



UNIVERSIDAD DE JAÉN
Facultad de Ciencias Experimentales

Trabajo Fin de Grado

**Caracterización de la
expresión diferencial de
microRNAs mediante análisis
bioinformático**

Alumno: Silvia Hernández Rosales

Junio, 2017



UNIVERSIDAD DE JAÉN
Facultad de Ciencias Experimentales

Trabajo Fin de Grado

Caracterización de la expresión diferencial de microRNAs mediante análisis bioinformático

Alumno: Silvia Hernández Rosales

Junio, 2017

ÍNDICE

ABSTRACT	5
RESUMEN.....	5
1. INTRODUCCIÓN.....	7
2. OBJETIVOS	11
3. MATERIAL Y MÉTODOS	13
3.1. Obtención de datos.....	13
3.2. Análisis exploratorio.....	14
3.3. Pre-procesamiento	15
3.4. Análisis estadístico	15
3.4.1. Análisis paramétrico	15
3.4.2. Análisis no paramétrico	15
3.5. Búsqueda de genes diana	16
3.6. Análisis funcional	16
3.7. Clasificadores y predictores.....	16
3.8. Esquema general	17
4. RESULTADOS	19
4.1. Análisis exploratorio.....	19
4.2. Análisis paramétrico	20
4.3. Análisis no paramétrico	20
4.4. Detección de genes diana	22
4.5. Análisis funcional	23
4.6. Clasificadores y predictores.....	23
5. DISCUSIÓN.....	27
6. CONCLUSIONES	29
7. BIBLIOGRAFÍA.....	31

ABSTRACT

The aim of this work is the analysis of microRNAs, key molecules that regulate the gene expression, using bioinformatic tools. To achieve this, and as an example, Systems Biology approaches were carried out to study endometriosis, one of the main causes of female infertility. Taken together, the main objectives are learning how to manage these analytical tools together to the possible detection of disease biomarkers related to the differential expression of microRNAs in women with endometriosis, when compared to healthy ones.

RESUMEN

El presente proyecto se basa en la aplicación de herramientas bioinformáticas para el análisis de la expresión de microRNAs, importantes moléculas reguladoras de la expresión génica. Para ello se han utilizado procedimientos propios de la Biología de Sistemas aplicados, como ejemplo, al estudio de la endometriosis, una de las principales causas de infertilidad femenina. Los principales objetivos son aprender a manejar estas herramientas de análisis y la detección de posibles biomarcadores de enfermedad a través de la detección de diferencias entre los perfiles de expresión de microRNAs de mujeres con endometriosis y mujeres sanas.

1. INTRODUCCIÓN

Los microRNAs (miRNAs) son moléculas de ARN de muy pequeño tamaño (constan de entre 20 y 25 nucleótidos) que impiden la síntesis de proteínas por parte de los ribosomas (Vergoulis et al., 2015). Su mecanismo de acción (Figura 1) se debe al hecho de que el microRNA se adhiere antes al ARN mensajero (ARNm) que el ARN de transferencia (ARNt), lo que provoca o bien que las nucleasas lo corten y quede inutilizado, o que el complejo microRNA-ARNm no sea reconocido por el ribosoma y, por lo tanto, no se sintetice la proteína.

Los microRNAs tienen un papel muy importante en procesos biológicos de regulación génica, y se consideran reguladores clave de la expresión génica a nivel post-transcripcional (Siristatidis et al., 2015). Una de las aplicaciones más prometedoras que se deriva de la propia naturaleza de estas pequeñas moléculas es el hecho de que pueden ser utilizados para inactivar determinados genes responsables de la aparición de una enfermedad. Cabe indicar que la expresión génica es el proceso mediante el cual una célula genera un producto génico a partir de la secuencia de nucleótidos existente en el ADN. Este proceso es necesario para la vida y cuenta con mecanismos complejos para su regulación (Watson et al., 2014).

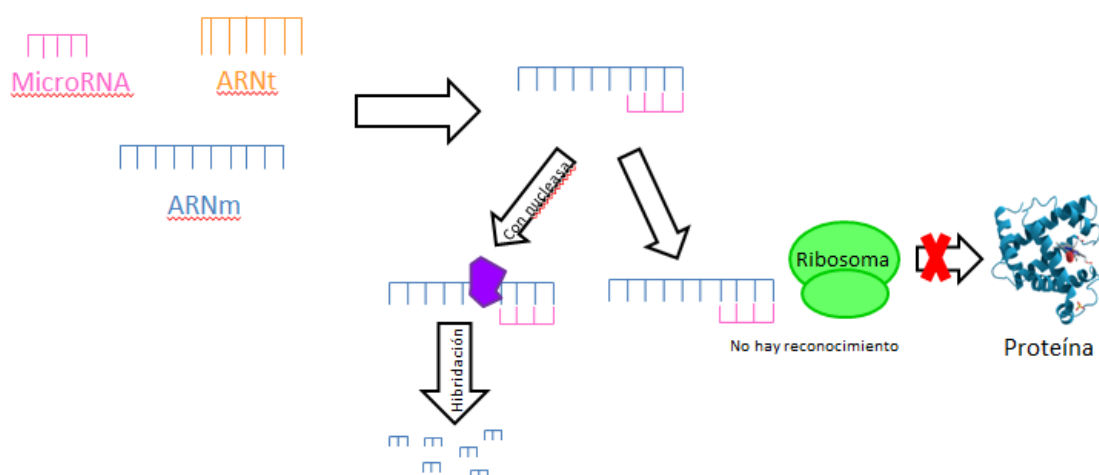


Figura 1. Esquema representativo del mecanismo de acción de los microRNAs.

La infertilidad constituye un problema social que afecta a una de cada seis parejas. Puede ser definida como la incapacidad de completar un embarazo después de un año de relaciones sexuales sin tomar medidas anticonceptivas (Brugo-Olmedo et al., 2003). Uno de los factores con mayor repercusión en la infertilidad es la edad a la que la pareja se siente preparada para tener hijos, lo que supone que cualquier factor clínico se pueda ver más intensificado; a partir de los 34-35 años la capacidad reproductiva de la mujer se ve muy afectada debido a que la presencia y calidad de los óvulos es cada vez menor.

Desde el punto de vista clínico, el estudio de la infertilidad puede abordarse atendiendo a diferentes factores. Si nos centramos en las alteraciones de la pareja infértil, destacan el proceso de ovulación, el estado del útero y la capacidad de migración del semen (Brugo-Olmedo et al., 2003). La Biología de la Reproducción es una disciplina de la Biología y de la Biomedicina que trata de dar respuesta a los problemas relacionados con la infertilidad y que ha experimentado una gran evolución en los últimos años. Entre las enfermedades que pueden causar infertilidad se encuentra la endometriosis, padecida por aproximadamente el 10% de las mujeres en edad reproductiva (Baranov et al., 2015).

