 2</i>	<b>1,508</b>	3,878	4,471
<b>HBA1</b>	<i>Hemoglobin subunit alpha 1</i>	<b>1,620</b>	12,357	13,052
<b>SLC16A3</b>	<i>Solute carrier family 16 member 3</i>	<b>1,561</b>	7,680	8,322
<b>S100A8</b>	<i>S100 calcium binding protein A8</i>	<b>2,098</b>	6,651	7,720
<b>HBA2</b>	<i>Hemoglobin subunit alpha 2</i>	<b>1,557</b>	12,439	13,078
<b>HBA1</b>	<i>Hemoglobin subunit alpha 1</i>	<b>1,528</b>	12,476	13,088
<b>HBA2</b>	<i>Hemoglobin subunit alpha 2</i>	<b>1,529</b>	12,450	13,063
<b>RGS2</b>	<i>Regulator of G protein signaling 2</i>	<b>1,519</b>	6,463	7,066
<b>LRRC3B</b>	<i>Leucine rich repeat containing 3B</i>	<b>1,550</b>	5,580	6,212
<b>ACTN3</b>	<i>Actinin alpha 3 (gene/pseudogene)</i>	<b>2,309</b>	8,830	10,037
<b>MNDA</b>	<i>Myeloid cell nuclear differentiation antigen</i>	<b>1,530</b>	5,736	6,350
<b>LDHB</b>	<i>Lactate dehydrogenase B</i>	<b>-1,707</b>	10,434	9,662
<b>LPL</b>	<i>Lipoprotein lipase</i>	<b>-1,632</b>	10,054	9,348
<b>LDHB</b>	<i>Lactate dehydrogenase B</i>	<b>-1,695</b>	11,008	10,246
<b>TECRL</b>	<i>Trans-2,3-enoyl-CoA reductase like</i>	<b>-1,702</b>	8,529	7,761
<b>LPL</b>	<i>Lipoprotein lipase</i>	<b>-1,561</b>	8,906	8,264
<b>SMAD9</b>	<i>SMAD family member 9</i>	<b>-1,664</b>	6,763	6,028
<b>ALDH6A1</b>	<i>Aldehyde dehydrogenase 6 family member A1</i>	<b>-1,524</b>	7,760	7,152
<b>SLC38A1</b>	<i>Solute carrier family 38 member 1</i>	<b>-1,790</b>	6,471	5,631
<b>MAFF</b>	<i>MAF bZIP transcription factor F</i>	<b>-1,629</b>	4,107	3,403
<b>IDI2</b>	<i>Isopentenyl-diphosphate delta isomerase 2</i>	<b>-1,536</b>	9,566	8,947
<b>H19</b>	<i>H19, imprinted maternally expressed transcript (non-protein coding)</i>	<b>-1,719</b>	10,059	9,277
<b>MIR675</b>	<i>MicroRNA 675</i>	<b>-1,719</b>	10,059	9,277
<b>C12ORF75</b>	<i>Chromosome 12 open reading frame 75</i>	<b>-1,728</b>	7,567	6,778

<b>SLC38A1</b>	<i>Solute carrier family 38 member 1</i>	<b>-2,295</b>	6,296	5,097
<b>C9ORF135</b>	<i>Chromosome 9 open reading frame 135</i>	<b>-1,795</b>	6,447	5,602

**Tabla 2.** Genes diferencialmente expresados obtenidos tras el análisis no paramétrico mediante la aplicación del test de *Rank Product*. Se muestra, para cada gen, su símbolo, nombre completo y los niveles medios de expresión de cada grupo experimental, así como su tasa de cambio (FC: *fold change*).

