

**ESCUELA TÉCNICA SUPERIOR DE INGENIERÍA  
AGRONÓMICA Y DEL MEDIO NATURAL**



**Grado en Ingeniería Agroalimentaria y del Medio Rural**

**ESTUDIO DEL TEJIDO OVÁRICO PARA MEJORAR LA  
PRODUCCIÓN A TRAVÉS DEL TAMAÑO DE CAMADA  
EN CONEJOS**

**Autora: Marta Serna García**  
**Directora: Dra. María Antonia Santacreu Jerez**  
**Directora: Dra. Rosa María Peiró Barber**  
**Directora científica: Dra. Eva Serna García**

**Diciembre 2016**



### **Estudio del tejido ovárico para mejorar la producción a través del tamaño de camada en conejos.**

El tamaño de camada es uno de los caracteres con mayor importancia económica en la producción del conejo de carne. La respuesta a la selección por tamaño de camada es baja, alrededor de 0,10 gazapos por generación. Una de las alternativas que se ha propuesto para mejorar la respuesta en tamaño de camada es la selección por tasa de ovulación, que es uno de los componentes más importantes del tamaño de camada.

En la UPV se ha llevado a cabo un experimento de selección por tasa de ovulación durante diez generaciones. En la generación base, se recuperaron embriones de las hembras y se vitrificaron. Tras 10 generaciones de selección, estos embriones se desvitrificaron y se transfirieron a hembras receptoras para reconstituir la población de la generación base y disponer de un grupo de hembras control que fueran contemporáneas de las hembras de la generación 10 de selección. La selección por tasa de ovulación ha tenido éxito, la línea seleccionada presenta un aumento de tres óvulos más que la línea control. Sin embargo, no se ha obtenido una respuesta correlacionada en tamaño de camada debido a un aumento en la mortalidad prenatal. Los factores propuestos para explicar el aumento de esta mortalidad en hembras con alta tasa de ovulación son: el aumento del número de óvulos inmaduros, el aumento de la variabilidad en el desarrollo embrionario y el aumento de la competencia entre los embriones o fetos por los recursos que afectan a su desarrollo.

El objetivo de este trabajo es estudiar cuál ha sido el efecto de la selección sobre la expresión génica diferencial en el tejido ovárico mediante el uso de microarrays. Los resultados de este trabajo permitirán buscar alternativas de mejora más eficaz del tamaño de camada que reporten mayores beneficios económicos a las explotaciones cunícolas.

**Palabras clave:** producción, explotaciones cunícolas, ovario, tamaño de camada, tasa de ovulación, microarrays.

### **Estudi del teixit ovàric per a millorar la producció per mitjà de la grandària de ventrada en conills.**

La grandària de ventrada és un dels caràcters amb major importància econòmica en la producció del conill de carn. La resposta a la selecció per grandària de ventrada és baixa, al voltant de 0,10 catxaps per generació. Una de les alternatives que s'ha proposat per a millorar la resposta en grandària de ventrada és la selecció per taxa d'ovulació, que és un dels components més importants de la grandària de ventrada.

En la UPV s'ha dut a terme un experiment de selecció per taxa d'ovulació durant deu generacions. En la generació base, es van recuperar embrions de les femelles i es van vitrificar. Després de 10 generacions de selecció, aquests embrions es van desvitrificar i es van transferir femelles receptores per a reconstituir la població de la generació base i disposar d'un grup de femelles control que foren contemporànies de les femelles de la generació 10 de selecció. La selecció per taxa d'ovulació ha tingut èxit, la línia seleccionada presenta un augment de tres òvuls més que la línia control. No obstant això, no s'ha obtingut una resposta correlacionada en grandària de ventrada degut a un augment en la mortalitat prenatal. Els factors proposats per a explicar l'augment d'aquesta mortalitat en femelles amb alta taxa d'ovulació són: l'augment del nombre d'òvuls immadurs, l'augment de la variabilitat en el desenvolupament embrionari i l'augment de la competència entre els embrions o fetus pels recursos que afecten el seu desenvolupament.

L'objectiu d'aquest treball és estudiar quin ha sigut l'efecte de la selecció sobre l'expressió gènica diferencial en el teixit ovàric per mitjà de l'ús de microarrays. Els resultats d'aquest treball permetran trobar alternatives de millora més eficaç de la grandària de ventrada que reporten majors beneficis econòmics a les explotacions cunícoles.

**Paraules clau:** producció, explotacions cunícoles, ovari, grandària de ventrada, taxa d'ovulació, microarrays.

### **Study of ovarian tissue to improve production through litter size in rabbits.**

Litter size is one of the most economically important traits in rabbit meat production. Selection response on litter size is low, about 0.10 kits per generation. Selection for ovulation rates has been proposed to improve the response in litter size. Ovulation rate is one of the most important components of litter size.

In UPV farms, one ovulation rate selection experiment has been carried out for ten generations. In the base generation, embryos were recovered from females and vitrified. After 10 generations of selection, these embryos were devitrified and transferred to recipient females to reconstitute a group of control females. Control females were contemporaneous with selected females from generation 10. Selection on ovulation rate has been successfully, selected line showed an increase of three ovules more than control line. However, a correlated response in litter size has not been obtained because of an increase in prenatal mortality. The factors proposed to explain the increase of this mortality in females with high ovulation rate are: the increase in the number of immature eggs, the increase in the variability in the embryonic development and the increase of the competition between the embryos or fetuses by the resources which affect their development.

The aim is to study the effect of selection on differential gene expression in ovarian tissue through the use of microarrays. Results of this work will allow searching for alternatives of more efficient improvement of the litter size that report greater economic benefits to the farms.

**Key words:** Production, rabbit farms, ovary, litter size, ovulation rate, microarrays.

# Índice general

---

<b>I. Introducción</b> .....	<b>1</b>
1. Características reproductivas de la coneja.....	1
2. La mejora del tamaño de camada .....	1
3. Selección por tasa de ovulación.....	2
3.1. Maduración del ovocito .....	2
3.2. Variabilidad del desarrollo embrionario.....	3
3.3. Competencia entre espacio y nutrientes .....	3
4. El transcriptoma y su análisis .....	4
5. Biochips de expresión .....	4
5.1. Tipos de biochips de expresión .....	5
5.2. Microarrays de expresión Affymetrix <i>GeneChip</i> <sup>®</sup> .....	7
5.2.1. Estructura de los microarrays <i>GeneChip</i> <sup>®</sup> .....	7
5.2.2. Fabricación de los microarrays <i>GeneChip</i> <sup>®</sup> .....	9
5.2.3. Preparación de la muestra y cuantificación del nivel de expresión en los microarrays <i>GeneChip</i> <sup>®</sup> .....	10
5.2.4. Microarray <i>GeneChip</i> <sup>®</sup> <i>Rabbit Gene 1.0 ST Array</i> .....	11
<b>II. Objetivos</b> .....	<b>15</b>
<b>III. Material y métodos</b> .....	<b>17</b>
1. Recogida del material biológico .....	17
2. Extracción del RNA total .....	17
3. Control de calidad, pureza y cuantificación del RNA total obtenido.....	18
3.1. Electroforesis capilar .....	18
3.2. Espectrofotometría .....	20
4. Protocolo <i>GeneChip</i> <sup>®</sup> <i>Rabbit Gene 1.0 ST</i> .....	20
5. Preprocesado de los datos de intensidad brutos .....	23
6. Control de calidad de los biochips.....	24
7. Análisis transcripcional.....	27
<b>IV. Resultados</b> .....	<b>29</b>
1. Control de calidad de las muestras .....	29
2. Control de pureza y cuantificación de RNA total .....	31
3. Control de calidad de los microarrays .....	32
4. Análisis transcripcional.....	39
4.1. Análisis de los componentes principales .....	39
4.2. Análisis diferencial entre los 2 grupos estudiados .....	40
<b>V. Conclusiones</b> .....	<b>47</b>
<b>VI. Referencias</b> .....	<b>49</b>

# Índice de figuras

---

<b>Figura 1:</b> Dogma Central de la Biología Molecular modificado de la propuesta inicial de Crick (1970). .....	4
<b>Figura 2:</b> Herramientas tecnológicas que permiten analizar la expresión génica. ....	4
<b>Figura 3:</b> Biochips de cDNA de dos colores. ....	6
<b>Figura 4:</b> Biochips de Affymetrix de un color. ....	7
<b>Figura 5:</b> Biochip <i>GeneChip</i> <sup>®</sup> de Affymetrix. ....	7
<b>Figura 6:</b> Complementariedad entre cadenas de nucleótidos .....	8
<b>Figura 7:</b> Estructura de las sondas de DNA sintetizadas en la superficie de los chips de vidrio. ....	8
<b>Figure 8:</b> Superficie de un biochip junto con la imagen de un biochip hibridado y ejemplo de una cuadrícula hibridada correspondiente a un oligonucleótido .....	9
<b>Figure 9:</b> Proceso de fabricación de los microarrays de expresión Affymetrix mediante el proceso de fotolitografía. ....	10
<b>Figure 10:</b> Esquemas de un protocolo típico de un biochip de expresión de affymetrix. ....	10
<b>Figure 11:</b> Proceso de hibridación entre las sondas del chip y los ácidos nucleicos de la muestra procesada .....	11
<b>Figure 12:</b> Distribución de las sondas en los transcritos que interrogan los biochips Affymetrix. ....	12
<b>Figura 13:</b> Esquema de trabajo para el análisis de expresión de los datos obtenidos a partir del escaneado de los biochips <i>GeneChip</i> <sup>®</sup> . ....	13
<b>Figure 14:</b> Esquema representativo de las tres fases obtenidas al adicionar el cloroformo a las muestras homogeneizadas. ....	18
<b>Figura 15:</b> Electroferogramas ejemplo de muestras de RNA con diferente integridad... ..	19
<b>Figura 16:</b> Gráfica de la absorbancia en nm de los ácidos nucleicos y de las proteína. ....	20
<b>Figura 17:</b> Esquema del protocolo <i>GeneChip</i> <sup>®</sup> <i>Rabbit Gene 1.0 ST</i> .....	22
<b>Figure 18:</b> Imagen digital obtenida del biochip (.DAT) tras el escaneado .....	24
<b>Figura 19:</b> Ejemplo de la acción del oligo B2 iluminando las esquinas del array .....	25
<b>Figura 20:</b> Ejemplo de la acción del oligo B2 mostrando el nombre del array. ....	25
<b>Figura 21:</b> Ejemplo de electroferograma modelo en el cual se observa el patrón de picos que se debe obtener en una muestra no degradada .....	29
<b>Figura 22:</b> Ejemplo de electroferograma degradado donde el 28S no se visualiza.: .....	29
<b>Figura 23:</b> Ejemplo de dos electroferogramas de las muestras realizadas con el Bioanalizador de Agilent .....	30
<b>Figura 24:</b> Esquina superior derecha de la imagen obtenida tras el escaneo y lectura de microarray .....	32
<b>Figura 25:</b> Parte de la imagen obtenida tras el escaneo y lectura de microarray .....	32

<b>Figura 26:</b> Desviación media absoluta de los residuos (“mad residual mean”) y media absoluta de la expresión relativa en logaritmo en base 2 (“RLE mean”) para las categorías “all probeset” y “positive control”, y área bajo la curva relativa (AUC) entre los controles positivos y negativos (“pos vs neg auc”).	33
<b>Figura 27:</b> Diagrama de cajas de los valores de expresión logarítmica relativa.	34
<b>Figura 28:</b> Diagrama de cajas de los valores de expresión logarítmicas antes de la normalización.	35
<b>Figura 29:</b> Diagrama de cajas de las señales de expresión logarítmicas normalizadas	36
<b>Figura 30:</b> Media de los valores de intensidad de los sets de sondas pertenecientes a la categoría de controles de marcaje obtenidos en cada chip.	37
<b>Figura 31:</b> Media de los valores de intensidad de los sets de sondas pertenecientes a la categoría de controles hibridación obtenidos en cada chip	38
<b>Figura 32:</b> Análisis de componentes principales de las seis muestras analizadas.	39
<b>Figura 33:</b> Planos tridimensionales del análisis de componentes principales de las seis muestras analizadas.	40
<b>Figura 34:</b> <i>Heatmap</i> de los 1.357 genes diferencialmente expresados entre las tres hembras de la población control y las tres hembras de la población seleccionada	41
<b>Figura 35:</b> Representación de 19 genes de los 1.357 genes diferencialmente expresados entre la población seleccionada y control estudiados que se relacionan con muerte celular ( <i>cell death</i> en inglés)	44

# Índice de tablas

---

<b>Tabla 1:</b> Productos requeridos para la realización del protocolo <i>GeneChip® Human Gene 1.0 ST</i> .....	21
<b>Tabla 2:</b> Esquema de trabajo para la realización del protocolo <i>GeneChip® Human Gene 1.0 ST</i> .....	21
<b>Tabla 3:</b> Valores RIN obtenidos para las muestras analizadas. ....	30
<b>Tabla 4:</b> Valores de pureza (Ratio A260/A280) y concentración para cada una de las muestras .....	31
<b>Tabla 5:</b> Genes cuya expresión entre las hembras de la población seleccionada y control presentan un p-valor <0,05 y <i>fold change</i>   1,5   .....	42-43

# I. Introducción

---

## 1. Características reproductivas de la coneja

La coneja es un mamífero plútopo, de ciclo reproductivo corto. Alcanza la pubertad entre las 18 y 20 semanas de vida. Posee una alta productividad media, en torno a los 8 nacidos vivos (Baselga y Blasco., 1989), dependiendo de la línea y la raza. No presenta ciclo estral definido, pertenece a un grupo de mamíferos en los que la ovulación se produce como consecuencia del coito.