La endometriosis es una enfermedad benigna, crónica, que solo afecta a mujeres. Podría definirse como la presencia de células endometriales fuera de la cavidad endometrial, en localizaciones como el ovario, útero o cavidad peritoneal. Este tejido tiene características especiales, ya que es capaz de implantarse, invadir tejidos adyacentes, proliferar y hacer que se diseminen las células a través de los vasos sanguíneos y linfáticos. La enfermedad puede producirse en un gran rango de edades, desde muy tempranas hasta la aparición de la menopausia (Makiyan, 2017).

Los síntomas asociados a este trastorno son principalmente el dolor pélvico, la dismenorrea y la infertilidad, aunque en determinadas ocasiones puede ser asintomática, lo que complica el diagnóstico. El tratamiento y el diagnóstico de la endometriosis están delimitados por la gravedad de la enfermedad. Por lo tanto, es necesario establecer una clasificación que tenga en cuenta la localización y la intensidad de dicha afección, debido a que son dos factores

que influyen en el diagnóstico. La clasificación más aceptada divide la endometriosis en leve, moderada, severa y extensa, valorando la presencia de implantes y/o adherencias a peritoneo, ovarios o trompas (Sociedad Española de Ginecología y Obstetricia, 2014).

En la actualidad, se utilizan como métodos de diagnóstico de la endometriosis la exploración física, los ultrasonidos o la tomografía axial computerizada (TAC), aunque el método más común para el diagnóstico de las lesiones endometriales es la laparoscopia, un procedimiento en el cual se examina la cavidad abdominal mediante una incisión quirúrgica en la piel, por debajo del ombligo, introduciendo un laparoscopio (instrumento de visualización). Este procedimiento, aunque conlleva molestias, tiene una recuperación más rápida frente a la cirugía abierta, método inequívoco de diagnóstico, pues perjudica en menor medida al paciente y facilita su recuperación (Gaitán et al., 2014).

Aún no se conoce la causa por la que se produce la endometriosis, por lo que se ha intentado abordar su estudio desde diferentes puntos de vista: análisis histológico, proteómica, transcriptómica y epigenómica. Por el momento, no existe ningún tratamiento óptimo que consiga alcanzar los 4 objetivos básicos: suprimir los síntomas (algias pélvicas, dismenorrea severa, etc.), restaurar la fertilidad, eliminar la endometriosis visible y evitar la progresión de la enfermedad (Sociedad Española de Ginecología y Obstetricia, 2014).

Un abordaje interesante en el análisis de las alteraciones moleculares subyacentes a la endometriosis es la búsqueda de diferencias en los perfiles de expresión de microRNAs entre mujeres con endometriosis y sanas. En este sentido, se ha detectado que el endometrio ectópico (aparición de endometrio fuera de la cavidad uterina) de mujeres con endometriosis, en comparación con las que no la padecen, presenta alteraciones moleculares entre las que se incluyen la expresión diferencial de microRNAs (Laudanski et al., 2013). En estos análisis son de gran utilidad las herramientas bioinformáticas.

La bioinformática es una disciplina científica que utiliza la tecnología computacional para realizar un análisis masivo de datos experimentales y, de esta forma, obtener información gracias a la cual responder a cuestiones biológicas (Esteban et al., 2008). Los datos experimentales obtenidos deben

ser procesados y analizados de modo adecuado. Para llevar a cabo el análisis de datos es necesario partir de una matriz de datos normalizados. A continuación se pueden incluir análisis exploratorios, análisis estadísticos, búsqueda de los genes diana, análisis funcionales y la detección de predictores y clasificadores. Todo esto permitirá distinguir la expresión diferencial en la situación biomédica objeto de estudio. Como abordaje de integración de análisis, es de gran ayuda la Biología de Sistemas, que nos permite profundizar en las interacciones moleculares y su relación con los procesos biológicos (Altmäe et al., 2014) lo cual, a su vez, puede apuntar hacia la detección de biomarcadores propios de la enfermedad.

Un biomarcador es un elemento clave que permite evaluar el estado biológico, en este caso si el paciente está sano o padece la enfermedad. Tienen especial interés en el diagnóstico de enfermedades que no muestran síntomas hasta que la enfermedad está avanzada, como es el caso de la endometriosis.

Hay varios tipos de biomarcadores, a saber: biomarcador de exposición, biomarcador de susceptibilidad y biomarcadores de respuesta o efecto. Los biomarcadores de susceptibilidad sirven como indicadores de sensibilidad individual al efecto de un xenobiótico o grupo de compuestos tóxicos. Un xenobiótico es una estructura química sintética que ha sido desarrollada en el laboratorio; se considera un biomarcador de exposición a un compuesto exógeno (o un metabolito) dentro del organismo que refleja la exposición de éste a un xenobiótico. El biomarcador de respuesta o efecto, por su parte, es indicativo de cambios bioquímicos en un organismo como resultado de la exposición a xenobióticos (Gil Hernández, 2000).

2. OBJETIVOS

El objetivo general de este Trabajo Fin de Grado es utilizar las herramientas bioinformáticas propias de la Biología de Sistemas para detectar diferencias en los niveles de expresión de microRNAs entre muestras de endometrio sano y endometriosis, de modo que pueda obtenerse información acerca de los procesos biológicos que están relacionados con este desorden ginecológico.

De este objetivo general se derivan una serie de objetivos específicos:

1. Conocer las herramientas bioinformáticas propias de la Biología de Sistemas.
2. Caracterizar niveles de expresión de microRNAs en mujeres sanas y con endometriosis.
3. Detectar los genes diana sobre los que actúan los microRNAs implicados en este trastorno.
4. Obtener los procesos biológicos relacionados con los genes diana de los microRNAs.
5. Generar clasificadores y predictores de endometriosis a partir de los resultados anteriores.

3. MATERIAL Y MÉTODOS

3.1. Obtención de datos

Se analizaron los datos del experimento GSE56414, obtenidos de la base de datos de acceso público *Gene Expression Omnibus* (GEO) (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE56414>), que corresponde a un estudio del perfil de expresión de microRNAs en mujeres con endometriosis y mujeres sanas. Dicho perfil de expresión se obtuvo utilizando el kit *Illumina Genome Analyzer* de la casa comercial Illumina, que permite la detección de la expresión de más de 2000 microRNAs. Para realizar nuestros análisis, partimos de los datos previamente normalizados por los investigadores que llevaron a cabo el procedimiento experimental (Saare et al., 2014).

El grupo de datos objeto de análisis consta de 11 muestras de tejido procedentes de dos pacientes con un diagnóstico de endometriosis moderada-severa (fase III-IV) confirmado histológicamente. De estas 11 muestras, 2 son de tejido endometrial eutópico, 5 de lesiones peritoneales y las 4 restantes corresponden a tejido adyacente aparente normal. Los fenotipos de las pacientes se recogen en la Tabla 1.