#### 4.4. Análisis funcional

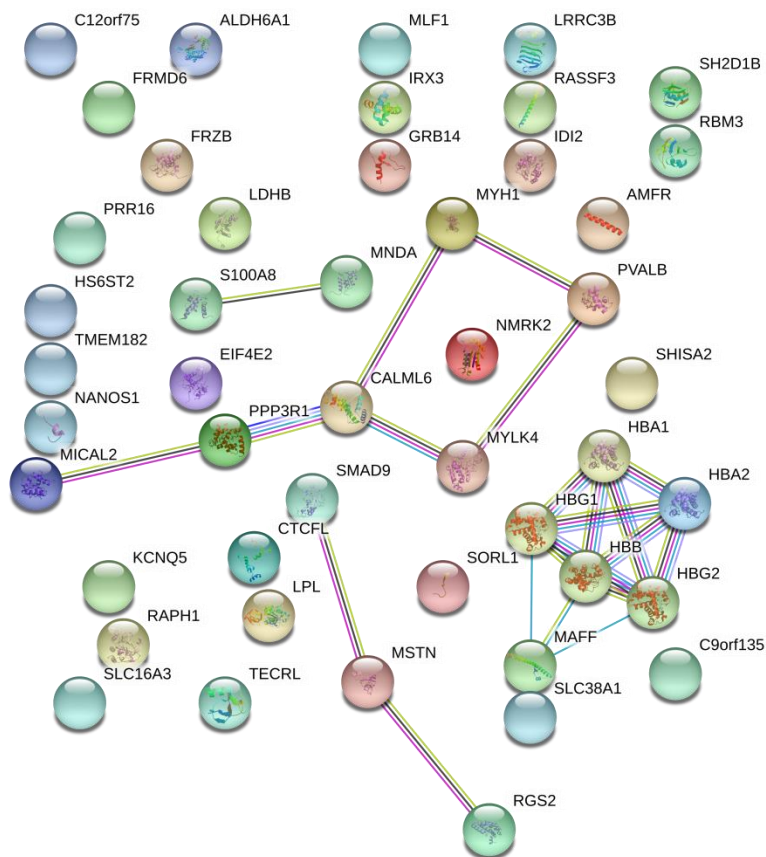
El análisis funcional nos permitió conocer los procesos biológicos asociados a los genes diferencialmente expresados obtenidos tras realizar el análisis no paramétrico por medio de *Rank Product* (Tabla 3). Encontramos diversas funciones biológicas estadísticamente significativas ( $p$ -valor  $< 0,05$ ), entre las cuales podemos destacar el transporte de oxígeno, en el cual encontramos cinco genes implicados.

GENES	PROCESOS BIOLÓGICOS	P-VALOR	FDR (%)
<b>HBG1, HBA2, HBA1, HBG2, HBB</b>	Transporte de oxígeno	4,09E-08	5,43E-05
<b>HBA2, HBA1, HBB</b>	Catabolismo del peróxido de hidrógeno	0,001075	1,417153
<b>HBA2, HBA1, HBB</b>	Regulación positiva de la muerte celular	0,002265	2,964746
<b>HBA2, HBA1, HBB</b>	Transporte de bicarbonato	0,005157	6,631959
<b>HBA2, HBA1, HBB</b>	Respuesta al peróxido de hidrógeno	0,006876	8,750627
<b>MAFF, HBG1, HBG2, HBB</b>	Coagulación de la sangre	0,010166	12,682746
<b>SORL1, HBA2, HBA1, HBB</b>	Endocitosis mediada por receptor	0,010468	13,035459
<b>HBA2, HBA1, HBB</b>	Heterooligomerización de proteínas	0,011634	14,385538
<b>HBA2, HBA1, HBB</b>	Eliminación de radicales superóxido o peróxido de hidrógeno	0,012649	15,545221
<b>MSTN, ACTN3</b>	Atrofia del músculo esquelético	0,019371	22,866206
<b>SLC16A3, LDHB</b>	Metabolismo del piruvato	0,052393	51,045415
<b>MSTN, NMRK2</b>	Regulación negativa de la diferenciación de mioblastos	0,054710	52,610362
<b>ALDH6A1, RGS2</b>	Diferenciación de las células de la grasa marrón	0,075313	64,628323
<b>RGS2, SORL1</b>	Regulación negativa de la actividad MAP quinasa	0,084329	68,941752
<b>RASSF3, MSTN, ACTN3</b>	Regulación del proceso de apoptosis	0,095206	73,496997

**Tabla 3.** Principales procesos biológicos en los que se encuentran implicados los genes diferencialmente expresados. FDR: *False Discovery Rate*.

#### 4.5. Interacción entre genes diferencialmente expresados

Una vez llevado a cabo el análisis funcional y con el fin de visualizar si entre los genes diferencialmente expresados se producían interacciones, llevamos a cabo la construcción de una red mediante la plataforma STRING, que puede observarse en la Figura 3. En la red observamos los distintos genes diferencialmente expresados representados en forma de esfera (nodos), y las interacciones entre ellos están representadas mediante líneas (aristas). De este modo podemos conocer entre qué genes existe relación y similitud en cuanto a las distintas funciones biológicas en las que participa cada uno de ellos.



**Figura 3.** Red generada mediante STRING que muestra las interacciones existentes entre algunos de los genes diferencialmente expresados.