Su aparato reproductor está constituido por dos ovarios, dos oviductos y un útero bicorne. Los dos cuernos uterinos, completamente separados entre sí, desembocan de forma independiente en la vagina.

La gestación tiene una duración de aproximadamente 31 días. En explotaciones con un sistema reproductivo semi-intensivo, las conejas serán montadas entre el noveno y el undécimo día post-parto, dando lugar a un solape parcial de gestación y lactación. El destete de los gazapos se realiza a los 28-35 días. Si la ovulación no es seguida por una gestación, puede generarse un estado de pseudogestación con una duración aproximada de 16-17 días.

## 2. La mejora del tamaño de camada

El tamaño de camada es uno de los caracteres con mayor importancia económica en la producción del conejo de carne (Cartuche et al., 2014). Es por ello que los programas de mejora genética en conejo disponen de líneas maternas en las que el objetivo de selección es el tamaño de camada (Baselga y Blasco, 1989). El criterio más utilizado para la selección del tamaño de camada es el número de gazapos en el momento del destete. De esta forma, se valora el número de animales nacidos vivos y la capacidad de la hembra de sacar adelante la camada hasta el destete.

El tamaño de camada es un carácter con una baja heredabilidad (alrededor de 0.10), de expresión tardía y que sólo se expresa en las hembras (Blasco et al., 1993; Laborda et al. 2011). En los experimentos de selección directa por tamaño de camada llevados a cabo en distintas líneas de conejo se han obtenido respuestas bajas, alrededor de 0,08 gazapos/generación (Blasco, 1996; García y Baselga, 2002). Así, la selección directa por tamaño de camada ha tenido menor éxito de lo esperado y la respuesta tan sólo se observa a largo plazo. Por este motivo, se han propuesto diferentes alternativas para mejorar de forma más eficaz el tamaño de camada.

Una de las alternativas propuestas para mejorar el tamaño de camada ha sido la selección por los principales componentes del tamaño de camada que son: la tasa de ovulación y la supervivencia prenatal. Se ha observado que, al seleccionar por tamaño de camada, la respuesta se produce bien por un aumento en la tasa de ovulación (García et al., 2000) o bien por un aumento de la supervivencia prenatal (García et al., 2000). Así, la selección indirecta se puede realizar mediante la selección por cualquiera de estos dos caracteres o por un índice que combine ambos.

En el apartado siguiente se va a hablar del único experimento de selección por tasa de ovulación realizado en conejo. Los animales de este trabajo proceden de este experimento.

### 3. Selección por tasa de ovulación

La selección por tasa de ovulación se ha propuesto como una vía de mejorar con más éxito el tamaño de camada. La tasa de ovulación presenta una heredabilidad superior a la del tamaño de camada. En conejo, las estimas de heredabilidad de la tasa de ovulación rondan el valor de 0.20 (Blasco et al., 1993; Laborda, 2011). Otro parámetro importante para predecir el éxito de la mejora del tamaño de camada por selección de la tasa de ovulación es la correlación genética entre ambos caracteres. El valor estimado de esta correlación es aproximadamente 0.35 aunque los errores de estimación son elevados (Blasco et al., 1993; Laborda, 2011). Para disminuir los errores de estimación de este parámetro son necesarios un elevado número de datos. La estima de la tasa de ovulación "in vivo" se suele hacer por laparoscopia, el tiempo empleado y el coste de la técnica es un problema para disponer de una gran base de datos.

En conejo se ha llevado a cabo un único experimento de selección por tasa de ovulación (Laborda et al. 2011, 2012a y 2012b). La selección se llevó a cabo durante diez generaciones. En la generación base, se vitrificaron embriones que se mantuvieron en nitrógeno líquido hasta que se desvitrificaron y transfirieron para reconstituir una población control coetáneo con hembras seleccionadas de la generación diez. El criterio de selección fue la tasa de ovulación estimada por laparoscopia como el número de cuerpos lúteos presentes en los ovarios a los 12 días de la segunda gestación. Además, al mismo tiempo se contaron el número de fetos presentes en ambos cuernos uterinos lo que permitió estimar la respuesta correlacionada de la supervivencia embrionaria y fetal. La mayoría de los autores están de acuerdo en considerar que en la supervivencia prenatal se pueden distinguir dos periodos, el periodo comprendido entre la ovulación y la implantación (supervivencia embrionaria) y el periodo comprendido desde la implantación, sobre el día siete de gestación, hasta el nacimiento (supervivencia fetal)

La selección por tasa de ovulación incrementó la tasa de ovulación en tres óvulos sin embargo no se tuvo éxito en mejorar el tamaño de camada. Después de diez generaciones, el tamaño de camada de las hembras de la línea seleccionada era similar al de hembras de la población control. Este fracaso en aumentar el tamaño de camada a pesar del aumento de la tasa de ovulación se debió a una disminución de la supervivencia prenatal (casi un 1% por generación).

El aumento de la mortalidad prenatal se produjo tanto en el periodo embrionario como en el periodo fetal, aunque la mortalidad fetal explicaba una parte importante de la mortalidad prenatal observada (Laborda et al 2011, 2012a y 2012b).

Los factores que han sido propuestos por diferentes autores para explicar el aumento de la mortalidad prenatal en hembras con alta tasa de ovulación son:

- La maduración del ovocito.
- La variabilidad en el desarrollo embrionario.
- La competencia por el espacio y los nutrientes que afectan a su desarrollo (Santacreu, 2006 en conejos; Geisert y Schmitt, 2002 en cerdos).

#### 3.1 Maduración del ovocito

Se ha observado un aumento del número de ovocitos inmaduros en hembras con alta tasa de ovulación tanto en experimentos de superovulación (Krisner, 2004; Koenig et al., 1986) como en un experimento de selección para aumentar la tasa de ovulación en cerdos

(Koenig et al., 1986). Los ovocitos inmaduros son ovocitos con menor capacidad para ser fecundados y si se fecundan dan lugar a embriones de mala calidad que pueden morir antes o después de la implantación lo que podría explicar la mortalidad observada en ambos periodos en el experimento de selección por tasa de ovulación en conejos.

### **3.2 Variabilidad en el desarrollo embrionario**

Las hembras con elevada tasa de ovulación podrían aumentar la variabilidad en el desarrollo embrionario debido a un aumento del tiempo en el proceso de ovulación. Los primeros ovocitos ovulados serían los primeros en ser fecundados y en enviar señales al útero. Se ha observado que el útero está sincronizado con los embriones más avanzados y que estos tienen mayor probabilidad de sobrevivir (Torres et al., 1984 en conejos; Pope, 1988 y Xie et al., 1990 en cerdos; Wilmut et al., 1986 y Al-Shorepy et al., 1992 en ratones).

La variabilidad en el desarrollo embrionario puede causar asincronía entre el estado de desarrollo del embrión y las secreciones uterinas, así algunos embriones no presentarán el desarrollo adecuado respecto al ambiente uterino donde se desarrollan (Wilmut et al., 1986; Torres et al., 1987; Bazer et al., 1990). Se ha sugerido que los embriones menos desarrollados pueden llegar a implantar y que es en el periodo post-implantación cuando tienen mayor probabilidad de morir (Wintenberger-Torres et al., 1974 en conejos; Pope et al., 1988; Wilde et al., 1988 cerdos).

### **3.3 Competencia por el espacio y los nutrientes**

La principal causa de mortalidad después de la implantación es la competencia entre los fetos por la disponibilidad del espacio y los nutrientes (Adams, 1962; Geisert y Schmitt, 2002). Este factor es especialmente importante en hembras con una alta tasa de ovulación, como por ejemplo las hembras de la línea seleccionada por tasa de ovulación, que presentan el útero atestado de embriones y fetos.

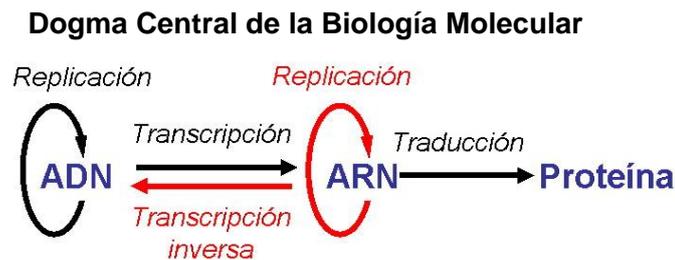
En conejos, se ha observado que una reducida vascularización por feto podría ser un factor importante para la supervivencia del feto. El suministro de sangre es especialmente importante en conejos debido a su tipo de placenta hemocorial, placenta de tipo invasiva donde el tejido fetal está directamente en contacto con la circulación sanguínea materna (Leiser and Kaufmann, 1994).

Además, se sabe que el pico de mayor mortalidad fetal se presenta entre los 7 y los 17 días de la gestación. Un momento clave en la supervivencia del feto es cuando la placenta corioalantoidea ha terminado su desarrollo y la nutrición del feto comienza a ser controlada por la misma, alrededor de día 12 de gestación.

En conclusión, el tamaño de camada es uno de los caracteres económicamente más importantes en la producción del conejo de carne. Se están buscando métodos de selección alternativos para incrementar de forma más eficiente el tamaño de camada. Una de las alternativas propuestas es la selección por tasa de ovulación, que es uno de los componentes del tamaño de camada. La selección para mejorar la tasa de ovulación ha sido efectiva pero no se ha mejorado el tamaño de camada; la selección ha incrementado la tasa de ovulación, pero también la mortalidad embrionaria y principalmente la mortalidad fetal. El estudio de la expresión génica del tejido ovárico de hembras de la línea seleccionada y control puede ayudar a comprender las posibles causas del incremento de la mortalidad embrionaria y fetal en la línea seleccionada durante diez generaciones para aumentar la tasa de ovulación.

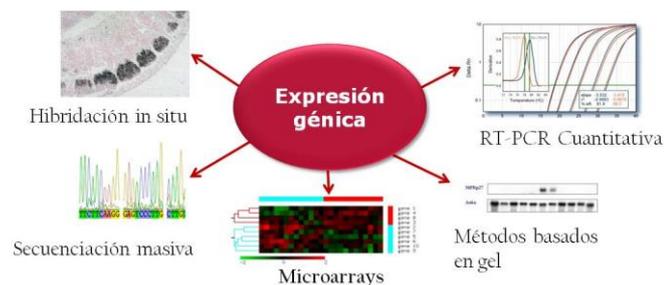
## 4. El transcriptoma y su análisis

El RNA mensajero (mRNA) es el ácido ribonucleico que contiene la información genética procedente del ácido desoxiribonucleico (DNA). Para la formación de una proteína, los genes son expresados por medio de la transcripción de forma que transfieren su información a este tipo de moléculas. Así, todo el mRNA transcrito bajo unas circunstancias específicas forma el transcriptoma. Estas moléculas serán más tarde traducidas a proteínas, estableciendo así el *Dogma Central de la Biología Molecular* (Crick, 1970) (Figura 1).