Código	Grupo	Localización del tejido
E101_1	Experimental (e1)	Endometrio eutópico de paciente con endometriosis
E101_2	Experimental (e2)	Lesión peritoneal endometriótica
E101_3	Control (c1)	Tejido peritoneal adyacente sano
E101_4	Experimental (e3)	Lesión peritoneal endometriótica
E47_1	Experimental (e4)	Endometrio eutópico de paciente con endometriosis
E47_3	Experimental (e5)	Lesión peritoneal endometriótica
E47_4	Control (c2)	Tejido peritoneal adyacente sano
E47_5	Experimental (e6)	Lesión peritoneal endometriótica
E47_6	Control (c3)	Tejido peritoneal adyacente sano
E47_8	Experimental (e7)	Lesión peritoneal endometriótica
E47_9	Control (c4)	Tejido peritoneal adyacente sano

Tabla 1. Tipo de muestras que integran el conjunto de datos.

Para realizar los análisis, las muestras fueron divididas en dos grupos:

- Control (n = 4): muestras de tejido peritoneal adyacente sano (E101_3, E47_4, E47_6 y E47_9).
- Experimental (n = 7): muestras de lesión peritoneal endometriótica y endometrio eutópico de pacientes con endometriosis (E101_1, E101_2, E101_4, E47_1, E47_3, E47_5, E47_8).

3.2. Análisis exploratorio

En primer lugar, realizamos un análisis exploratorio con el objetivo de detectar la existencia de algún patrón de distribución en nuestros datos. Para ello, utilizamos como herramienta el programa libre MeV-TM4 (*Multiple Experiment Viewer – TM4 Microarray Software Suite*) (<http://www.jcvi.org/cms/research/software/>) (Saeed et al., 2006) y como técnicas de análisis exploratorio, el agrupamiento jerárquico (*clustering*) y el análisis de componentes principales (PCA).

La primera de ellas consiste en hacer agrupaciones de los perfiles que más se parecen entre sí, atendiendo a los valores de expresión, con la generación de una estructura en forma de árbol denominada dendrograma. Todas las “ramas” de este árbol aportan información acerca de la manera en la que se encuentran agrupadas las muestras, de modo que las distancias entre ellas generan grupos de similitud. Como parámetro de distancia se utilizó la euclídea y, como método de agrupamiento, el de tipo completo (*complete linkage*).

El PCA es una técnica estadística multivariante, que permite reducir la información contenida en un número elevado de variables a un número inferior de las mismas, perdiendo la menor cantidad de información posible. Así pues, se obtienen combinaciones lineales de variables independientes entre sí, y que pueden definir grupos homogéneos (Salinas, 2006). Desde el punto de vista gráfico, aporta información exploratoria relacionada con la similitud de las muestras, en nuestro caso debido a los niveles de expresión de los microRNAs.

3.3. Pre-procesamiento

Como parte del procesamiento de los datos, se llevó a cabo una transformación a escala logarítmica de los mismos, que es el modo más simple de estabilización de la varianza y además permite uniformar la escala en las tasas de expresión (Esteban et al., 2008), comprimiendo de esta manera los valores altos y expandiendo los pequeños.

3.4. Análisis estadístico

Con el objetivo de detectar microRNAs diferencialmente expresados entre las muestras objeto de estudio, se aplicaron dos tipos de tests estadísticos: paramétricos y no paramétricos. Todos los análisis se realizaron utilizando el software TM4 y, de manera adicional, los procedimientos fueron replicados en *R Statistics Software* (<http://cran.us.r-project.org/>).

3.4.1. Análisis paramétrico

La prueba estadística elegida para llevar a cabo esta parte del análisis fue el test de Welch, un tipo de test paramétrico que asume que las varianzas entre los grupos son desiguales, detectando así diferencias en la expresión de miRNAs entre dos tipos de muestras. Se considerarán como diferencialmente expresados aquellos microRNAs cuyo p-valor corregido (FDR: *false discovery rate*) sea inferior a 0'05 y, al mismo tiempo, cuyo valor absoluto de la tasa de cambio (FC: *fold change*) sea superior a 1'5.

3.4.2. Análisis no paramétrico

En lo que respecta a los métodos no paramétricos se escogió el *Rank Product* (<https://www.bioconductor.org/packages/release/bioc/html/RankProd.html>), un método no paramétrico que se basa en obtener los genes expresados mediante el establecimiento de rangos y umbrales de significación fiables. Como criterio para considerar un microRNA como diferencialmente expresado, decidimos que lo será aquel cuyo valor de pfp (*porcentaje de falsos positivos*) sea inferior a 0'05 y cuyo valor absoluto de la tasa de cambio sea superior a 1'5.

3.5. Búsqueda de genes diana

Una vez obtenidos los miRNAs diferencialmente expresados, se procedió a la detección de los genes diana correspondientes. Para ello, se hizo uso de miRDB (<http://mirdb.org/miRDB/index.html>), una base de datos online para la predicción de dianas y anotación funcional de microRNAs (Wong & Wang, 2015; Wang, 2016). Todas las dianas en miRDB son predichas por una herramienta bioinformática, MirTarget, desarrollada mediante el análisis de miles de interacciones de microRNAs y dianas en experimentos de secuenciación de alta resolución. Además, una de las grandes ventajas de esta base de datos es que permite detectar de manera simultánea los genes diana de un grupo dado de microRNAs, en lugar de hacerlo de forma individual.

Para la búsqueda de genes diana se utilizaron los parámetros por defecto a excepción del *score*: de forma general, la base de datos excluye aquellos genes que tienen una puntuación de predicción inferior a 60 y, en nuestro caso, este valor fue modificado a una puntuación de 90, con el objetivo de restringir el número de dianas.

3.6. Análisis funcional

Los listados de genes diana de los microRNAs diferencialmente expresados fueron utilizados para llevar a cabo un análisis funcional, con el objetivo de identificar los procesos biológicos en los que participan dichos genes. Para ello, hicimos uso de la base de datos de acceso público DAVID (*Database for Annotation, Visualization and Integrated Discovery*; <https://david.ncifcrf.gov/>) (Huang et al., 2009).

3.7. Clasificadores y predictores

Los clasificadores y predictores son algoritmos encargados de detectar y agrupar las muestras en distintas categorías en función de las características de estas. Una vez que se obtuvieron los microRNAs diferencialmente expresados entre ambos grupos experimentales, y con el objetivo de generar clasificadores y predictores de lesiones endometriales, se utilizó el software libre WEKA (<http://www.cs.waikato.ac.nz/ml/weka/>) (Frank et al., 2004), una

herramienta de minería de datos. Como parámetro de clasificación nos basamos en el área bajo la curva ROC, el cual es de aplicación biomédica general al reflejar tanto la sensibilidad como la especificidad de clasificación (<https://www.bioestadistica.uma.es/analisis/roc1/>).