Tras observar la red comprobamos que existen pocas interconexiones y encontramos cuatro grupos de genes interconectados. El primero lo componen los genes S100A8 y MNDA. Un segundo grupo está integrado por seis genes que son: MYH1, CALML6, MICAL2, PVALB, MYLK4 y PPP3R1; podemos ver que existe un agrupamiento cíclico entre estos tres últimos. En el tercero encontramos tres genes:

SMAD9, MSTN y RGS2, entre ellos hay relación lineal. Por último, existe un grupo formado por seis genes diferencialmente expresados: MAFF, HBA1, HBA2, HBG1, HBB y HBG2, los cinco últimos mencionados forman un módulo de alta interacción, debido al hecho de que participan en los mismos procesos biológicos, tal y como indica el análisis funcional.

## 5. DISCUSIÓN

El Síndrome del Ovario Poliquístico es una endocrinopatía responsable de provocar infertilidad por anovulación, siendo uno de sus síntomas comunes la resistencia a la insulina. Existen diversos estudios, además de este Trabajo de Fin de Grado, que han comparado la expresión de genes en pacientes con PCOS que presentan resistencia a la insulina y sujetos control sanos mediante el uso de enfoques globales tales como la transcriptómica y la proteómica, utilizando microarrays de ADN (Skov et al., 2007; Kaur et al., 2012; Wang et al., 2015), lo que pone de manifiesto la utilidad que estas herramientas presentan frente al análisis de una gran cantidad de genes. A pesar del gran volumen de estudios realizados, la etiología y patogenia del PCOS siguen sin estar claras (Lan et al., 2015).

Nuestros datos provienen del estudio realizado por V. Skov y colaboradores en el año 2007 (Skov et al., 2007), en el cual se detectaron 3730 genes expresados a la baja y 2407 sobreexpresados mediante la utilización del software estadístico R. Además, en este trabajo previo se demostró que existía una expresión significativa a la baja de genes implicados en el proceso de fosforilación oxidativa y del gen PGC-1 $\alpha$ , lo cual puso de manifiesto la asociación existente entre la resistencia a la insulina y el metabolismo oxidativo de las mitocondrias, que se encuentra alterado en el músculo esquelético de mujeres con PCOS. Cabe destacar que los resultados obtenidos se validaron mediante una PCR a tiempo real y la utilización de otras herramientas informáticas diferentes (Skov et al., 2007).

En el presente Trabajo de Fin de Grado encontramos 65 genes diferencialmente expresados, de los cuales 50 se corresponden con genes sobreexpresados y los 15 restantes, con genes expresados a la baja. De este modo, encontramos una cantidad mucho menor de genes con expresión diferencial. Sin duda, este hecho se debe al análisis computacional diferencial que hemos llevado a cabo, más estricto y, en nuestra opinión, adecuado al ser de tipo no paramétrico, dado el bajo número de muestras.

Por otro lado, el primer estudio analizó diversos mecanismos moleculares obteniendo los más significativos, mientras que en este caso se analizaron diversos genes y a partir de los diferencialmente expresados se conocieron los procesos

biológicos a los que están asociados. Así pues, obtuvimos varios procesos biológicos significativos ( $p$ -valor  $< 0,05$ ), de entre los cuales destaca el transporte de oxígeno, el cual se encuentra asociado con la fosforilación oxidativa, anteriormente mencionada, dado que el oxígeno actúa como aceptor al final de la cadena de transporte de electrones. Además de este, encontramos diversas funciones más relacionadas en ambos estudios como el proceso de apoptosis y otras relacionadas con diversos tejidos como el muscular.

Resulta interesante que en este proyecto se hayan obtenido dos genes sobreexpresados relacionados con la atrofia del músculo esquelético (MSTN y ACTN3), pues se pone de manifiesto la resistencia a la insulina presente en estas muestras de mujeres con PCOS, debido a que el músculo esquelético es el principal sitio de captación de glucosa mediada por insulina. Al existir dicha resistencia, el músculo esquelético captaría menos glucosa, pudiendo producirse su atrofia. Además, también encontramos que el gen MSTN está relacionado con el proceso de regulación negativa de los mioblastos (células precursoras de las células musculares), junto con otro gen llamado NMRK2.