**Figura 1:** Dogma Central de la Biología Molecular. Modificado de la propuesta inicial de Crick (1970). (<http://pendientedemigracion.ucm.es/info/genetica/grupod/Transcripcion.htm>)

Actualmente distinguen dos eras; la era pre-genómica en la que se estudia individualmente el gen, mediante las técnicas de hibridación *in situ*, *Northern blotting* (métodos basados en gel) o RT-PCR (reacción en cadena de la polimerasa en tiempo real, acrónimo del inglés real *time polymerase chain reaction*). La segunda era, la llamada post-genómica, permite estudiar de manera simultánea todo el genoma mediante, por ejemplo, el uso de herramientas como microarrays y secuenciación masiva (Figura 2). Estas tecnologías nos permiten estudiar perfiles genéticos que van desde 100 a miles de genes o incluso el genoma total, es decir, análisis globales de expresión en un solo ensayo para obtener información de todos los genes del organismo en unas condiciones específicas.



**Figura 2.** Herramientas tecnológicas que permiten analizar la expresión génica.

## 5. Biochips de expresión

El desarrollo y financiación en los últimos años de un gran número de proyectos de secuenciación masiva, de entre los cuales cabe destacar el Proyecto Genoma Humano (Venter, et al., 2001), ha permitido el desarrollo de nuevas técnicas tanto en el campo de la biología como en el campo de la bioinformática. Los avances logrados en estas áreas

han permitido el desarrollo de una nueva mentalidad en la que existe una visión global de los procesos biológicos, dando lugar a las denominadas ciencias “ómicas”.

Dentro de estas ciencias “ómicas” se pueden incluir la genómica, la transcriptómica, la proteómica y la metabolómica, entre otras.

A finales de los años 80 cuatro científicos (Stephen Fodor, Michael Pirrung, Leighton Read y Lubert Stryer) desarrollaron uno de los mayores logros en biología molecular experimental: una revolucionaria tecnología para la determinación y cuantificación de DNA en una muestra.

Stephen Fodor y sus colaboradores describieron, por primera vez, el término *microarray* y publicaron su invención en la revista Science en 1991 (Fodor, et al., 1991). A este término de *microarray* (el de mayor aceptación a nivel internacional) también se le conoce como biochip, chip de DNA, o *Genechip*. Además, Stephen Fodor fundó la empresa Affymetrix la cual es pionera en comercializar biochips, ya sean de expresión génica como de caracterización del DNA.

El concepto básico de un biochip es sencillo: Consiste en un gran número de moléculas de DNA fijadas y ordenadas sobre un soporte sólido de manera que forman una matriz de secuencias en dos dimensiones. A estos fragmentos de DNA de una sola hebra inmovilizados en el soporte, se les denomina sondas (*probes* en inglés). Por otro lado, el RNA total problema es extraído de un tipo celular o tejido de interés. Estos ácidos nucleídos extraídos de las muestras a analizar (*target sequences* en inglés) se marcan por diversos métodos y se incuban dentro del biochip donde están todas las sondas, permitiendo la hibridación (reconocimiento y unión entre moléculas complementarias). Dicha hibridación entre la sonda y la muestra, basada en el modelo de pares de bases de Watson and Crick, proporciona una medida cuantitativa de la abundancia de cada una de las secuencias. Esta información es capturada digitalmente y analizada para obtener información biológica. La comparación de los patrones de hibridación mediante herramientas informáticas permite la identificación de los genes diferenciales entre muestras.

Por tanto, los biochips proporcionan una poderosa herramienta en la biología molecular, cuyas aplicaciones incluyen diversas áreas como las ciencias agronómicas, las ciencias básicas, las ciencias biomédicas, etc.

## 5.1 Tipos de biochips de expresión

Existen principalmente dos tipos de biochips de uso común: los biochips de DNA complementario (cDNA) y los biochips de oligonucleótidos.

Los biochips de cDNA están formados por una colección de secuencias de genes, normalmente productos de PCR (100-2.000 pares de bases (pb), procedentes de secuencias de cDNA y de ESTs (marcador de secuencia expresada, acrónimo del inglés *Expressed Sequence Tags*), y colocados individualmente sobre una matriz sólida. Con este tipo de biochips, dos tipos de muestra pueden ser comparadas en un mismo ensayo. Para ello, tras la preparación de ambas muestras, se lleva a cabo el marcaje con diferentes fluoróforos y la hibridación en un único array (Figura 3).

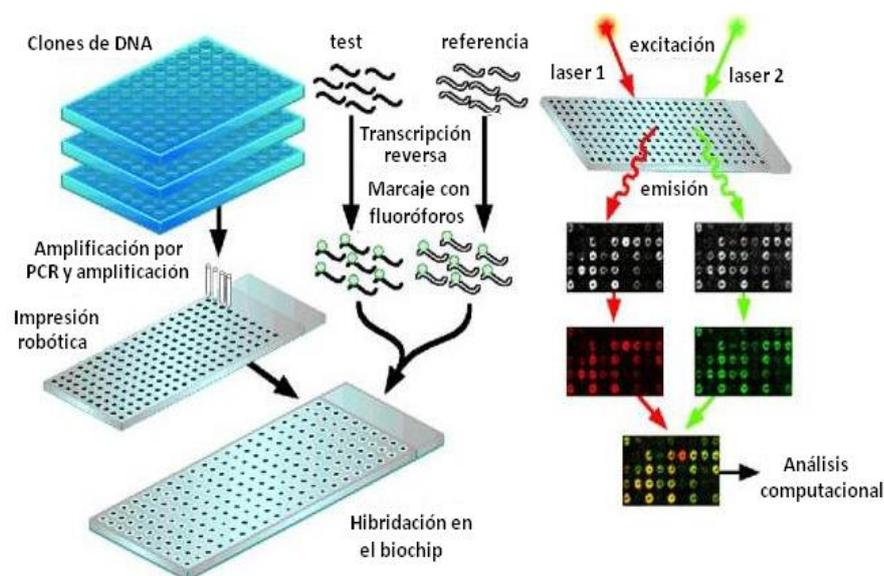
Con respecto a las ventajas de los biochips de cDNA, se trata de una tecnología relativamente accesible y rentable. Tanto la hibridación como la captura de datos no necesita equipamiento especializado, y los arrays prefabricados son relativamente baratos. Por último, la gran longitud de las secuencias de la muestra a analizar (2 kbp) incrementa la sensibilidad de detección (elevada intensidad de señal).

En cuanto a los inconvenientes, pueden existir problemas de competitividad ya que se hibridan dos muestras dentro de un mismo microarray. Además, la longitud de las sondas hace que puede dar lugar a fallos de especificidad a la hora de detectar genes individuales y el estado del DNA de doble cadena en el microarray está mal definido, de forma que podría haber restricción de contactos con la matriz y podrían producirse enlaces cruzados que afectarían a la hibridación. Por último, la fabricación del microarray depende de los sets de clones, con lo cual, incluso utilizando sets muy bien mantenidos, estos son propensos a la contaminación, condicionando los resultados del análisis.

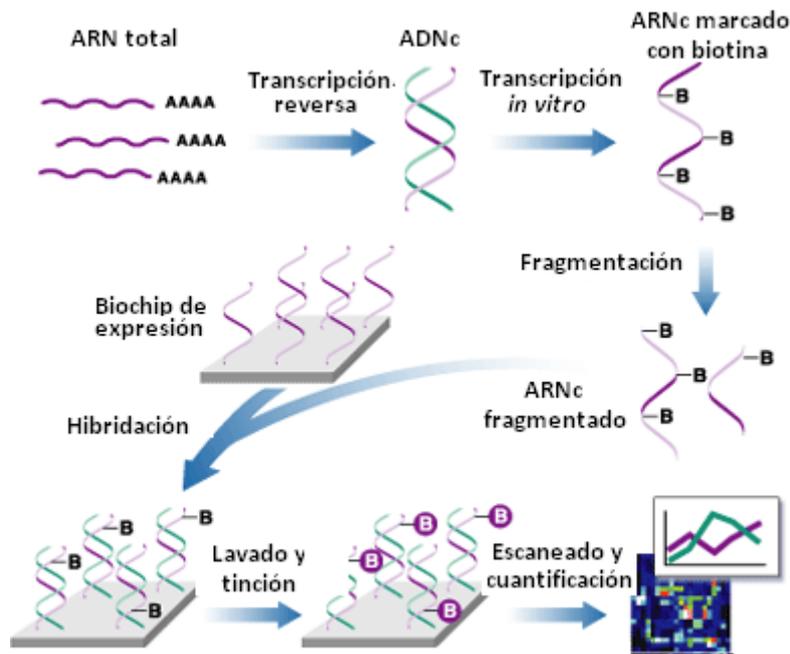
Por otra parte, los biochips de oligonucleótidos se fabrican mediante la síntesis *in situ* de sondas, por ejemplo, fotolitografía, de cadena única sobre la matriz, en base a la información de secuencias anotadas en bases de datos. La empresa líder en este campo es Affymetrix, por lo que la descripción de este tipo de array se realiza en base a sus biochips denominados Affymetrix *GeneChip*<sup>®</sup>. A diferencia de los microarrays de cDNA, solo una muestra puede ser hibridada en un array, por lo que las muestras se analizan en ensayos individuales (Figura 4).

Con respecto a las ventajas del uso de los microarrays Affymetrix, dado que las sondas están basadas enteramente en información de secuencia de bases de datos, no conlleva problemas en la preparación de set de clones, y por tanto tampoco problemas de contaminación. Además, el uso de secuencias cortas (25 pares de bases) minimiza la hibridación cruzada, aumentando así la especificidad. Por otro lado, para cada transcrito se usan múltiples secuencias cortas que representan partes significativamente únicas dentro del genoma, lo que permite distinguir variantes de *splicing* y genes estrechamente relacionados, así como aumentar la sensibilidad. Por tanto, aunque las secuencias de corta longitud podrían producir una disminución de la sensibilidad en comparación con los microarrays de cDNA, el uso de múltiples sondas para cada transcrito (*probeset en inglés*) permite conseguir finalmente una alta sensibilidad.

En cuanto a las desventajas, la fabricación de los microarrays de Affymetrix necesita el uso de un equipamiento especializado, no siendo posible la fabricación por parte del investigador, reduciendo por tanto la flexibilidad en el diseño del microarrays.



**Figura 3:** Biochips de cDNA de dos colores. Imagen modificada de Sánchez & Ruíz de Villa, A Tutorial Review of Microarray Data Analysis, 2008.



**Figura 4:** Microarray de expresión de Affymetrix. Imagen modificada de Sánchez & Ruíz de Villa, A Tutorial Review of Microarray Data Analysis, 2008.

## 5.2 Microarrays de expresión Affymetrix *GeneChip*<sup>®</sup>

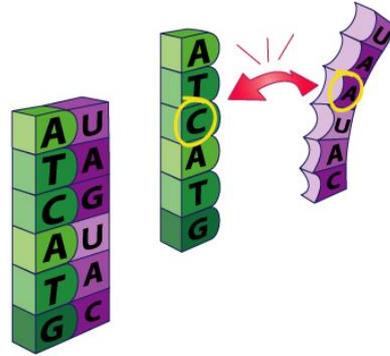
### 5.2.1. Estructura de los microarrays *GeneChip*<sup>®</sup>

Los microarrays de expresión de Affymetrix son denominados *GeneChip*<sup>®</sup> (Figura 5) y se caracterizan por tratarse de arrays de oligonucleótidos de DNA de alta densidad.



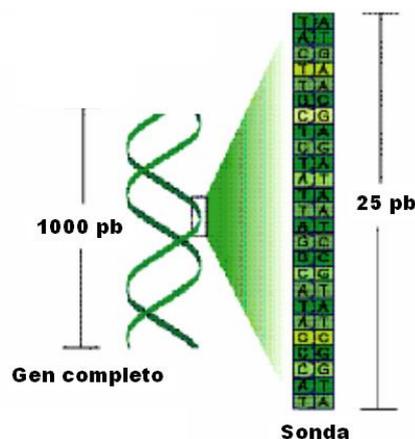
**Figura 5:** Biochip *GeneChip*<sup>®</sup> de Affymetrix.

Estos microarrays usan la atracción química natural entre el DNA (en el array) y las moléculas de DNA de cadena única (sDNA) o RNA objetivo (procedentes de la muestra problema a analizar) para determinar el nivel de expresión de un determinado gen. Así, cuando existe una perfecta complementariedad entre dos cadenas de nucleótidos, ambas se mantendrán unidas. Sin embargo, si las bases no son complementarias, ambas cadenas no se unirán. Incluso si una única base de las cadenas no empareja adecuadamente la hibridación no se producirá (Figura 6).