3.8. Esquema general

En la Figura 2 se representa el esquema general del procesamiento seguido.

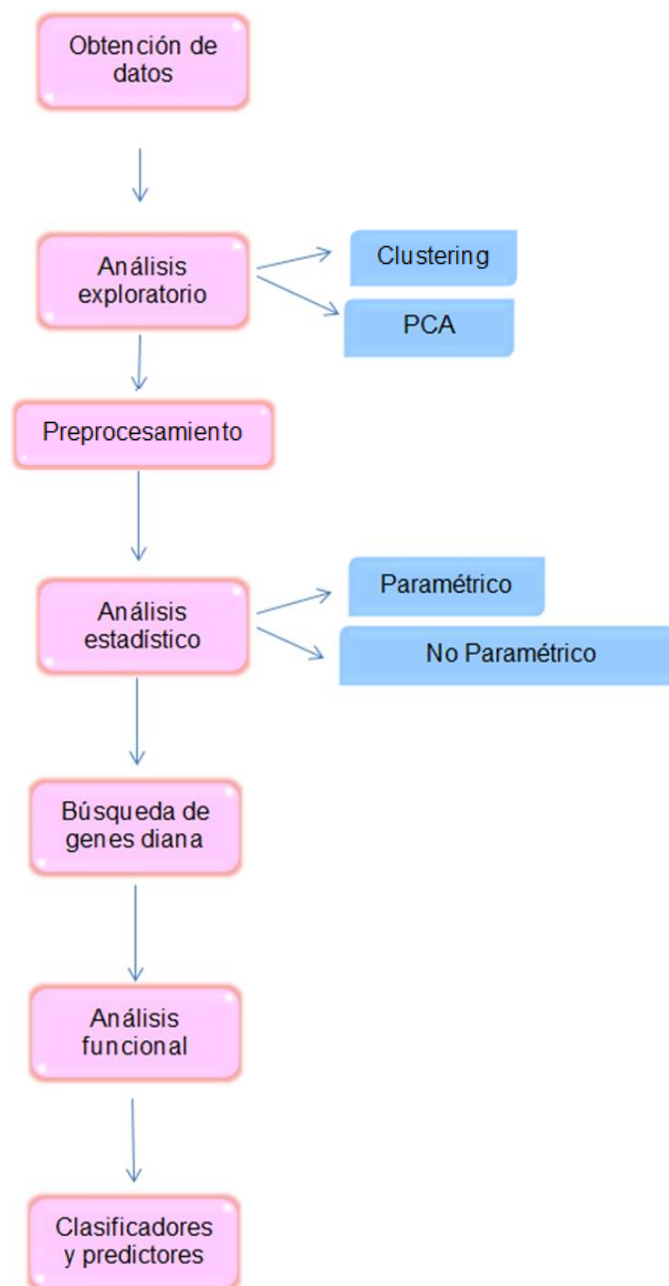


Figura 2. Esquema de trabajo general en un análisis siguiendo una estrategia de Biología de Sistemas. PCA: análisis de componentes principales.

4. RESULTADOS

4.1. Análisis exploratorio

El objetivo del análisis exploratorio es tratar de detectar patrones de similitud de las muestras. Este análisis se llevó a cabo mediante la realización de un agrupamiento jerárquico y un análisis de componentes principales.

Los resultados del agrupamiento jerárquico (Figura 3) muestran la existencia de un patrón de distribución entre las muestras control y las muestras de pacientes con endometriosis, si bien el agrupamiento no es completamente diferencial.

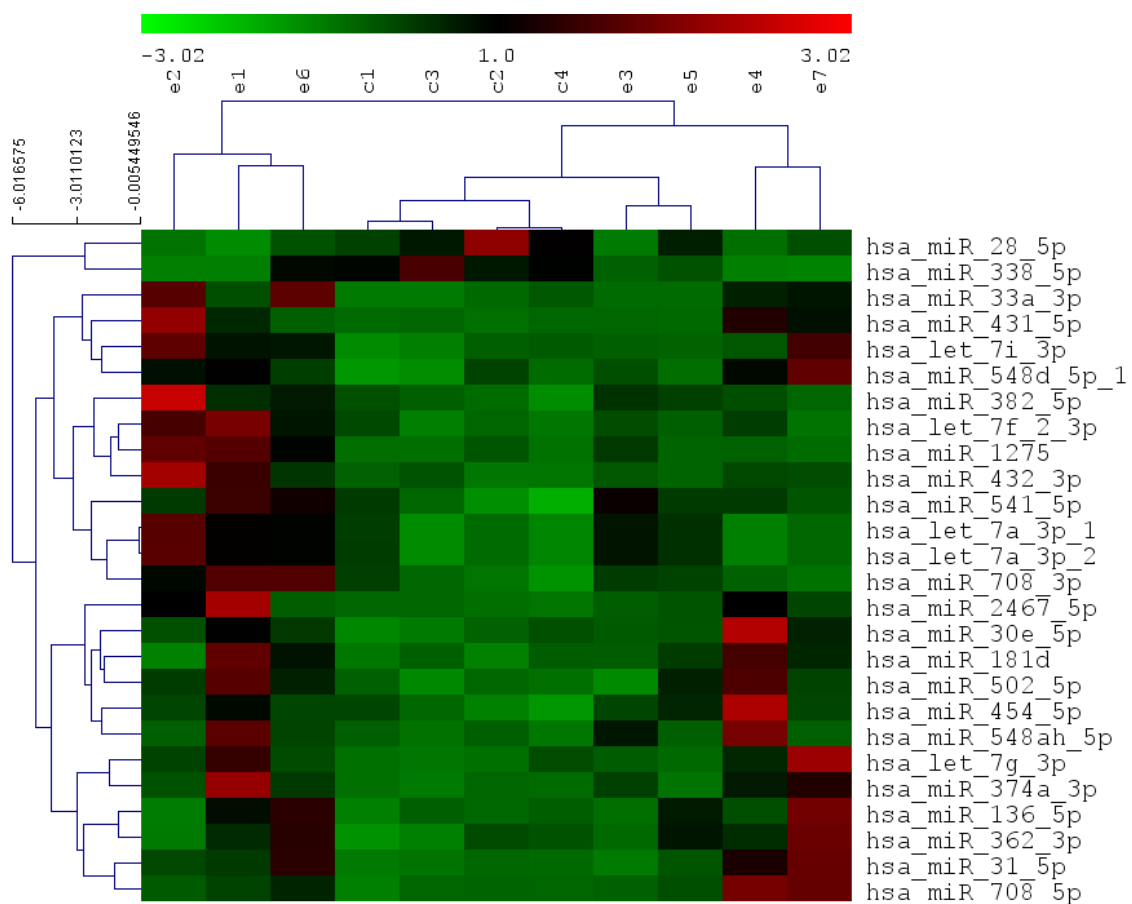


Figura 3. Resultado de la realización de un cluster jerárquico, en el cual se detecta cómo las muestras se agrupan según el perfil de expresión de microRNAs significativos por el test de Welch para p -valores < 0.05 no corregidos.