Además de lo anterior, diversos hallazgos proporcionan evidencias de que la vía de señalización de MAP quinasa juega un papel importante en PCOS (Lan et al., 2015). De este modo, existe otro proceso biológico destacable, se trata de la regulación negativa de la actividad MAP quinasa (MAPK), en la cual se encuentran implicados los genes con expresión diferencial RGS2 y SORL1. En este sentido, estudios previos han determinado que la activación anormal de la vía de señalización MAPK en distintos tejidos puede contribuir a la producción excesiva de andrógenos en mujeres con PCOS (Nelson-Degrave et al., 2005; Corbould et al., 2006).

Otra función de interés a la que se encuentran asociada los genes diferencialmente expresados ALDH6A1 y RGS2 es la diferenciación de las células de grasa marrón. Este tipo celular compone el tejido adiposo marrón, cuya principal función es el mantenimiento de la temperatura corporal (Cannon y Nedergaard, 2004). Sin embargo, varios estudios han demostrado que en humanos existe tejido adiposo marrón metabólicamente activo (Cypess et al., 2009; Virtanen et al., 2009). Además,

se demostró en ratas con PCOS, en las cuales el tejido adiposo marrón no tiene actividad, que el trasplante de pequeñas cantidades de dicho tejido provocaba la activación del mismo y, en consecuencia, la mejora significativa de características clave del PCOS como la resistencia a la insulina, el ciclo estral irregular y la infertilidad (Yuan et al., 2016). Esto pone de manifiesto la conexión existente entre las células de grasa marrón y la patología en cuestión.

Finalmente, dado todo lo anteriormente descrito, podemos afirmar que los microarrays resultan de una gran utilidad en la búsqueda de los mecanismos genéticos subyacentes a las enfermedades, en este caso de Síndrome del Ovario Poliquístico, aunque aún queda mucho por indagar al respecto.

## **6. CONCLUSIONES**

Podemos extraer diversas conclusiones de este Trabajo de Fin de Grado, entre las cuales podemos destacar la utilidad de los microarrays para el estudio de diferentes patologías mediante la comparación de individuos sanos e individuos enfermos, lo que permite indagar en los mecanismos genéticos y moleculares que dan lugar a las enfermedades.

Por otra parte, encontramos que existe a nuestra disposición una enorme variedad de herramientas y programas propios de la Biología de Sistemas, así como de bases de datos de uso público que hacen posible que se puedan llevar a cabo este tipo de estudios. Entre ellos cabe destacar los aquí utilizados como son MeV – TM4, programa que nos permitió la realización de los análisis exploratorio y estadístico; DAVID, plataforma que nos ha permitido llevar a cabo el análisis funcional de los genes en cuestión, y STRING, herramienta que nos facilitó información acerca de las relaciones existentes entre los diferentes genes.

Por último, este estudio ha hecho posible detectar los procesos biológicos subyacentes a los genes con expresión diferencial en el Síndrome del Ovario Poliquístico, utilizando procedimientos estadísticos y computacionales estrictos y, en nuestra opinión, más adecuados que los utilizados por otros autores que analizaron los mismos datos.

## 7. BIBLIOGRAFÍA

- Alekseyenko AV. Multivariate Welch t-test on distances. *Bioinformatics*. 2016; 32(23):3552–3558.
- Bachelot A. Polycystic Ovarian Syndrome: clinical and biological diagnosis. *Annales de Biologie Clinique (Paris)*. 2016; 74(6):661–667.
- Barrero P. Aplicaciones de la técnica de microarrays en ciencias biomédicas: presente y futuro. *Química Viva*. 2005; 4(3):91–100.
- Bellver J, Rodríguez-Taberner L, Robles A, Muñoz E, Martínez F, Landeras J, Acevedo B. Polycystic Ovary Syndrome throughout a woman's life. *Journal of Assisted Reproduction and Genetics*. 2018; 35(1):25–39.
- Brazas MD, Yim DS, Yamada JT, Ouellette BF. The 2011 Bioinformatics Links Directory update: more resources, tools and databases and features to empower the bioinformatics community. *Nucleic Acids Research*. 2011; 39(Web server issue):W3–W7.
- Brown PO, Botstein D. Exploring the new world of the genome with DNA microarrays. *Nature Genetics*. 1999; 21(1 Suppl):33–37.
- Cannon B, Nedergaard J. Brown adipose tissue: Function and physiological significance. *Physiological Reviews*. 2004; 84(1):277–359.
- Chalmel F, Rolland AD, Niederhauser-Wiederkehr C, Chung SS, Demougin P, Gattiker A, Moore AJ, Patard JJ, Wolgemuth DJ, Jegou B, Primig M. The conserved transcriptome in human and rodent male gametogenesis. *Proceedings of the National Academy of Sciences of the United States of America*. 2007; 104(20):8346–8351.
- Chan IS, Ginsburg GS. Personalized medicine: progress and promise. *Annual Review of Genomics and Human Genetics*. 2011; 12:217–244.
- Clark AM, Ledger W, Galletly C, Tomlinson L, Blaney F, Wang X, Norman RJ. Weight loss results in significant improvement in pregnancy and ovulation rates in anovulatory obese women. *Human Reproduction (Oxford, England)*. 1995; 10(10):2705–2712.