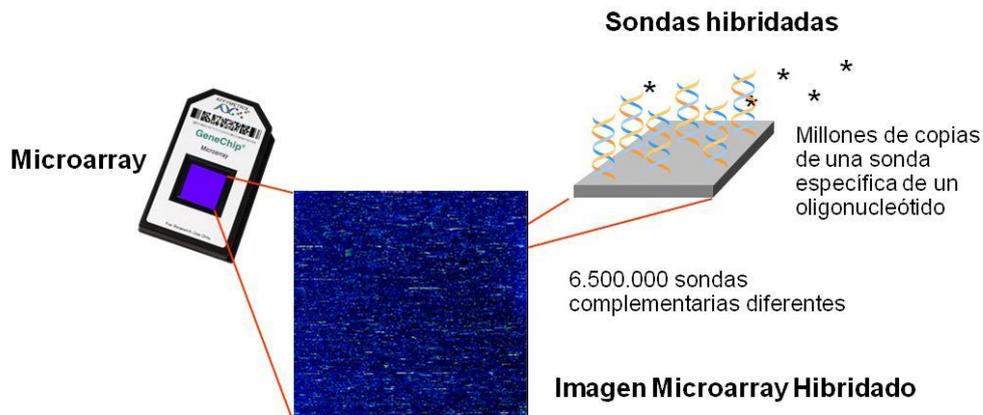
**Figura 6:** Complementariedad entre cadenas de nucleótidos (Affymetrix, 2005).

Para que los microarrays cumplan su función de medir la expresión de cada gen, el primer paso consiste en sintetizar la cadena de DNA denominada sonda en la superficie de un chip de cuarzo. Estas cadenas, de 25 pb representan una pequeña sección característica y específica de un gen (Figura 7), de modo que no podrán emparejarse con ninguna otra secuencia de otro gen a lo largo del genoma. Así, cuando una molécula se une a dicha sonda se conoce que gen está expresado en la muestra, ya que la única hibridación posible es con ese gen.



**Figura 7:** Estructura de las sondas de DNA sintetizadas sobre la superficie de cuarzo de los chips. Imagen modificada de Affymetrix, 2005.

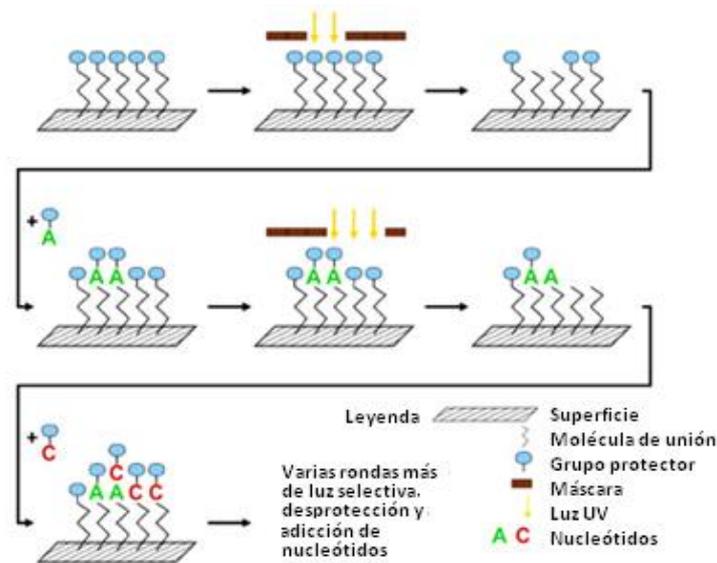
La superficie de un microarrays de Affymetrix es de aproximadamente 1 cm por 1 cm. Dicha superficie está dividida formando una cuadrícula donde cada partición tiene unas dimensiones de 5  $\mu\text{m}$  por 5  $\mu\text{m}$  y contiene millones de réplicas de un solo tipo de sonda. Por otra parte, pueden llegar a contener hasta aproximadamente 6,5 millones de sondas diferentes representando todo el genoma (Figura 8).



**Figura 8:** Superficie de un biochip junto con la imagen de un biochip hibridado y ejemplo de una cuadrícula hibridada correspondiente a un oligonucleótido. Imagen modificada de Affymetrix, 2005.

### 5.2.2. Fabricación de los microarrays GeneChip®

El proceso de fabricación de estos microarrays es mediante la técnica de fotolitografía. Ésta es una técnica aplicada ampliamente en la industria de los semiconductores que combina la síntesis química en fase sólida con técnicas de fabricación fotolitográfica. Se utiliza una máscara para exponer selectivamente la luz sobre una oblea de silicio. La luz ultravioleta, por tanto, elimina grupos protectores fotolábiles en posiciones definidas. Las zonas desprotegidas es donde se van a ir añadiendo las bases de interés. Así, aquellas cadenas no “desprotegidas” no serán capaces de recibir un nuevo nucleótido. Cualquier secuencia puede ser construida en cualquier posición sobre la oblea de silicio. Este proceso debe ser muy preciso ya que es importante que los nucleótidos sean añadidos en el orden correcto para formar la sonda correcta. El número de pasos se corresponde con la longitud del oligonucleótido (y no el número total de oligonucleótidos). Por ello se puede aumentar el número de genes sin aumentar el número de pasos, y es así, cuando se llega finalmente, a la secuencia de longitud total de 25 pares de bases, que es la longitud final de cada sonda (Figura 9). En resumen, esta técnica nos permite la síntesis de miles de millones de oligonucleotidos de cDNA, en lugares definidos, sobre la superficie de cuarzo usando la luz ultravioleta. El proceso conlleva una alta precisión debido a la complejidad, multitud y variedad de sondas que han de ser sintetizadas, por lo que está altamente automatizado con la ayuda de computadores y robots.



**Figura 9:** Proceso de fabricación de los biochips de expresión Affymetrix mediante el proceso de fotolitografía. Imagen modificada de Affymetrix, 2005.

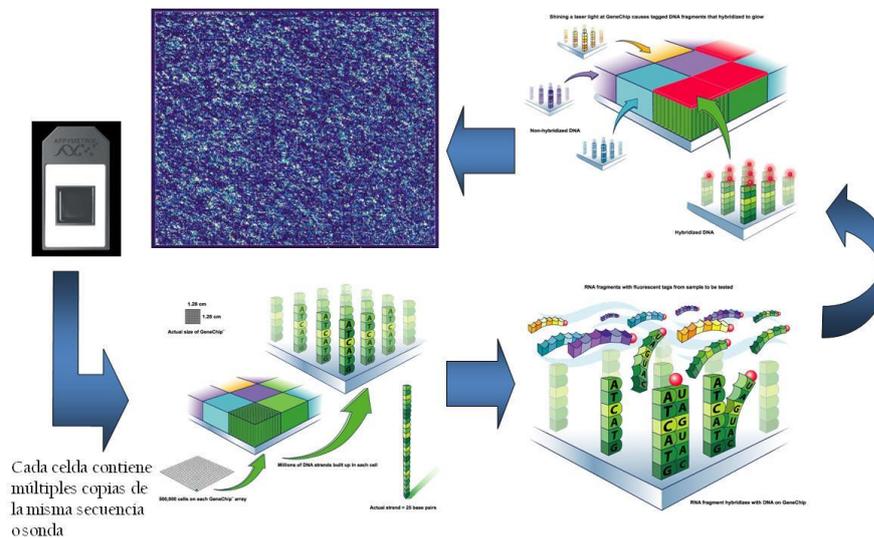
### 5.2.3. Preparación de la muestra y cuantificación del nivel de expresión en los microarrays *GeneChip*<sup>®</sup>

Una vez diseñado el chip, se lleva a cabo la extracción del RNA total de la muestra biológica y que depende del protocolo que especifique el biochip *GeneChip*<sup>®</sup> utilizado. Normalmente es un protocolo con múltiples pasos (amplificación, fragmentación, marcaje, etc.) (Figura 10). Durante el paso de marcaje, las moléculas añadidas actuarán como soporte de moléculas fluorescentes que serán introducidas en el microarray posteriormente.



**Figura 10:** Esquema de un protocolo típico de un biochip de expresión de Affymetrix.

Tras llevar a cabo el protocolo, la muestra preparada se introduce en el microarray para permitir la hibridación. Durante la hibridación, las cadenas de RNA o sDNA de la muestra intentan buscar sondas complementarias a lo largo del microarray (Figura 11).



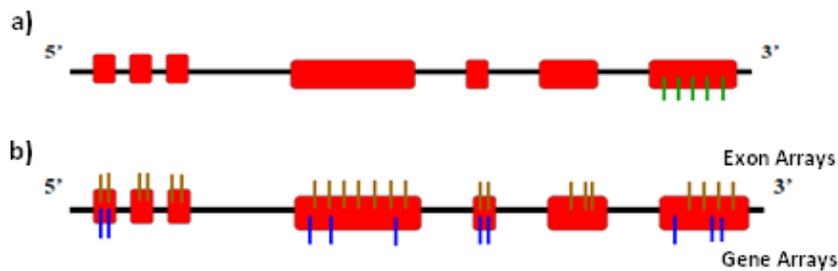
**Figura 11:** Proceso de hibridación entre las sondas del chip y los ácidos nucleicos de la muestra procesada (Affymetrix, 2005).

Tras la hibridación, se llevan a cabo una serie de lavados del chip para eliminar hibridaciones no específicas, así como para eliminar todo el RNA/sDNA no unido a sondas. Durante estos lavados también se lleva a cabo el paso de tinción mediante el cual se producirá la unión de las moléculas fluorescentes a las cadenas de RNA/sDNA unidas en el microarray. El último paso del procedimiento es el escaneo y lectura del microarray. El escáner contiene un láser que incidirá sobre el microarray dando lugar a la señal fluorescente final representando el nivel de expresión de cada transcrito.

#### 5.2.4. Microarray GeneChip® Rabbit Gene 1.0 ST Array

La nueva generación de microarrays comercializados por Affymetrix permiten una cobertura global del transcrito (Microarrays *whole-transcript* en inglés).

En esta nueva generación de microarrays cada uno de los genes está representado en el microarray por un set de sondas extendidas a lo largo de toda la longitud del gen, proporcionando una visión de la expresión más completa y precisa que los microarrays antiguos de expresión 3' (Figura 12).



**Figura 12:** Distribución de las sondas en los transcritos que interrogan los biochips Affymetrix. a) Distribución de las sondas en un ensayo clásico en el extremo 3'. b) Distribución de las sondas en un ensayo de cobertura total del transcrito.

Además del cambio en la distribución de las sondas en los transcritos, esta nueva generación de microarrays posee otras nuevas características. En el protocolo de preparación de la muestra diana, se genera DNA de cadena única en vez de cRNA, por lo que, tras el proceso de hibridación, se formarán dúplex de DNA-DNA, en vez de dúplex RNA-DNA, lo que proporciona una mayor especificidad.

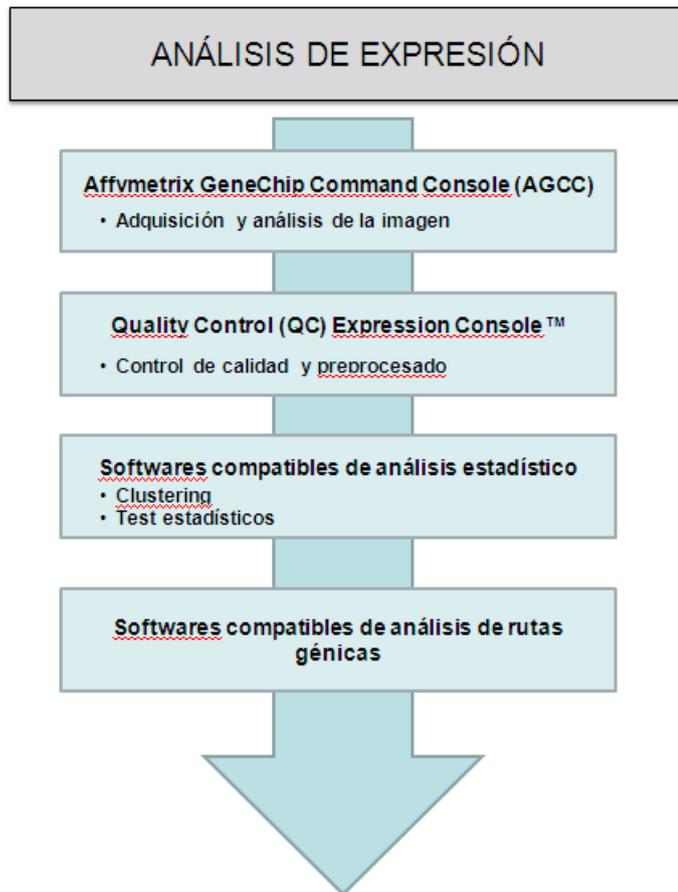
Otra característica nueva con respecto a los microarrays anteriores es la ausencia de los valores denominados *Mismatch* (MM), de forma que solo utiliza los denominados *Perfect match* (PM). En los microarrays anteriores a esta nueva generación para cada posición del transcrito objetivo existen dos tipos de sondas: una sonda PM y otra denominada MM. La sonda PM está formada por oligonucleótidos de DNA de cadena simple complementario al DNA objetivo, mientras que la sonda MM está formada por oligonucleótidos de DNA de cadena simple que es idéntica a la PM excepto por la substitución de una única base en la posición central del oligonucleótido. Así, en la última generación de biochips Affymetrix se han eliminado las sondas MM, de forma que se ahorra espacio en el microarray que puede ser utilizado para aumentar el número de sondas PM.