El análisis de componentes principales (Figura 4), confirma la no existencia de un patrón de distribución claro entre las muestras control y las muestras de pacientes con endometriosis.

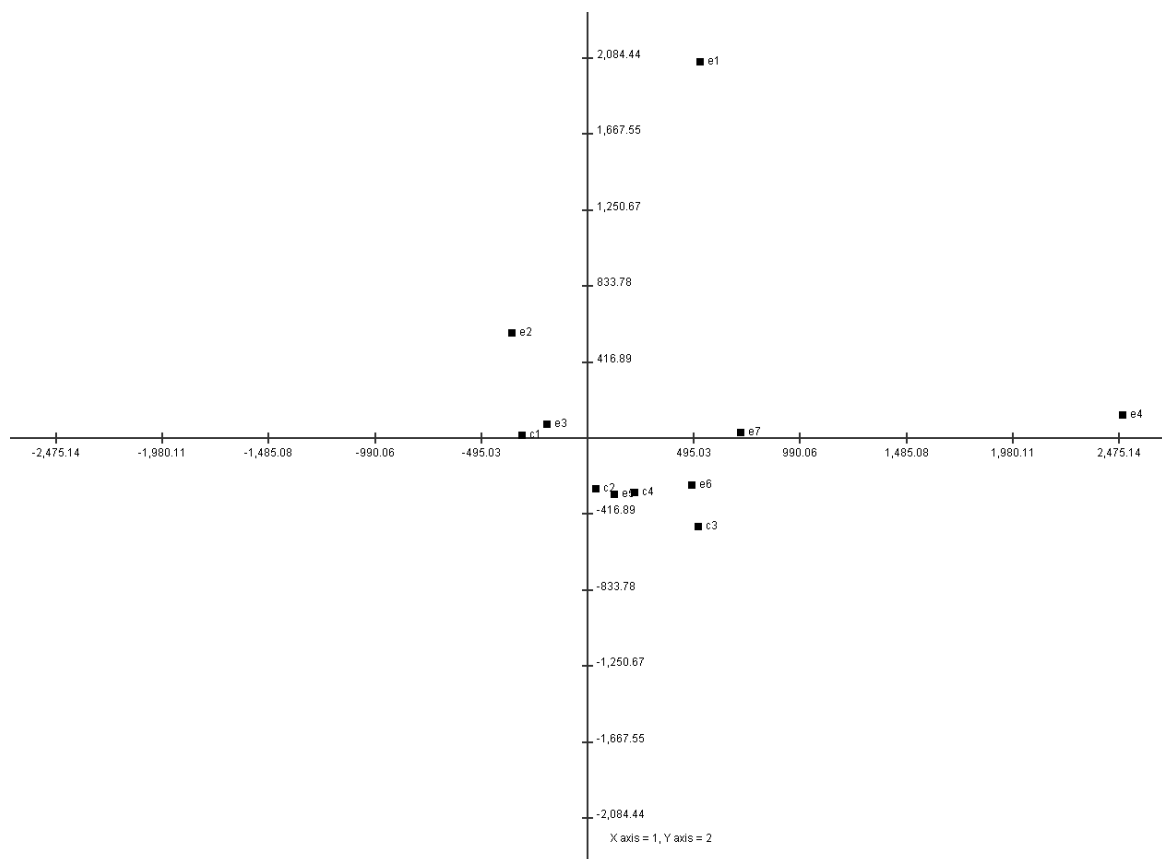


Figura 4. Resultado de la realización de un análisis de componentes principales (PCA), en el cual se agrupan las muestras sin seguir un patrón específico.

4.2. Análisis paramétrico

Al aplicar el test de Welch, no se obtuvo ningún microRNA diferencialmente expresado que se ajustase a los criterios que habíamos establecido (p-valor corregido inferior a 0'05 y valor absoluto de la tasa de cambio superior a 1'5).

4.3. Análisis no paramétrico

Al realizar el análisis no paramétrico mediante el test *Rank Product*, se obtuvieron 11 microRNAs diferencialmente expresados que se ajustasen a los criterios establecidos (pfp inferior a 0'05 y valor absoluto de la tasa de cambio superior a 1'5). En la Tabla 2 se muestra el listado completo de microRNAs y los valores de ambos parámetros. La tasa de cambio se obtiene a través del cálculo de la diferencia de los niveles medios de expresión entre los grupos.

MicroRNA	Tasa de cambio	p-valor corregido(pfp)
hsa-miR-449a	5,27	0,02
hsa-miR-34c-5p	4,01	<0.01
hsa-miR-141-3p	3,99	0,02
hsa-miR-203	2,97	0,02
hsa-miR-183-5p	2,62	0,05
hsa-miR-21-3p	2,22	<0.01
hsa-miR-182-5p	2,13	0,04
hsa-miR-210	2,06	<0.01
hsa-miR-452-5p	1,82	<0.01
hsa-miR-542-3p	1,79	0,01
hsa-miR-142-3p	1,53	0,02

Tabla 2. Grupo de microRNAs diferencialmente expresados tras aplicar el test estadístico no paramétrico Rank Product.

miRNA	Control	Experimental
hsa-miR-449a	0,35±0,20	68,22±43,80
hsa-miR-34c-5p	10,29±3,55	566,31±252,78
hsa-miR-141-3p	12,80±8,95	695,12±479,63
hsa-miR-203	2,93±0,57	57,41±31,61
hsa-miR-183-5p	23,32±7,49	319,22±216,50
hsa-miR-21-3p	395,58±95,36	3624,47±2739,60
hsa-miR-182-5p	46,00±11,99	387,12±199,35
hsa-miR-210	40,26±14,52	316,35±208,59
hsa-miR-452-5p	280,75±67,35	1724,45±1160,56
hsa-miR-542-3p	58,97±15,90	352,75±182,96
hsa-miR-142-3p	15,81±8,55	73,23±24,81

Tabla 3. Valores medios de expresión y errores estándar de los 11 microRNAs diferencialmente expresados. Control: media control ± error estándar grupo control; Experimental: media experimental ± error estándar grupo experimental.

Todos los análisis posteriores se llevan a cabo teniendo en cuenta solo este grupo de microRNAs, pues consideramos que representan de manera

significativa los cambios que tienen lugar en el perfil de expresión de un tipo de muestras frente a otro. Dichos cambios aparecen reflejados en la Figura 5, que representa los valores medios de expresión para cada microRNA en ambos grupos experimentales. Para la realización de dicha gráfica, se tomaron como valores la media de expresión para cada microRNA. Tal y como puede observarse tanto en la Tabla 3 como en la Figura 5, todos los microRNAs que detectamos como diferencialmente expresados aumentan su expresión en las muestras procedentes de tejido dañado por la endometriosis.