- Collins FS. Genome-sequencing anniversary. *Faces of the genome*. *Science*. 2011; 331(6017):546.
- Conway GS, Honour JW, Jacobs HS. Heterogeneity of the Polycystic Ovary Syndrome: clinical, endocrine and ultrasound features in 556 patients. *Clinical Endocrinology*. 1989; 30(4):459–470.
- Corbould A, Zhao H, Mirzoeva S, Aird F, Dunaif A. Enhanced mitogenic signaling in skeletal muscle of women with Polycystic Ovary Syndrome. *Diabetes*. 2006; 55(3):751–759.
- Cypess AM, Lehman S, Williams G, Tal I, Rodman D, Goldfine AB, Kuo FC, Palmer EL, Tseng YH, Doria A, Kolodny GM, Kahn CR. Identification and importance of brown adipose tissue in adult humans. *The New England Journal of Medicine*. 2009; 360(15):1509–1517.
- Diamanti-Kandarakis E, Kouli C, Tsianateli T, Bergiele A. Therapeutic effects of metformin on insulin resistance and hyperandrogenism in Polycystic Ovary Syndrome. *European Journal of Endocrinology*. 1998; 138(3):269–274.
- Díaz-Beltrán L, Cano C, Wall DP, Esteban FJ. Systems Biology as a comparative approach to understand complex gene expression in neurological diseases. *Behavioral Sciences (Basel, Switzerland)*. 2013; 3(2):253–272.
- Dunaif A. Insulin resistance and the Polycystic Ovary Syndrome: mechanism and implications for pathogenesis. *Endocrine Reviews*. 1997; 18(6):774–800.
- Dunaif A, Graf M, Mandeli J, Laumas V, Dobrjansky A. Characterization of groups of hyperandrogenic women with acanthosis nigricans, impaired glucose tolerance, and/or hyperinsulinemia. *The Journal of Clinical Endocrinology and Metabolism*. 1987; 65(3):499–507.
- Ehrmann DA. Polycystic Ovary Syndrome. *The New England Journal of Medicine*. 2005; 352(12):1223–1236.
- Ellis PJ, Furlong RA, Conner SJ, Kirkman-Brown J, Afnan M, Barratt C, Griffin DK, Affara NA. Coordinated transcriptional regulation patterns associated with

- infertility phenotypes in men. *Journal of Medical Genetics*. 2007; 44(8):498–508.
- Esteban FJ, Cano C, De la Haza I, Cano-Ortiz A, Mendizábal NV, Goñi J, Horcajadas JA. Análisis bioinformático de datos: aplicación en microarrays. *Cuadernos de Medicina Reproductiva*. 2008; 14(1):87–96.
- Feig C, Kirchoff C, Ivell R, Naether O, Schulze W, Spiess AN. A new paradigm for profiling testicular gene expression during normal and disturbed human spermatogenesis. *Molecular Human Reproduction*. 2007; 13(1):33–43.
- Fernald GH, Capriotti E, Daneshjou R, Karczewski KJ, Altman RB. Bioinformatics challenges for personalized medicine. *Bioinformatics (Oxford, England)*. 2011; 27(13):1741–1748.
- Franks S, Mason H, Willis D. Follicular dynamics in the Polycystic Ovary Syndrome. *Molecular and Cellular Endocrinology*. 2000; 163(1-2):49–52.
- Franks S, McCarthy M. Genetics of ovarian disorders: Polycystic Ovary Syndrome. *Reviews in Endocrine and Metabolic Disorders*. 2004; 5(1):69–76.
- Fruzzetti F, Campagna AM, Perini D, Carmina E. Ovarian volumen in normal and hyperandrogenic adolescent women. *Fertility and Sterility*. 2015; 104(1):196–199.
- Galperin MY, Cochrane GR. The 2011 Nucleic acids research database issue and the online molecular biology database collection. *Nucleic Acids Research*. 2011; 39(Database issue):D1–D6.
- Gentleman R, Carey V, Huber W, Irizarry R, Dudoit S. *Bioinformatics and computational biology solutions using R and Bioconductor*. New York: Springer. 2005.
- Glintborg D, Hermann AP, Andersen M, Hagen C, Beck-Nielsen H, Veldhuis JD, Henriksen JE. Effect of pioglitazone on glucose metabolism and luteinizing hormone secretion in women with Polycystic Ovary Syndrome. *Fertility and Sterility*. 2006; 86(2):385–397.