Dentro de estos nuevos microarrays actualizados se encuentra el *GeneChip® Rabbit Gene 1.0 ST Array* ofreciendo un análisis global de todo el genoma de conejo.

Este microarray identifica 23.282 genes mediante 496.321 sondas distintas, con una mediana de 22 sondas por gen. Para ello, utiliza secuencias de transcritos anotados procedentes de las bases de datos Orycun 2.0 y Ensembl (mayo 2009).

Por otra parte, los microarrays son fabricados de manera que puedan aportar información del rendimiento del procesado de la muestra, de la hibridación y del escaneado de la imagen. Con este propósito, cada uno de ellos contiene un conjunto de sondas que actúan como controles de calidad de la muestra. El valor de la intensidad que se obtiene de ellos es utilizado para generar los parámetros de control de calidad del microarray (ver apartado 6).

Además de toda la instrumentación y reactivos necesarios para llevar a cabo el análisis de las muestras, Affymetrix proporciona una serie de softwares para el análisis de los datos obtenidos que permiten desde la captura de la imagen recogida por el escáner hasta un software para llevar a cabo controles de calidad de los datos obtenidos. Finalmente, la información obtenida por Affymetrix también es compatible con un gran número de softwares que permiten realizar análisis estadísticos robustos y análisis de rutas biológicas a partir de los datos obtenidos, permitiendo la extracción y visualización de la información que muestra la tecnología de los microarrays (Figura 13).



**Figura 13:** Esquema de trabajo para el análisis de expresión de los datos obtenidos a partir del escaneado de los biochips *GeneChip*®.



## II. Objetivos

---

- Análisis transcriptómico en tejido ovárico de una línea seleccionada por tasa de ovulación en conejos mediante el uso de microarrays de expresión génica de Affymetrix.
- Identificación del perfil genómico global y de los genes más relevantes y significativos en la línea seleccionada por tasa de ovulación en conejos.



### **III. Material y métodos**

---

#### **1. Recogida del material biológico**

Se han analizado muestras de tejido ovárico de conejas nulíparas sacrificadas a las 16 horas post-monta. Tres conejas pertenecían a una línea seleccionada por tasa de ovulación durante 10 generaciones y las otras tres conejas pertenecían a la población base (grupo control).

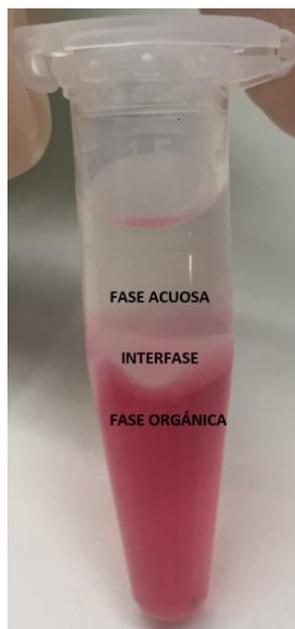
Tras el sacrificio, se pesaron entre 34-40 mg de tejido ovárico al que se le añadió 1 ml de RNA later®.

Las muestras se congelaron a -80°C hasta el momento de realizar la extracción del RNA total.

#### **2. Extracción del RNA total**

El protocolo utilizado para la extracción de RNA total está basado en el método descrito por Chomczynski y Sacchi (Chomczynski, 1987). Para la homogeneización del tejido ovárico (34-40 mg), se sumergen las muestras en 1 ml de "Tripure Isolation Reagent®" (Roche Group S.A.), solución que contiene una mezcla estable de fenol y tiocianato de guanidina (solución de lisis) y se utiliza un homogeneizador para tubos de 1,5-2 ml (Ultra Turrax). A continuación, se añade un volumen de cloroformo en relación 1:5 con la solución de lisis que se mezcla homogéneamente mediante la inversión repetida y vigorosa. Posteriormente, se deja reposar durante 3 minutos a temperatura ambiente. Tras centrifugar durante 15 minutos a 14.000 r.p.m. (revoluciones por minuto) se obtienen tres fases (Figura 14): la fase acuosa más superficial, incolora, es la que contiene el RNA total de la muestra. Esta fase acuosa se encuentra separada de otra fase orgánica de color rosáceo por una interfase blanquecina de composición proteica. En la fase orgánica se encuentran disueltos tanto el DNA de la muestra como la solución de lisis (restos de "Tripure Isolation Reagent®") y diferentes compuestos fenólicos de las muestras.

Finalmente, se recoge la fase acuosa pipeteando con cuidado para no contaminar el RNA con el resto de las fases.



**Figura 14.** Esquema representativo de las tres fases obtenidas al adicionar el cloroformo a las muestras homogeneizadas. **Fase acuosa:** Fracción incolora que contiene el RNA total de la muestra. **Interfase proteica:** de color blanquecino y de muy pequeño volumen. **Fase orgánica:** de color rosáceo debido al Tripure Isolation Reagent © que contiene DNA y compuestos fenólicos.

Para realizar la fase de precipitación del RNA, se añade al sobrenadante un volumen de isopropanol en relación 1:2 con la solución de lisis añadida durante la fase de homogeneización, y 2  $\mu\text{L}$  de glucógeno, que actúa como coadyuvante del proceso y aporta al precipitado un color blanquecino. Esta mezcla se incubó toda la noche a una temperatura de  $-20\text{ }^{\circ}\text{C}$  para permitir que se forme el precipitado de RNA. Posteriormente, se centrifugan las muestras durante 15 minutos a 14.000 r.p.m. obteniéndose el pellet en el fondo del microtubo. Este pellet se lava con un volumen de etanol al 75%, previamente preparado y frío, en relación 1:1 con la solución de lisis, se agita con un vortex y se vuelve a centrifugar a 14.000 r.p.m. durante 15 minutos. Tras la última centrifugación se elimina el etanol completamente, eliminando todo el volumen posible ayudándonos con una pipeta y dejando evaporar el resto de volumen de etanol durante 5 minutos a temperatura ambiente.

Finalmente, el pellet se resuspende en 100  $\mu\text{L}$  de agua estéril libre de RNasas y se almacenan las muestras obtenidas a  $-80\text{ }^{\circ}\text{C}$  hasta su posterior uso.

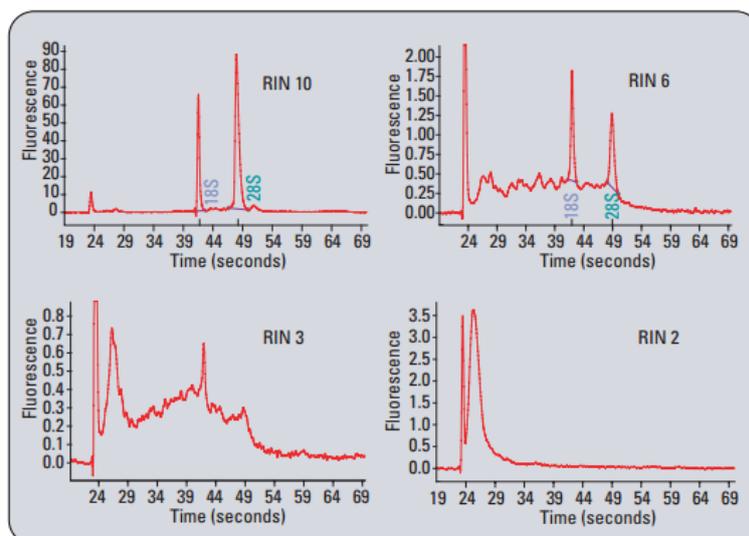
### 3. Control de la calidad del RNA total obtenido

#### 3.1. Electroforesis capilar

La integridad del RNA total extraído se determina con el Bioanalizador Agilent 2100 (Agilent Technologies), el cual ofrece un amplio rango de kits de análisis. Se utilizó el kit RNA 6000 Nano LabChip®, apropiado para el análisis de muestras de RNA. Cada chip contiene microcanales interconectados que se utilizan para la separación electroforética de ácidos nucleídos en función del tamaño. Los análisis basados en chips ofrecen diferentes ventajas sobre las tecnologías existentes, entre las que cabe destacar, la reducida cantidad de muestra necesaria y de materiales peligrosos, así como el aumento de la velocidad de análisis y de la precisión de los datos.

La preparación de este tipo de chips se lleva a cabo siguiendo las instrucciones del fabricante. Los reactivos necesarios para la preparación de los LabChip® se dejaron 30 minutos a temperatura ambiente antes de su utilización. Para preparar el gel se utilizó una columna de 550  $\mu$ l, la cual se centrifuga a 4.000 r.p.m. durante 10 minutos. Posteriormente, se recoge el gel filtrado y se divide en alícuotas de 65  $\mu$ l. Tras la adición de 1  $\mu$ l de colorante al gel, la mezcla gel-colorante es inyectada dentro del sistema de canales. Tras la adición de 5  $\mu$ l del marcador endógeno (DNA fragmentado en un tamaño de 50 pb y usado para alinear todas las muestras) tanto en el pocillo del marcador de pesos moleculares como en los pocillos de las muestras, se introduce en cada uno de los pocillos de muestra 1  $\mu$ l de RNA total de las correspondientes muestras a analizar y 1  $\mu$ l de marcador de pesos moleculares (mezcla de RNA de diferentes tamaños a distinta concentración) en su correspondiente pocillo. Las muestras de RNA y el marcador de pesos moleculares son anteriormente preparados mediante una reacción de desnaturalización a 70 °C durante 2 minutos. Finalmente, los chips se agitan durante 1 minuto a máxima potencia en un vórtex con un acoplador apropiado para el LabChip®, y se introducen en el Agilent 2100 Bioanalyzer. Para llevar a cabo la visualización de los resultados, se utiliza el software 2100 expert v. B.02.08.SI648.

Para la determinación de la integridad del RNA total, el bioanalizador usa fluorescencia entre 670 nm y 700 nm. Para cada muestra, el software calcula el valor de integridad RIN (acrónimo del inglés *RNA Integrity Number*) mediante un algoritmo que evalúa diferentes parámetros de integridad en las medidas electroforéticas del RNA (Agilent Technologies, 2005). La integridad del RNA es un parámetro de gran importancia en los estudios de expresión génica. Tradicionalmente se ha evaluado calculando el ratio 28S/18S de RNA ribosómico (rRNA), ya que la degradación del RNA provoca un descenso en el ratio de bandas entre los ribosomas 28S y 18S. Sin embargo, gran cantidad de estudios han demostrado la inconsistencia de este método. El algoritmo RIN toma valores entre 1 (muestra muy degradada) y 10 (muestra íntegra), y permite comparar directamente las muestras de RNA y garantizar la reproducibilidad de los experimentos (Figura 15). Los valores de RIN igual o superior a 7 se consideran óptimos para un análisis de todo el transcriptoma.



**Figura 15:** Electroferogramas ejemplo de muestras de RNA con diferentes integridades y por tanto con diferentes valores de RIN. La integridad de la muestra no solo es determinada por el ratio de las bandas ribosomales entre 28S y 18S sino por todas las características del electroferograma. Se puede observar desde una muestra con una integridad total (RIN=10) hasta una muestra con una gran degradación (RIN=2) (Mueller, Lightfoot, & Schroeder, 2004).