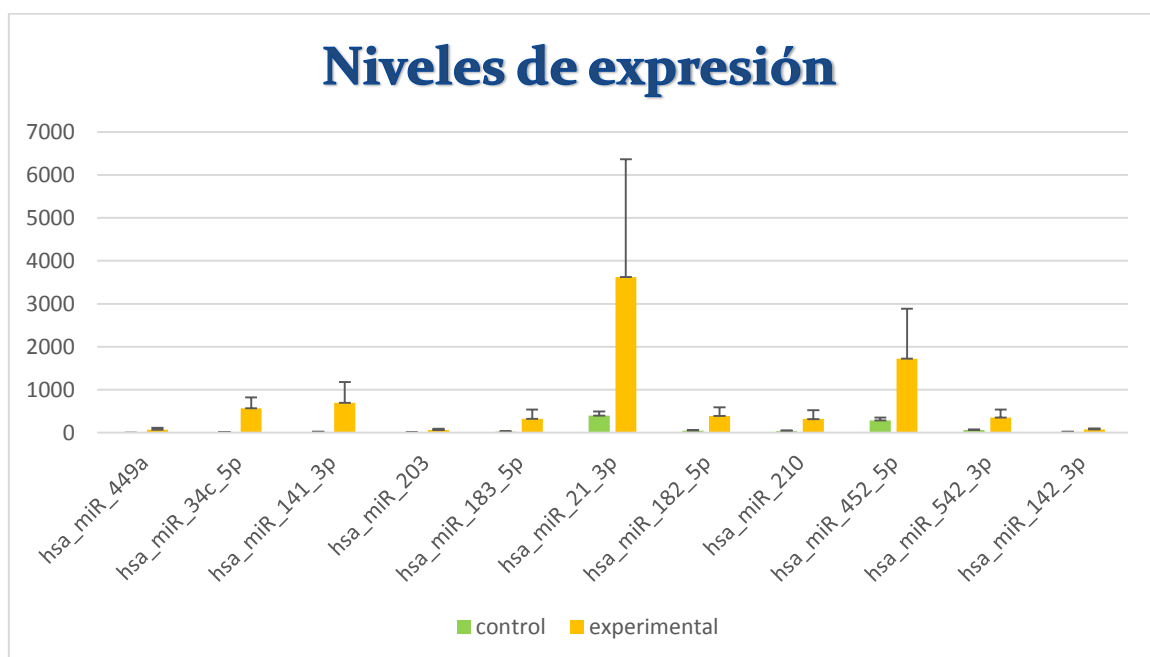


Figura 5. Representación de los niveles medios de expresión de cada microRNA en cada grupo experimental. Las barras de error representan el error estándar de los valores.

4.4. Detección de genes diana

Tras obtener los microRNAs diferencialmente expresados, se procedió a la identificación de los genes diana de los mismos. Para ello, se utilizó la base de datos miRDB. Se hicieron comprobaciones del número de genes diana que se obtenían si se modificaba el parámetro score (indicador de la precisión de predicción) y, como puede observarse en la Tabla 4, a medida que aumenta esta puntuación el número de genes diana va disminuyendo, pues aumenta la

especificidad de la secuencia diana para cada microRNA. Así pues, consideramos finalmente 682 genes diana para el posterior análisis funcional.

Score	60	75	90
Número de genes diana	2870	1697	682

Tabla 4. Número de genes diana obtenidos para cada puntuación de score en miRDB. Como se puede observar, a medida que se va modificando el parámetro score, la cantidad de genes diana cambia, siendo menor el número conforme aumenta el parámetro.

4.5. Análisis funcional

A partir del grupo de 682 genes obtenidos como resultado de la búsqueda de dianas de los 11 microRNAs diferencialmente expresados, se llevó a cabo un análisis funcional con el objetivo de conocer los procesos biológicos en los que se encuentran involucrados este grupo de genes.

Como se puede apreciar en la Tabla 5, los genes diana principalmente están relacionados con procesos de transcripción, proliferación celular y desarrollo embrionario.

4.6. Clasificadores y predictores

Para la detección de microRNAs como clasificadores y predictores se utilizaron dos grupos de microRNAs (Tabla 6):

- Grupo “mirs_11signif”: Corresponde al grupo completo de 11 microRNAs obtenidos como diferencialmente expresados al aplicar el test de Rank Product.
- Grupo “mirs_3signif”: un subconjunto del grupo a), integrado por 3 microRNAs (miR-449a, miR-34c-5p y miR-141-3p), que son los únicos que coinciden con los también detectados como diferencialmente expresados por parte de los autores de los experimentos en los que nos

basamos. Dichos autores emplearon estos 3 microRNAs para la validación experimental mediante técnicas de PCR (Saare et al., 2014).

En la Tabla 6 se muestra cómo tanto los algoritmos “LogitBoost” como “MultiClassClassifier” generan los mayores valores de área bajo la curva ROC (sensibilidad y especificidad). En la Tabla 7 se muestran los resultados detallados de clasificación exacta para el algoritmo “MultiClassClassifier”.

Proceso biológico	p-valor	Genes implicados
Regulación negativa de la transcripción del promotor de la ARN polimerasa II	$7,47 \cdot 10^{-7}$	CREBRF, MEF2C, TSHZ3, FRK, FGF9, WWC2, ZEB2, FOXO3, ZEB1, GLI2, TAGLN3, ATF2
Regulación positiva de la transcripción del promotor de la ARN polimerasa II	$3,53 \cdot 10^{-6}$	CCNT2, MEF2C, E2F5, MORF4L2, ZEB2, FOXO3, ZEB1, GLI2, FOXO4, SETX, SIN3A
Regulación negativa de la proliferación celular	$1,95 \cdot 10^{-5}$	DLC1, FRK, KMT2A, STRN, ZEB1, FOXO4, CBFA2T3, TGFB2, SPRY2, CUL5, PKD2
Adhesión célula-célula	$5,81 \cdot 10^{-5}$	RTN4, FLRT3, FMNL2, TWF1, LDHA, ZC3H15, KIF5B, DOCK9, YWHAB, PKN2, KTN1
Regulación negativa de la adhesión célula-matriz	$7,35 \cdot 10^{-5}$	CORO1C, DLC1, ARHGAP6, CLASP2, THBS1, SRC
Detención del ciclo celular	$7,51 \cdot 10^{-5}$	KMT2E, LAMTOR3, KAT2B, TGFB1, FOXO4, TCF7L2, TGFB2, BRINP1, CUL5, CDKN1B
Regulación negativa de la vía de señalización apoptótica extrínseca	$1,07 \cdot 10^{-4}$	CTTN, ITGA6, ITGAV, TGFB1, SGMS1, THBS1, TCF7L2, SRC
Secreción del glutamato	$1,40 \cdot 10^{-4}$	SYT1, SLC1A2, GLS, NTRK2, PPFIA1, VAMP2, UNC13B
Desarrollo embrionario del útero	$1,41 \cdot 10^{-4}$	GNA13, ZFAND5, ELL, TGFB1, GABPA, MBNL1, KIDINS220, GLI2, ITGB1, FOXP1
Diferenciación osteoblástica	$2,51 \cdot 10^{-4}$	MEF2C, DNAJC13, FGF9, MRC2, CREB3L1, LEF1, VCAN, GLI2, WWTR1

Tabla 5. Principales procesos biológicos en los que se encuentran implicados los genes diana de los 11 microRNAs diferencialmente expresados. FDR: false discovery rate.