- Groth D, Hartmann S, Klie S, Selbig J. Principal components analysis. *Methods in molecular Biology*. 2013; 930:527–47.
- Hamilton-Fairley D, Kiddy D, Watson H, Paterson C, Franks S. Association of moderate obesity with a poor pregnancy outcome in women with Polycystic Ovary Syndrome treated with low dose gonadotrophin. *British Journal of Obstetrics and Gynaecology*. 1992; 99(2):128–131.
- Handel MA, Schimenti JC. Genetics of mammalian meiosis: regulation, dynamics and impact on fertility. *Nature Reviews. Genetics*. 2010; 11(2):124–136.
- Harris TW, Antoshechkin I, Bieri T, Blasiar D, Chan J, Chen WJ, De La Cruz N, Davis P, Duesbury M, Fang R, Fernandes J, Han M, Kishore R, Lee R, Müller HM, Nakamura C, Ozersky P, Petcherski A, Rangarajan A, Rogers A, Schindelman G, Schwarz EM, Tuli MA, Van Auken K, Wang D, Wang X, Williams G, Yook K, Durbin R, Stein LD, Spieth J, Sternberg PW. WormBase: a comprehensive resource for nematode research. *Nucleic Acids Research*. 2010; 38(Database issue):D463–D467.
- Hickey T, Chandy A, Norman RJ. The androgen receptor CAG repeat polymorphism and X-chromosome inactivation in Australian Caucasian women with infertility related to Polycystic Ovary Syndrome. *The Journal of Clinical Endocrinology and Metabolism*. 2002; 87(1):161–165.
- Hojlund K, Beck-Nielsen H. Impaired glycogen synthase activity and mitochondrial dysfunction in skeletal muscle: markers or mediators of insulin resistance in type 2 diabetes. *Current Diabetes Reviews*. 2006; 2(4):375–395.
- Hwang K, Yatsenko AN, Jorgez CJ, Mukherjee S, Nalam RL, Matzuk MM, Lamb DJ. Mendelian genetics of male infertility. *Annals of the New York Academy of Sciences*. 2010; 1214:E1–E17.
- Jiang D, Tang C, Zhang A. Cluster analysis for gene expression data: A survey. *IEEE Transaction on Knowledge and Data Engineering*. 2004; 16(11):1370–1386.
- Jiao X, Sherman BT, Huang da W, Stephens R, Baseler MW, Lane HC, Lempicki RA. DAVID-WS: a stateful web service to facilitate gene/protein list analysis. *Bioinformatics (Oxford, England)*. 2012; 28(13):1805–1806.

- Johnstone EB, Rosen MP, Neril R, Trevithick D, Sternfeld B, Murphy R, Addauan-Andersen C, McConnell D, Pera RR, Cedars MI. The polycystic ovary post-Rotterdam: a common, age-dependent finding in ovulatory women without metabolic significance. *The Journal of Clinical Endocrinology and Metabolism*. 2010; 95(11):4965–4972.
- Kaur S, Archer KJ, Devi MG, Kriplani A, Strauss JF, Singh R. Differential gene expression in granulosa cells from Polycystic Ovary Syndrome patients with and without insulin resistance: identification of susceptibility gene sets through network analysis. *The Journal of Clinical Endocrinology and Metabolism*. 2012; 97(10):E2016–E2021.
- Kiddy DS, Hamilton-Fairley D, Bush A, Short F, Anyaoku V, Reed MJ, Franks S. Improvement in endocrine and ovarian function during dietary treatment of obese women with Polycystic Ovary Syndrome. *Clinical Endocrinology*. 1992; 36(1):105–111.
- Kiddy DS, Sharp PS, White DM, Scanlon MF, Mason HD, Bray CS, Polson DW, Reed MJ, Franks S. Differences in clinical and endocrine features between obese and non-obese subjects with Polycystic Ovary Syndrome: an analysis of 263 consecutive cases. *Clinical Endocrinology*. 1990; 32(2):213–220.
- Kocabas AM, Crosby J, Ross PJ, Out HH, Beyhan Z, Can H, Tam WL, Rosa GJ, Halgren RG, Lim B, Fernandez E, Cibelli JB. The transcriptome of human oocytes. *Proceedings of the National Academy of Sciences of the United States of America*. 2006; 103(38):14027–14032.
- Lan CW, Chen MJ, Tai KY, Yu D, Yang YC, Jan PS, Yang YS, Chen HF, Ho HN. Functional microarray analysis of differentially expressed genes in granulosa cells from women with Polycystic Ovary Syndrome related to MAPK/ERK signaling. *Scientific Reports*. 2015; 5:14994.
- Lee JR, Magee DM, Gaster RS, LaBaer J, Wang SX. Emerging protein array technologies for proteomics. *Expert Review of Proteomics*. 2013; 10(1):65–75.
- Legro RS. Polycystic Ovary Syndrome and cardiovascular disease: a premature association? *Endocrine Reviews*. 2003; 24(3):302–312.