### 3.2. Espectrofotometría

Para llevar a cabo tanto la cuantificación de la concentración como la determinación del grado de pureza de las muestras se utilizó el espectrofotómetro GeneQuant Pro™ (GE Healthcare). Para cada muestra se realiza una disolución 1:3 del RNA total (1  $\mu$  de la muestra obtenida y 2  $\mu$ l de agua libre de RNAsas). Las diluciones obtenidas se introducen en capilares de cuarzo para realizar la medición en el espectrofotómetro. La medida de referencia o blanco se realiza con 3  $\mu$ l de agua libre de RNAsas.

Los ácidos nucleídos absorben eficientemente luz ultravioleta debido a la presencia en sus cadenas de bases aromáticas nitrogenadas. La absorción de radiación ultravioleta del DNA y del RNA es una característica específica de estas moléculas, y por tanto es usada eficientemente para determinar su concentración. Así, para calcular la cantidad de RNA en una muestra se utiliza la siguiente conversión: 1 unidad de absorbancia a 260 nm de longitud de onda ( $A_{260}$ ) de RNA de cadena simple es igual a 40  $\mu$ g/mL.

Para comprobar la pureza de la muestra, se mide espectrofotométricamente la absorbancia a 260 nm ( $A_{260}$ ) y a 280 nm ( $A_{280}$ ), ya que mientras los ácidos nucleídos tienen su máximo de absorción a 260 nm, las proteínas lo tienen a 280 nm. De esta manera, observando la relación entre ambos valores de absorbancia ( $A_{260}/A_{280}$ ) podremos saber si nuestro RNA total extraído tiene una alta pureza o, si, por lo contrario, está contaminado con restos de proteínas. Un ratio  $A_{260}/A_{280}$  cercano a 2,0 indica una muestra de gran pureza (Figura 16).

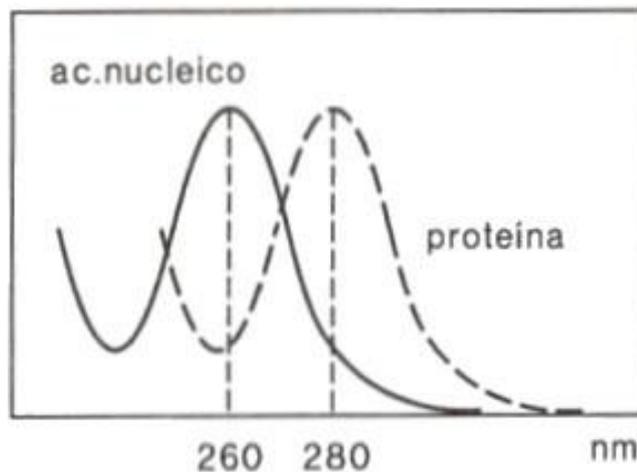


Figura 16: Gráfica de la absorbancia en nm de los ácidos nucleicos y las proteínas.

## 4. Protocolo GeneChip® Rabbit Gene 1.0 ST Array

Se ha utilizado el microarray GeneChip® Rabbit Gene 1.0 ST Array en el presente trabajo por disponer de una química más actualizada y eficiente.

El protocolo para la realización del microarray GeneChip® Rabbit Gene 1.0 ST Array se lleva a cabo mediante una serie de procesos que incluyen la síntesis del cDNA de cadena simple con sentido a partir del RNA total extraído de las muestras, su fragmentación y marcaje, y finalmente la hibridación y escaneo del microarray. Durante todo el proceso se necesita una serie de instrumentación, software y reactivos específicos que se muestra en el Tabla 1.

**Tabla 1:** Productos requeridos para la realización del protocolo *GeneChip® Rabbit Gene 1.0 ST Array*.

<b>INSTRUMENTACIÓN</b>			
Escáner	<i>GeneChip® Scanner 3000 7G</i>		
Estación de fluidos	<i>GeneChip® Fluidics Station 450</i>		
Horno de hibridación	<i>GeneChip® Hybridization Oven 645</i>		
<b>SOFTWARE</b>			
Adquisición y análisis de imágenes	<i>Software Affymetrix</i>	<i>GeneChip®</i>	<i>Command Console® (AGCC)</i>
<b>REACTIVOS</b>			
Preparación de las muestras	<i>Kit GeneChip® WT PLUS Reagent</i>		
Fragmentación y marcaje final	<i>Kit GeneChip® WT Terminal Labeling</i>		
Control del marcaje	<i>Kit GeneChip® Poly-A RNA Control</i>		
Control de la hibridación	<i>Kit GeneChip® Hybridization Control</i>		
Hibridación, lavado y tinción	<i>Kit GeneChip® Hybridization, Wash, and Stain</i>		

El esquema de trabajo a seguir para llevar a cabo el protocolo completo en cada muestra y está dividido en tres días (Tabla 2), variable dependiendo del número de muestras, tras los cuales se obtiene una imagen (archivo .DAT) que se procesa para obtener los datos de expresión.

**Tabla 2:** Esquema de trabajo para la realización del protocolo *GeneChip® Rabbit Gene 1.0 ST Array*.

### **DÍA 1**

- Síntesis de la primera cadena de cDNA
- Síntesis de la segunda cadena de cDNA
- Síntesis del cRNA mediante transcripción in vitro

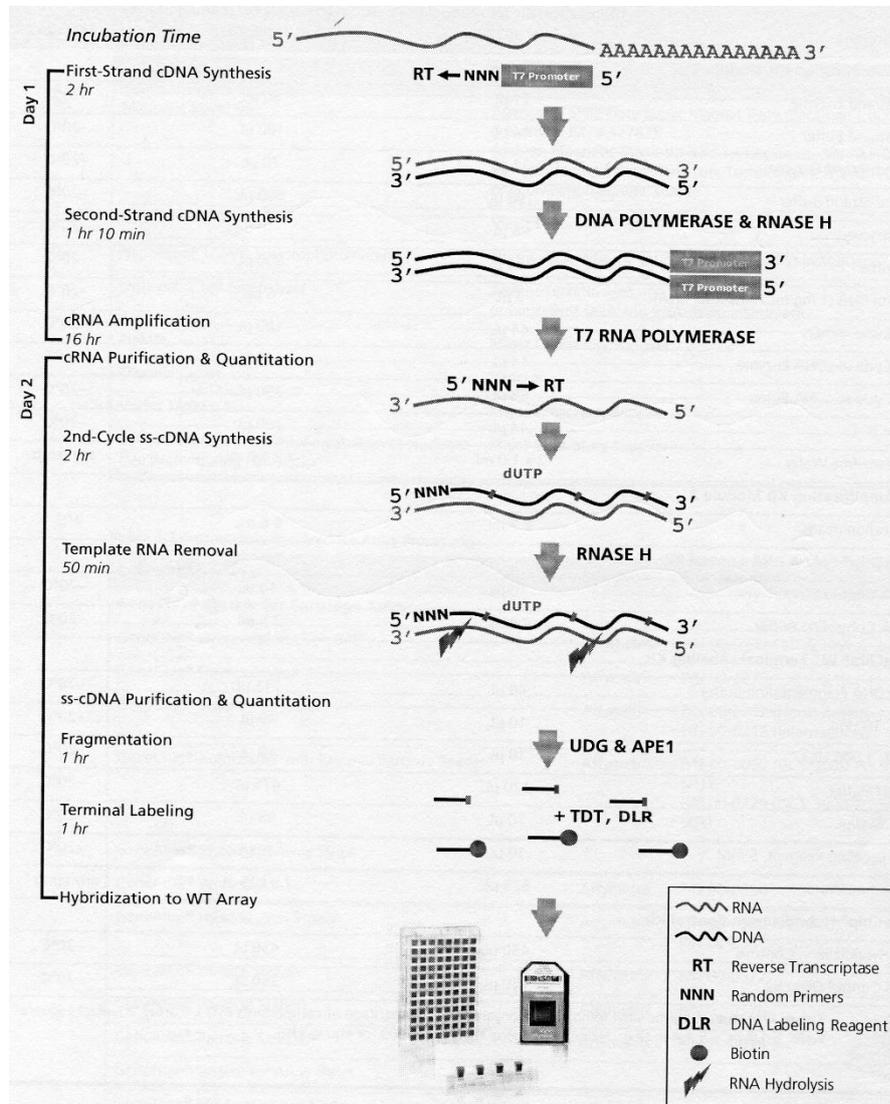
### **DÍA 2**

- Purificación del cRNA
- Evaluación de la concentración del cRNA
- Segundo ciclo de síntesis de cDNA
- Hidrólisis con RNAasa H
- Purificación del cDNA de simple cadena
- Evaluación de la concentración del cDNA
- Fragmentación y marcaje del cDNA de simple cadena
- Hibridación de las muestras en los biochip (comienzo)

### **DÍA 3**

- Hibridación de la muestra en el biochip (finalización)
- Lavados y tinción
- Escaneo de los microarrays

El protocolo se resume en la figura 17 donde se muestran las diferentes etapas del proceso:



**Figura 17:** Esquema del protocolo *GeneChip® Rabbit Gene 1.0 ST Array* modificado de Affymetrix, 2009-2010.

El protocolo de amplificación, fragmentación y marcaje se realiza siguiendo las instrucciones del fabricante mediante el protocolo *GeneChip® WT PLUS Reagent Kit* (Affymetrix).

Se inicia mediante un proceso de transcripción reversa de RNA a cDNA. A partir del cDNA se sintetiza y amplifica la cadena antisentido del cRNA mediante la transcripción *in vitro* usando la secuencia de la T7 RNA polimerasa. La cadena de cDNA es sintetizada por transcripción reversa del cRNA usando cebadores aleatorios. Después de esta síntesis, se purifica y se añaden los desoxirribonucleótidos trifosfato (dNTPs) para preparar el DNA de

cadena simple para su fragmentación y marcaje. La cadena de DNA es tratada con la combinación de UDG (uracil-DNA glicosilasa) y APE1 (endonucleasa 1 apurínica/apirimidínica) que reconocen específicamente los residuos de desoxiuridina trifosfato (dUTP) generando fragmentos entre 25 y 200 bases de longitud. Tras la fragmentación, el DNA es marcado mediante la enzima desoxinucleotidil transferasa terminal (TdT), cuyo reactivo lleva unido covalentemente moléculas de biotina. La muestra fragmentada y marcada es añadida al cocktail de hibridación que contiene reactivos y controles de hibridación. Cada cocktail será inyectado en un biochip de GeneChip® Rabbit Gene 1.0 ST Array (Affymetrix) que integra a 23,282 genes, y se lleva a cabo la reacción de hibridación en el horno Genechip Hybridization Oven 645 (Affymetrix) a 45°C durante 16 horas.

En esta etapa las muestras se hibridan por complementariedad a las sondas presentes en el microarray. Para eliminar las uniones inespecíficas y poder generar la señal fluorescente. Los microarrays se pasan por una estación de lavado Genechip Fluidics Station 450 (Affymetrix), y en ella la molécula ficoeritrina-streptavidina se une a la biotina presente en la muestra de forma covalente siguiendo el protocolo FS450\_0002. Finalmente, los microarrays se digitalizan utilizando el Scanner GCS 3000 7G (Affymetrix), obteniéndose una imagen integrada por el conjunto de intensidades de fluorescencia emitidas a partir del perfil de expresión de todos los genes presentes con la ayuda del software *Affymetrix GeneChip® Command Console*.