Tester:	weka.experiment.PairedCorrectedTester										
Analysing:	Area_under_ROC										
Datasets:	2										
Resultsets:	11										
Confidence:	0.05 (two tailed)										
Sorted by:	-										
Date:	3/31/17 10:10 AM										
Dataset	(1) meta	(2) meta	(3) meta	(4) meta	(5) meta	(6) meta	(7) meta	(8) meta	(9) meta	(10) met	(11) met
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
mirs_11signif	(10) 0.95	0.80	0.50 *	0.60 *	0.50 *	0.80	0.95	1.00	0.95	0.80	0.50 *
mirs_3signif	(10) 0.95	0.80	0.50 *	0.60 *	0.50 *	0.80	0.95	1.00	0.95	1.00	0.50 *
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
	(v/ *)	(0/2/0)	(0/0/2)	(0/0/2)	(0/0/2)	(0/2/0)	(0/2/0)	(0/2/0)	(0/2/0)	(0/2/0)	(0/0/2)

Tabla 6. Resultados de la generación de clasificadores y predictores aplicando diferentes algoritmos mediante el software WEKA. En cada columna se indica la puntuación tipo área bajo la curva ROC para cada uno de los algoritmos. (1): AdaBoost; (2): AttributeSelectedClassifier; (3): Bagging; (4): ClassificationViaClustering; (5): ClassificationViaRegression; (6): END; (7): Filtered.Classifier; (8): LogitBoost; (9): MultiBoostAB; (10): MultiClassClassifier; (11): MultiScheme.

```
MultiClassClassifier

Classifier 1
Logistic Regression with ridge parameter of 1.0E-8
Coefficients...
      Class
variable   c
=====
hsa_mir_449a -0.9998
hsa_mir_34c_5p -0.2899
Intercept    8.184

Odds Ratios...
      Class
variable   c
=====
hsa_mir_449a    0.368
hsa_mir_34c_5p  0.7483

Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      11           100      %
Incorrectly Classified Instances    0           0      %
Kappa statistic                     1
Mean absolute error                 0.0443
Root mean squared error             0.081
Relative absolute error             8.8586 %
Root relative squared error         15.7468 %
Total Number of Instances          11

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall  F-Measure  ROC Area  Class
      1      0      1      1      1      1      c
      1      0      1      1      1      1      e
Weighted Avg.    1      0      1      1      1      1

=== Confusion Matrix ===
 a b  <-- classified as
 4 0 | a = c
 0 7 | b = e
```

Tabla 7. Resultado del clasificador y predictor "MultiClassClassifier".

Tal y como se muestra en la Tabla 7, de los tres microRNAs seleccionados solo dos (miR-449a y miR-34c-5p) presentan capacidad predictiva y de clasificación para el algoritmo seleccionado. Es más, solo con estos dos microRNAs se alcanzan los valores máximos de predicción y clasificación.

5. DISCUSIÓN

El objetivo del trabajo es obtener, mediante un enfoque de Biología de Sistemas, el perfil de expresión de microRNAs en endometriosis como ejemplo para el aprendizaje y uso de herramientas bioinformáticas. La endometriosis es una enfermedad benigna, crónica, que solo afecta a mujeres y que podría definirse como la presencia de células endometriales fuera de la cavidad endometrial, en localizaciones como el ovario, útero o cavidad peritoneal. Aún no se conoce la causa por la que se produce este desorden; sin embargo, una de las aproximaciones que puede ser útil para encontrar diferencias entre los perfiles moleculares de mujeres que padecen la enfermedad, frente a mujeres sanas, es el análisis de la expresión de los microRNAs. Los microRNAs (miRNAs) son moléculas de ARN de muy pequeño tamaño que impiden la síntesis de proteínas por parte de los ribosomas (Vergoulis et al., 2015).

Diferentes estudios han mostrado que existe un retraso considerable desde que comienzan a aparecer los síntomas de la endometriosis hasta que se produce el diagnóstico (Sociedad Española de Ginecología y Obstetricia, 2014). Hasta la fecha, se han relacionado diferentes biomarcadores con la enfermedad, pero aún no se ha encontrado uno lo suficientemente sensible y específico (Wang et al., 2014). Por lo tanto, consideramos que el uso de la Biología de Sistemas aplicado al análisis de microRNAs puede ser muy útil para avanzar en el estudio de este desorden ginecológico. Con este fin, y gracias a la aplicación de tests estadísticos de tipo no paramétrico, se obtuvieron 11 microRNAs como diferencialmente expresados al comparar muestras de lesiones endometriales con tejido sano. Concretamente, todos los microRNAs aumentaban sus niveles de expresión en las muestras de lesiones endometriales.

Dentro de este grupo de 11 microRNAs destacamos especialmente 3, por ser los que muestran valores de tasas de cambio iguales o superiores a 4, es decir, cuadruplican sus valores de expresión en las muestras de lesiones con respecto a las de tejido sano. Se trata de los microRNAs miR-449a, miR-34c-5p y miR-141-3p. Este conjunto de microRNAs coincide con el detectado por parte de los autores de los microarrays que hemos utilizado en el análisis (Saare et

al., 2014); cabe destacar el hecho de que estos autores validaron experimentalmente la expresión diferencial de estos microRNAs mediante un procedimiento distinto, en este caso a través de amplificación por PCR. Por el contrario, en otro estudio se detectó que dos de estos tres microRNAs (miR-141-3p y miR-34c-5p) disminuyen su expresión en muestras de lesiones endometriales cuando se comparan con tejidos sanos (Ohlsson Teague et al., 2009). Consideramos que estas diferencias pueden deberse a la variabilidad biológica propia y la baja potencia estadística (pocos replicados biológicos) de los estudios realizados. A pesar de ello, consideramos que profundizar en el estudio del miR-449a puede proporcionar información interesante acerca de los mecanismos por los que se produce este desorden ginecológico. Es más, según nuestros resultados, los microRNAs miR-449a y miR-34c-5p dan lugar a los máximos valores de predicción y clasificación, apuntando a ellos como posibles dianas terapéuticas.