- Mardis ER. Next-generation DNA sequencing methods. *Annual Review of Genomics and Human Genetics*. 2008; 9(1):387–402.
- Markowski EP, Markowski CA. C39.5. Evaluation of some alternatives to the welch test. *Journal of Statistical Computation and Simulation*. 1993; 44(3–4):262–264.
- Matzuk MM, Lamb DJ. The biology of infertility: research advances and clinical challenges. *Nature Medicine*. 2008; 14(11):1197–1213.
- McCartney ChR, Marshall JC. Polycystic Ovary Syndrome. *The New England Journal of Medicine*. 2016; 375(1):54–64.
- Nandi A, Chen Z, Patel R, Poretsky L. Polycystic Ovary Syndrome. *Endocrinology and Metabolism Clinics of North America*. 2014; 43(1):123–147.
- Nelson-Degrave VL, Wickenheisser JK, Hendricks KL, Asano T, Fujishiro M, Legro RS, Kimball SR, Strauss JF, McAllister JM. Alterations in mitogen-activated protein kinase kinase and extracellular regulated kinase signaling in theca cells contribute to excessive androgen production in Polycystic Ovary Syndrome. *Molecular Endocrinology (Baltimore, Maryland)*. 2005; 19(2):379–390.
- Oberthuer A, Juraeva D, Li L, Kahlert Y, Westermann F, Eils R, Berthold F, Shi L, Wolfinger RD, Fischer M, Brors B. Comparison of performance of one-color and two-color gene-expression analyses in predicting clinical endpoints of neuroblastoma patients. *The Pharmacogenomics Journal*. 2010; 10(4):258–266.
- Okoniewski MJ, Miller CJ. Comprehensive analysis of affymetrix exon arrays using Bioconductor. *PLoS Computational Biology*. 2008; 4(2):e6.
- Orio F Jr, Giallauria F, Palomba S, Cascella T, Manguso F, Vuolo L, Russo T, Tolino A, Lombardi G, Colao A, Vigorito C. Cardiopulmonary impairment in young women with Polycystic Ovary Syndrome. *The Journal of Clinical Endocrinology and Metabolism*. 2006; 91(8):2967–2971.
- Ostermeier GC, Dix DJ, Miller D, Khatri P, Krawetz SA. Spermatozoal RNA profiles of normal fertile men. *Lancet (London, England)*. 2002; 360(9335):772–777.

- Pasquali R, Antenucci D, Casimirri F, Venturoli S, Paradisi R, Fabbri R, Balestra V, Melchionda N, Barbara L. Clinical and hormonal characteristics of obese amenorrheic hyperandrogenic women before and after weight loss. *The Journal of Clinical Endocrinology and Metabolism*. 1989; 68(1):173–179.
- Petersen KF, Shulman GI. Etiology of insulin resistance. *The American Journal of Medicine*. 2006; 119(5 Suppl 1):S10–S16.
- Polson DW, Kiddy DS, Mason HD, Franks S. Induction of ovulation with clomiphene citrate in women with Polycystic Ovary Syndrome: the difference between responders and nonresponders. *Fertility and Sterility*. 1989; 51(1):30–34.
- Primig M. The bioinformatics tool box for reproductive biology. *Biochimica et Biophysica Acta*. 2012; 1822(12):1220–1895.
- Robinson S, Kiddy D, Gelding SV, Willis D, Niththyanathan R, Bush A, Johnston DG, Franks S. The relationship of insulin insensitivity to menstrual pattern in women with hyperandrogenism and polycystic ovaries. *Clinical Endocrinology*. 1993; 39(3):351–355.
- Rung J, Brazma A. Reuse of public genome-wide gene expression data. *Nature Reviews Genetics*. 2013; 14(2):89–99.
- Saeed AI, Sharov V, White JA, Li J, Howe EA, Liang W, Bhagabati NK, Braisted JC, Thiagarajan M, Quackenbush J. TM4 Microarray Software Suite. *Methods in Enzymology*. 2006; 411:134–193.
- Shafiee MN, Chapman C, Barrett D, Abu J, Atiomo W. Reviewing the molecular mechanisms which increase endometrial cancer (EC) risk in women with Polycystic Ovarian Syndrome (PCOS): Time for paradigm shift? *Gynecologic Oncology*. 2013; 131(2):489–492.
- Shulman GI, Rothman DL, Jue T, Stein P, DeFronzo RA, Shulman RG. Quantitation of muscle glycogen synthesis in normal subjects and subjects with non-insulin-dependent diabetes by <sup>13</sup>C nuclear magnetic resonance spectroscopy. *The New England Journal of Medicine*. 1990; 322(4):223–228.