## 5. Preprocesado de los datos de intensidad brutos

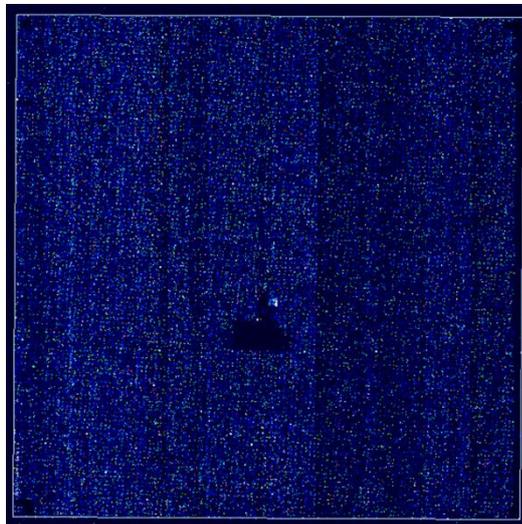
Tras el escaneado de los biochips, obtenemos una imagen del biochip que es la imagen digital bruta, (almacenada en un archivo .DAT). Posteriormente, mediante el software de Affymetrix *GeneChip® Command Console*, se lleva a cabo un procesado de dicha imagen, dando lugar a un archivo. CEL que contiene, para cada celda, un valor de intensidad de fluorescencia. Este fichero es el resultado de la segmentación de la imagen obtenida y contiene la media y desviación estándar de los niveles de intensidad, así como la localización de la sonda dentro del microarray es un archivo de datos a nivel de sonda. Sin embargo, en los microarrays Affymetrix *GeneChip®* cada valor de intensidad no corresponde a la expresión de un gen, ya que hay múltiples valores (sondas) por cada gen. Así, estos valores deben ser combinados para cada set de sondas (*probeset*), generando un valor de expresión por cada gen presente en el chip (sumarización, del inglés *summarization*). Para llevar a cabo esta sumarización se necesita no solo la información que aporta el archivo. CEL sino también conocer la correspondencia entre las sondas y los nombres de los genes. Esta información aparece en el fichero. CDF. Por eso, a partir del preprocesado de los datos a nivel de sonda se obtienen los datos de expresión para cada gen contenido en el microarray. En el presente trabajo se utiliza para el preprocesado el algoritmo RMA (Media robusta de multiples microarrays, acrónimo del inglés *Robust Multiarray Average*) que es el indicado para este tipo de microarrays y que permite la corrección del ruido de fondo basándose solo en valores PM.

## 6. Control de calidad de los biochips.

Antes de llevar a cabo el análisis de expresión génica, se debe realizar un adecuado control de calidad para cada uno de los chips, con el objetivo de asegurar que los datos obtenidos tienen la calidad suficiente para ser utilizados en el análisis. Para llevar a cabo dicho control de calidad, se siguen una serie de pasos que se detallan a continuación.

### 1) Análisis de la imagen del biochip.

En primer lugar, se lleva a cabo la inspección de la imagen obtenida (.DAT) a través del software de Affymetrix *GeneChip® Command Console*, lo cual proporciona información acerca del balance del color, de la uniformidad en la hibridación y en los spots, de si el efecto de ruido es mayor del esperado y de la existencia de artefactos (Figura 18).

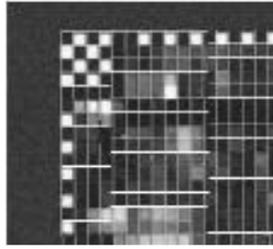


**Figure 18:** Imagen digital obtenida del biochip (.DAT) tras el escaneado.

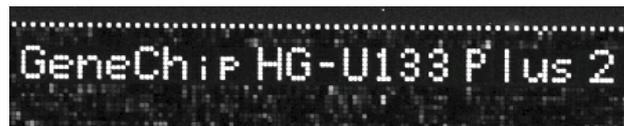
### 2) Control de la hibridación del oligo B2

Para que el software *GeneChip® Command Console* lleve a cabo el procesamiento de la imagen y de lugar al archivo con los datos brutos de intensidad (.CEL), se debe producir el alineamiento de una gradilla sobre la imagen. Para ello se utiliza el oligo B2, un oligo sintético de control añadido al cóctel de hibridación y que permite el alineamiento automático de la gradilla sobre la imagen, ya que los límites de la zona donde se encuentran las sondas son fácilmente identificables mediante la acción de dicho oligo. De esta forma es necesaria una correcta hibridación del oligo B2 para asegurar un correcto alineamiento por parte del software *GeneChip® Command Console* y así asegurar que los datos resultantes son correctos. La correcta hibridación del oligo B2 se observa analizando la imagen a través de:

- A. El patrón de tablero de ajedrez en cada esquina del array (Figura 19). En el chip *GeneChip® Rabbit Gene 1.0 ST* la imagen es dividida en una cuadrícula de 9x9 cuadrados donde cada uno de ellos contiene los patrones de hibridación del oligo B2 en cada esquina.
- B. El nombre de la array, que se encuentra en la esquina superior izquierda o media-alta de la matriz (Figura 20).



**Figura 19:** Ejemplo de la acción del oligo B2 iluminando las esquinas del array.



**Figura 20:** Ejemplo de la acción del oligo B2 mostrando el nombre del array.

Variaciones en la intensidad de hibridación del oligo B2 a lo largo del array son normales y no indican variaciones en la eficiencia de hibridación.

### 3) Controles de calidad de los datos obtenidos.

Para comprobar la calidad de los datos obtenidos durante el preprocesado de los datos de intensidad en el software *Affymetrix Expression Console*, se calculan diversos parámetros de calidad que permiten identificar tanto microarrays de baja calidad como la fuente del problema:

#### A. Parámetros a nivel de sonda.

Son aquellos parámetros obtenidos con los datos procedentes del archivo .CEL, es decir datos sin preprocesar (a nivel de sonda). El parámetro “**pm\_mean**” proporciona la media de las intensidades para todas las sondas presentes en el array. Este valor permite determinar si ciertos microarrays son inusualmente tenues o brillantes, lo cual aunque no supone un problema de por sí, sí merecen un exámen más detallado.

#### B. Parámetros a nivel de set de sondas

Estos parámetros están basados en valores de señal para set de sondas, es decir, valores obtenidos tras el preprocesado. Principalmente, se muestran los valores de señal obtenidos para los genes control de hibridación y para los genes control de marcaje ya que en ambos casos es importante comprobar los adecuados niveles de expresión obtenidos. Los controles de hibridación y de marcaje son los siguientes:

#### -Controles de hibridación: *bioB*, *bioC*, *bioD*, y *cre*.

*BioB*, *bioC*, y *bioD* representan genes de la ruta metabólica de síntesis de la biotina en *E. scherichia coli* y *cre* es el gen de la recombinasa del bacteriófago P1.

El kit control de hibridación contiene una mezcla de transcritos de cRNA marcados con biotina de los genes *bioB*, *bioC*, *bioD* y *cre*. Estos controles son añadidos dentro del cóctel

de hibridación, independientemente de la preparación de la muestra, y son usados para evaluar la eficiencia de hibridación de la muestra en arrays de genes eucariotas. Además, estos controles de hibridación se pueden utilizar para evaluar indirectamente la calidad del RNA.

El nivel de expresión para los cuatro controles bacterianos debe ser *bioB*<*bioC*<*bioD*<*cre*.

#### **-Controles (poli-A) de marcaje: *lys*, *phe*, *thr* y *dap***

Los controles de RNA con cola poli-A pueden ser usados para monitorizar el proceso de etiquetado. Cada array eucariota *GeneChip*<sup>®</sup> contiene sets de sondas de varios genes de la bacteria *Bacillus subtilis* los cuales están ausentes en las muestras de eucariotas (*lys*, *phe*, *thr* y *dap*). Estos controles son sintetizados *in vitro*, modificados con la adición de colas poli-A y clonados en los vectores pBluescript que contienen la secuencia del promotor T3. Amplificando estos controles con la RNA polimerasa T3, se generan RNAs, los cuales se añaden a la muestra a analizar. De este modo, se pueden realizar controles en el proceso de preparación de la muestra, evaluándolos como genes control internos.

Al igual que en caso anterior el nivel de expresión para los cuatro controles poli-A debe ser *lys*<*phe*<*thr*<*dap*.

#### **C. Parámetros a nivel de categorías**

Para llevar a cabo exámenes mas detallados, se dispone de serie de parámetros denominados "Parámetros por categorías".

Los principales son las siguientes:

-**"all\_probeset"**: Contiene todos los sets de sondas analizadas. Los parámetros obtenidos para esta categoría serán los más representativos de la calidad general de los datos obtenidos.

-**"bac\_spike"**: Contiene es el set de sondas que hibrida con los controles de hibridación bacterianos (*bioB*, *bioC*, *bioD* y *cre*). Esta categoría es útil para identificar problemas con la hibridación o el microarray. Los valores obtenidos para esta categorías tendrán más variabilidad que otras categorías debido al número limitado de sets de sondas en esta categoría.

-**"polya\_spike"**: Contiene es el set de sondas que hibridan con los controles de marcaje poliadenilados (*lys*, *phe*, *thr*, y *dap*). Por tanto, esta categoría es útil para identificar problemas en la preparación de la muestra objetivo. Los valores obtenidos para esta categoría tendrán, al igual que ocurre con la categoría anterior, más variabilidad debido al limitado número de set de sondas que contiene.

-**"pos\_control"**: Contiene el conjunto de sets de sondas diseñadas para evaluar regiones exónicas de genes constitutivos. Por tanto, esta categoría es usada para estimar el ratio de verdaderos positivos.

- **"neg\_control"**: Contiene el conjunto de sets de sondas diseñadas para evaluar regiones intrónicas de genes constitutivos. Por tanto, esta categoría es usada para estimar el ratio de falsos positivos.

Las categorías "pos\_control" y "all\_probeset" son útiles para tener una visión global de la calidad de los datos de cada microarray. Los parámetros basados en estas categorías reflejan por tanto la calidad general del experimento (muestra inicial, procesado de la

muestra, array, hibridación, escaneado, alineamiento, etc) y la naturaleza de los datos que se utilizan en análisis estadísticos posteriores. Por otro lado, la categoría “polya\_spike” es útil para identificar problemas con la preparación de la muestra, mientras que la categoría “bac\_spike” es útil para la identificación de problemas con la hibridación y el array.

## 7. Análisis transcripcional

Los pasos descritos en las secciones anteriores son pasos preparatorios para el análisis final de los datos, el análisis transcripcional global.

Para la realización del análisis transcriptómico, se ha utilizado el software comercial Partek Genomics Suite 6.6, el cual proporciona una solución completa y optimizada para el eficiente y rápido análisis de datos genómicos mediante el Análisis Componentes Principales (ACP).

### -Análisis de Componentes principales.

El ACP es una técnica utilizada para reducir la dimensionalidad de un conjunto de datos. Intuitivamente la técnica sirve para hallar las causas de variabilidad de un conjunto de datos y ordenarlas por importancia. Técnicamente un ACP busca la proyección según la cual los datos quedan mejor representados en términos de mínimos cuadrados. Así, se transforma el conjunto original de variables en otro conjunto de nuevas variables **incorreladas** entre sí (no tienen repetición o redundancia en la información) llamado conjunto de componentes principales. Al analizar la expresión de miles de genes se debe realizar ACP para reducir la dimensionalidad y utilizar únicamente las componentes principales que explican mejor la variabilidad de las muestras. Estas nuevas variables son combinaciones lineales de las anteriores y se van construyendo según el orden de importancia en cuanto a la variabilidad total que recogen las muestras.

### -Análisis de expresión diferencial se analizara mediante un análisis de la varianza (ANOVA).

A la hora de aceptar o rechazar la hipótesis nula (no existe expresión diferencial entre las diferentes condiciones) se utiliza el estadístico p-valor. Dado que en este tipo de análisis no se evalúa un solo gen aislado, sino que se evalúan miles simultáneamente pueden aparecer cientos de falsos positivos. Para evitarlo, existen diferentes métodos que controlan la tasa de falsos positivos en el presente trabajo se utiliza el procedimiento propuesto por Benjamini and Hochberg (Benjamini & Hochberg, 1995) donde se asume que al contrastar muchas hipótesis se pueden tolerar algunos errores tipo I, siempre que el número de estos errores sea pequeño en relación con el número de hipótesis que se rechazan. Es lo que se conoce como control de FDR (Proporción esperada de falsos positivos, acrónimo del inglés *False Discovery Rate*), definido como la proporción de falsos positivos entre todos los genes inicialmente identificados como expresados diferencialmente. El método esencialmente se basa en corregir el p-valor original, obteniendo un p-valor ajustado. En el presente trabajo se utiliza un valor  $\alpha=0,05$  como tasa de error por lo que se considera significativa la expresión de los p-valores ajustados  $\leq 0,05$ .