Por otro lado, el análisis de enriquecimiento funcional permitió identificar genes relacionados con alteraciones propias de la infertilidad. Concretamente, destacamos el gen GNA13, relacionado con el desarrollo embrionario del útero. Es interesante citar que este mismo gen (GNA13) se ha encontrado en un estudio relacionado con el linfoma de Burkitt, un tumor en el que se produce la alteración de este gen (Greenough & Dave, 2014). La relación entre la implicación de este gen tanto en el tumor como en endometriosis puede deberse a las características especiales del tejido endometrial, dada su capacidad de implantarse tras invadir tejidos adyacentes, a través de la proliferación y diseminación de las células a través de los vasos sanguíneos y linfáticos. Los procesos biológicos que más destacan en relación con los genes implicados en endometriosis son: transcripción; proliferación celular; y desarrollo embrionario. Aún no se conoce la causa de la enfermedad, pero existe un trabajo reciente (Baranov et al., 2015) en el que se han formulado diferentes hipótesis que indican que la causa de dicha enfermedad puede ser de origen mutacional, dando lugar a procesos inflamatorios, angiogénesis, apoptosis y proliferación celular.

6. CONCLUSIONES

Los resultados obtenidos gracias a la aplicación de la Biología de Sistemas nos permiten ampliar los conocimientos acerca de este tema, pues hemos sido capaces de detectar microRNAs que podrían ser utilizados como biomarcadores no invasivos y posibles dianas terapéuticas.

Sugerimos que este estudio podría realizarse con un número mayor de muestras, para de esta forma poder seguir verificando que los resultados son aplicables a gran escala.

Por último, nuestros resultados indican la importancia del estudio molecular de la endometriosis con el fin de detectar biomarcadores y posibles dianas terapéuticas que eviten que se tenga que recurrir a técnicas quirúrgicas invasivas de diagnóstico.

7. BIBLIOGRAFÍA

- Altmäe S, Esteban FJ, Stravreus-Evers A, Simón C, Giudice L, Lessey BA, Horcujadas JA, Macklon NS, D'Hooghe T, Campoy C, Fauser BC, Salamonsen LA, Salumets A, 2014. Guidelines for the design, analysis and interpretation of “omics” data: focus on human endometrium. *Hum Reprod Update*, 20:12-28.
- Baranov VS, Ivaschenko TE, Liehr T, Yarmolinskaya MI, 2015. Systems genetics view of endometriosis: a common complex disorder. *Eur J Obstet Gynecol Reprod Biol*, 185:59–65.
- Brugo-Olmedo S, Chillik C, Kopelman S, 2003. Definición y Causas de la Infertilidad. *Revista Colombiana de Obstetricia y Ginecología*, 54:227–247.
- Esteban FJ, Cano C, De la Haza I, Cano-Ortiz A, Mendizábal NV, Goñi J, Horcujadas JA, 2008. Análisis bioinformático de datos: aplicación en microarrays. *Cmr*, 14:87-96.
- Frank E, Hall MA, Trigg L, Holmes G, Witten IH, 2004. Data mining in bioinformatics using Weka. *Bioinformatics*, 20:2479-2481.
- Gaitán HG, Reveiz L, Farquhar C, Elias VM, 2014. Laparoscopy for the management of acute lower abdominal pain in women of childbearing age. *Cochrane Database Syst Rev*, 5:CD007683.
- Gil Hernández F, 2000. El papel de los biomarcadores en Toxicología Humana. *Revista de Toxicología*, 17:19–26.
- Greenough A, Dave SS, 2014. New clues to the molecular pathogenesis of Burkitt lymphoma revealed through next-generation sequencing. *Curr Opin Hematol*, 4:326–32.
- Huang DW, Sherman BT, Lempicki RA, 2009. Bioinformatics enrichment tools: Paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res*, 37:1–13.
- Laudanski P, Charkiewicz R, Kuzmicki M, Szamatowicz J, Charkiewicz A, Niklinski J, 2013. MicroRNAs expression profiling of eutopic proliferative

- endometrium in women with ovarian endometriosis. *Reprod Biol Endocrinol*, 11:78.
- Makiyan Z, 2017. Endometriosis origin from primordial germ cells. *Organogenesis*, in press.
- Ohlsson Teague EM, Van der Hoek KH, Van der Hoek MB, Perry N, Wagaarachchi P, Robertson SA, Print CG, Hull LM, 2009. MicroRNA-Regulated Pathways Associated with Endometriosis. *Mol Endocrinol*, 23:265–275.
- Saare M, Rekker K, Laisk-Podar T, Sõritsa D, Roost AM, Simm J, Velthut-Meikas A, Samuel K, Metsalu T, Karro H, Sõritsa A, Salumets A, Peters M, 2014. High-throughput sequencing approach uncovers the miRNome of peritoneal endometriotic lesions and adjacent healthy tissues. *PLoS ONE*, 9:e112630.
- Saeed AI, Sharov V, White J, Li J, Liang W, Bhagabati N, Braisted J, Klapa M, Currier T, Thiagarajan M, Sturn A, Snuffin M, Rezantsev A, Popov D, Ryltsov A, Kostukovich E, Borisovsky I, Liu Z, Vinsavich A, Trush V, Quackenbush J, 2006. TM4 Microarray Software Suite. *Methods Enzymol*, 411:134–193.
- Salinas PH, 2006. Análisis De Componentes Principales. *Revista chilena de obstetricia y ginecología*, 71:1–11.
- Siristatidis C, Vogiatzi P, Brachnis N, Liassidou A, Iliodromiti Z, Bettocchi S, Chrelias C, 2015. microRNAs in assisted reproduction and their potential role in IVF failure. *In vivo*, 29:75-169.
- Sociedad Española de Ginecología y Obstetricia, 2014. Endometriosis. *Progresos de Obstetricia y Ginecología*, 57: 436-444.
- Vergoulis, T, Kanellos I, Kostoulas N, Georgakilas G, Sellis T, Hatzigeorgiou A, Dalamagas T, 2015. MirPub: A database for searching microRNA publications. *Bioinformatics*, 31:1502–1504.
- Wang L, Liu HY, Shi HH, Lang JH, Sun W, 2014. Urine peptide patterns for

non-invasive diagnosis of endometriosis: A preliminary prospective study. *Eur J Obstet Gynecol Reprod Biol*, 177:23–28.

Wang X, 2016. Improving microRNA target prediction by modeling with unambiguously identified microRNA-target pairs from CLIP-ligation studies. *Bioinformatics*, 32:1316–1322.

Watson JD, Baker TA, Bell SP, Gann A, Levine M, Losick R, 2014. Molecular biology of the gene, 7th edition. *Boston: Pearson/CSH Press*.

Wong N, Wang X, 2015. miRDB: An online resource for microRNA target prediction and functional annotations. *Nucleic Acids Res*, 43:146–152.