- Skov V, Glintborg D, Knudsen S, Jensen T, Kruse TA, Tan Q, Højlund K. Reduced expression of nuclear-encoded genes involved in mitochondrial oxidative metabolism in skeletal muscle of insulin-resistant women with Polycystic Ovary Syndrome. *Diabetes*. 2007; 56(9):2349–2355.
- Sobie EA, Lee YS, Jenkins SL, Iyengar R. Systems biology--biomedical modeling. *Science Signaling*. 2011; 4(190):tr2.
- Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A, Santos A, Tsafou KP, Kuhn M, Bork P, Jensen LJ, von Mering C. STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Research*. 2015; 43(Database issue):D447–D452.
- The Rotterdam ESHRE/ASRM-Sponsored PCOS Consensus Workshop Group. Revised 2003 consensus on diagnostic criteria and long-term health risks related to Polycystic Ovary Syndrome (PCOS). *Human Reproduction* (Oxford, England). 2004; 19(1):41–47.
- Vargas E, Díaz-Beltrán L, Esteban FJ. Integración y análisis de datos mediante biología de sistemas. 2018. (Sometido).
- Venter JC, Adams MD, Smith HO. The sequence of the human genome. *Science*. 2015; 61(9):1207–1208.
- Virtanen KA, Lidell ME, Orava J, Heglind M, Westergren R, Niemi T, Taittonen M, Laine J, Savisto NJ, Enerbäck S, Nuutila P. Functional brown adipose tissue in healthy adults. *The New England Journal of Medicine*. 2009; 360(15):1518–1525.
- Wang RS, Maron BA, Loscalzo J. Systems Medicine: Evolution of systems biology from bench to bedside. *Wiley Interdisciplinary Reviews Systems Biology and Medicine*. 2015; 7(4):141–161.
- Wheeler DA, Srinivasan M, Egholm M, Shen Y, Chen L, McGuire A, He W, Chen YJ, Makhijani V, Roth GT, Gomes X, Tartaro K, Niazi F, Turcotte CL, Irzyk GP, Lupski JR, Chinault C, Song XZ, Liu Y, Yuan Y, Nazareth L, Qin X, Muzny DM, Margulies M, Weinstock GM, Gibbs RA, Rothberg JM. The complete

genome of an individual by massively parallel DNA sequencing. *Nature*. 2008; 452(7189):872–876.

White DM, Polson DW, Kiddy D, Sagle P, Watson H, Gilling-Smith C, Hamilton-Fairley D, Franks S. Induction of ovulation with low-dose gonadotropins in Polycystic Ovary Syndrome: an analysis of 109 pregnancies in 225 women. *The Journal of Clinical Endocrinology and Metabolism*. 1996; 81(11):3821–3824.

Willis DS, Watson H, Mason HD, Galea R, Brincat M, Franks S. Premature response to luteinizing hormone of granulosa cells from anovulatory women with Polycystic Ovary Syndrome: relevance to mechanism of anovulation. *The Journal of Clinical Endocrinology and Metabolism*. 1998; 83(11):3984–3991.

Yuan X, Hu T, Zhao H, Huang Y, Ye R, Lin J, Zhang C, Zhang H, Wei G, Zhou H, Dong M, Zhao J, Wang H, Liu Q, Lee HJ, Jin W, Chen ZJ. Brown adipose tissue transplantation ameliorates Polycystic Ovary Syndrome. *Proceedings of the National Academy of Sciences of the United States of America*. 2016; 113(10):2708–2713.