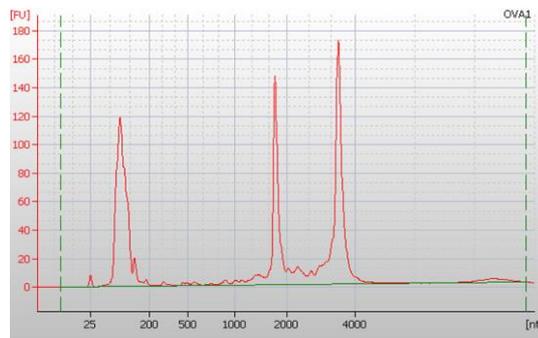


# IV. Resultados

---

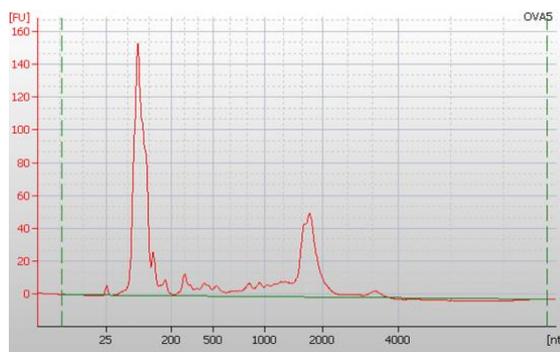
## 1. Control de calidad de las muestras

Una vez recogidas las seis muestras de tejido ovárico y realizada la extracción del RNA total de cada una de ellas, se lleva a cabo un control de la integridad de las muestras mediante el bioanalizador Agilent 2100 (Agilent Technologies) para comprobar la calidad de las mismas. El electroferograma (Figura 21) permite una inspección visual de la integridad del RNA donde se observan los picos correspondientes a 5S, 18S y 28S y, por tanto, indica que la muestra no está degradada y es óptima para el estudio. Los valores obtenidos para el ratio 28S/18S de cada muestra analizada en el presente trabajo se observa en la Tabla 3.



**Figura 21:** Ejemplo de electroferograma modelo en el cual se observa el patrón de picos que se debe obtener en una muestra no degradada.

En nuestro experimento nos encontramos con muestras degradadas que no pasaron el control de calidad de integridad del RNA y, por tanto, se descartaron en el estudio (Figura 22).



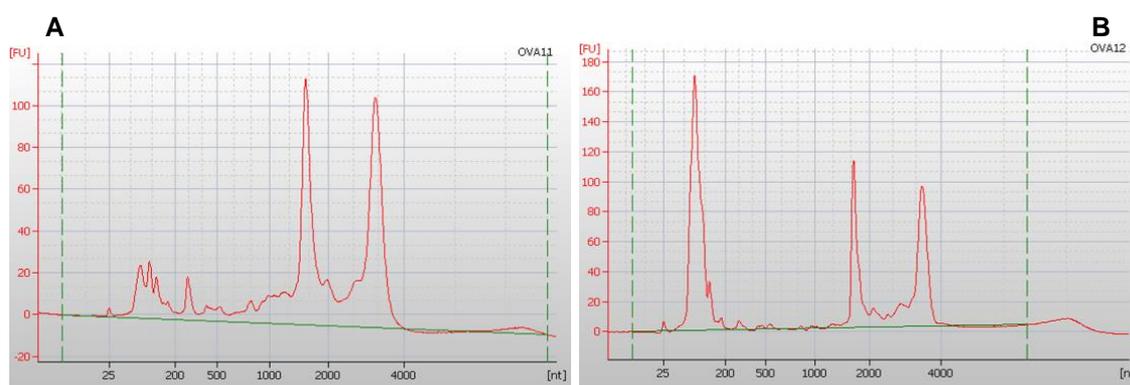
**Figura 22:** Ejemplo de electroferograma degradado donde el 28S no se visualiza.

La tabla 3 también muestra los valores de RIN para la mayoría de las muestras analizadas. Solo una de las muestras presenta valor RIN adecuado mientras en el resto de muestras no se puede especificar. Este resultado es debido a que cuando la región 5S es muy grande no puede calcular el valor de RIN de forma adecuada.

**Tabla 3:** Valores del ratio 28S/18S y RIN (*RNA Integrity Number*) obtenidos para las muestras de tejido ovárico analizadas pertenecientes a la población base (grupo control) y generación 10 (grupo selección) seleccionadas por tasa de ovulación.

Muestras	Grupo control/selección	RIN	Ratio 28S/18S
1	Control	Sin especificar	1,30
2	Control	Sin especificar	0,60
3	Control	Sin especificar	0,60
4	Selección	Sin especificar	1,00
5	Selección	8,3	1,50
6	Selección	Sin especificar	1,10
<b>Media</b>		-----	<b>1,02</b>
<b>Desviación estándar</b>		-----	<b>0,37</b>

Al comprar los electroferogramas de la muestra que presenta un valor de RIN óptimo (RIN 8,3) con las otras muestras se observan patrones similares y óptimos para el estudio del transcriptoma (Figura 23).



**Figura 23:** Ejemplo de dos electroferogramas de las muestras realizadas con el Bioanalizador de Agilent. A. Electroferograma de una muestra con valor de RIN 8,3. B. Electroferograma de una muestra con valor de RIN no especificado.

## 2. Control de pureza y cuantificación de RNA total

El control de pureza y cuantificación del RNA total de las seis muestras se realiza mediante el miniespectrofotómetro GeneQuant pro™ (Biochrom). Los valores del ratio A260/A280 (Tabla 4) para todas las muestras son cercanos a 2,00, con una media de 1,96, indicando que las preparaciones presentan una pureza adecuada. La concentración de RNA total en las muestras es elevada y similar, atribuyéndolo a un rendimiento del proceso de extracción muy homogéneo.

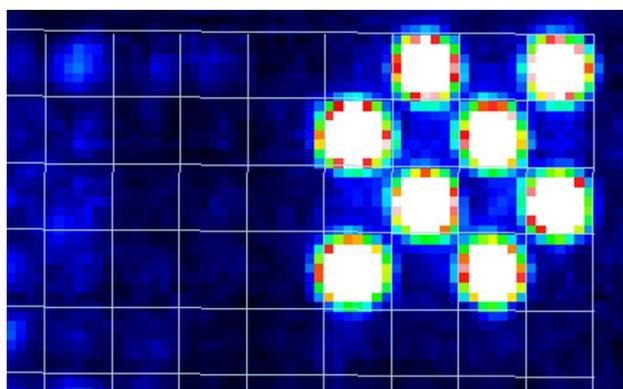
**Tabla 4:** Valores de concentración y pureza (ratio A260/A280) para las muestras de tejido ovárico analizadas en el presente trabajo.

<b>Muestras</b>	<b>Grupo control/selección</b>	<b>Concentración (ng/ul)</b>	<b>Ratio A260/A280</b>
<b>1</b>	Control	787	1,94
<b>2</b>	Control	1829	1,90
<b>3</b>	Control	1826	1,90
<b>4</b>	Selección	1908	2,05
<b>5</b>	Selección	1471	1,90
<b>6</b>	Selección	1022	2,04
<b>Media</b>		<b>1474</b>	<b>1,96</b>
<b>Desviación estándar</b>		<b>472</b>	<b>0,07</b>

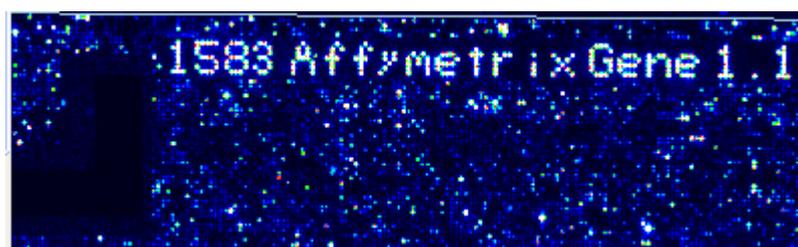
### 3. Control de calidad de los biochips

Tras escaneado los seis microarrays, se inspeccionaron las imágenes de los archivos .DAT mediante el software *GeneChip® Command Console®* (AGCC). Así, se comprueba la inexistencia de artefactos y la existencia de una intensidad global uniforme en la matriz. Posteriormente, se analizan los patrones resultantes de la hibridación del oligo B2.

Las seis muestras analizadas en el presente trabajo muestran unos resultados adecuados, ya que en todas ellas se observaba el patrón de tablero de ajedrez en cada esquina de las 81 zonas en las que se divide el microarray (Figura 24), así como el nombre del microarray en el borde inferior izquierdo del microarray (Figura 25). Como consecuencia de la correcta hibridación del oligo B2, la cuadrícula fue alineada correctamente de forma automática por el software, de modo que se llevó a cabo la creación de los archivos .CEL para cada una de las muestras analizadas.



**Figura 24:** Esquina superior derecha de la imagen obtenida tras el escaneo y lectura de biochip. Se observa el patrón de ajedrez típico formado por la hibridación del oligo B2.



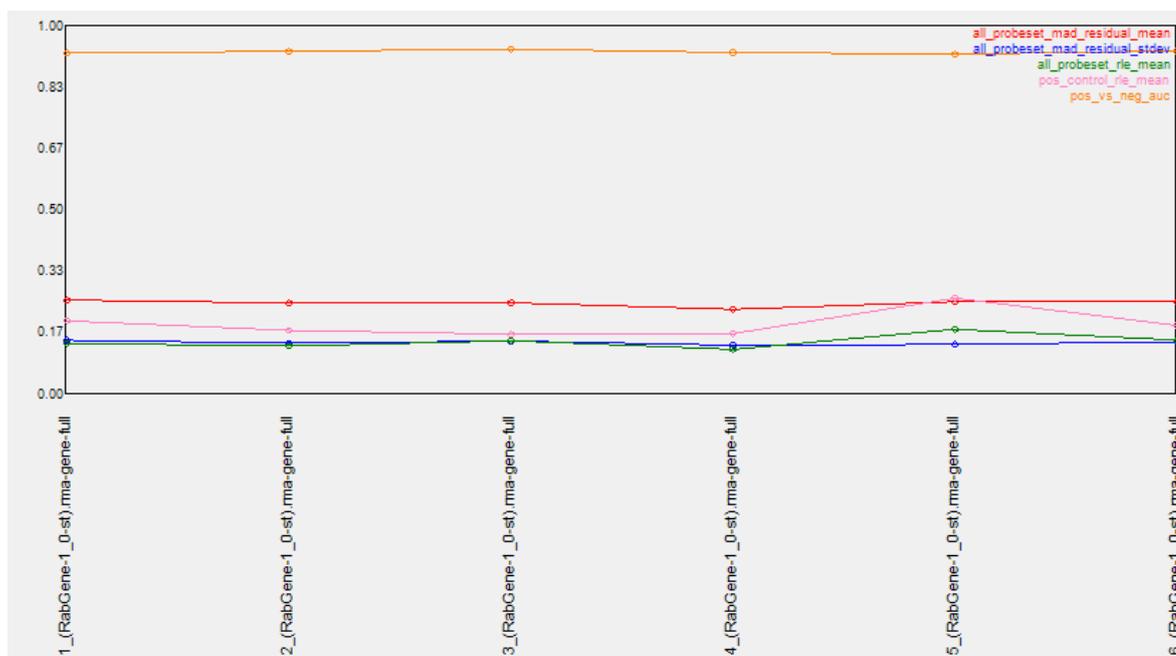
**Figura 25:** Parte de la imagen obtenida tras el escaneo y lectura de biochip donde se observa el nombre formado por la hibridación del oligo B2.

Con el uso del software *Expression Console*™ 1.4.1.46 de Affymetrix se realiza el pre-procesado de las muestras mediante el método RMA. Este software además proporciona una tabla con todos los parámetros necesarios para llevar a cabo un control de calidad de los datos de expresión obtenidos.

### -Visión global de la calidad de los datos de expresión

Las categorías “positive control” y “all probeset” son útiles para tener una visión global de la calidad de los datos de cada chip. Los parámetros basados en estas categorías por tanto reflejan la calidad general del experimento y la naturaleza de los datos que se utilizan en los análisis estadísticos posteriores.

En primer lugar, en la Figura 26 se observa el parámetro “pos vs neg auc”, el cual permite conocer en qué grado las señales de las sondas separan los controles positivos de los controles negativos. En consecuencia, valores cercanos a 1 reflejan una separación perfecta. La media obtenida en seis microarrays son de  $0,93 \pm 0,01$ , muy próximos a 1 y por encima del valor umbral (valor umbral= 0,8), lo que permite verificar una correcta separación entre falsos positivos y verdaderos positivos.



**Figura 26:** Desviación media absoluta de los residuos (“*mad residual mean*”) y media absoluta de la expresión relativa en logaritmo en base 2 (“*RLE mean*”) para las categorías “*all probeset*” y “*positive control*”, y área bajo la curva relativa (AUC) entre los controles positivos y negativos (“*pos vs neg auc*”).

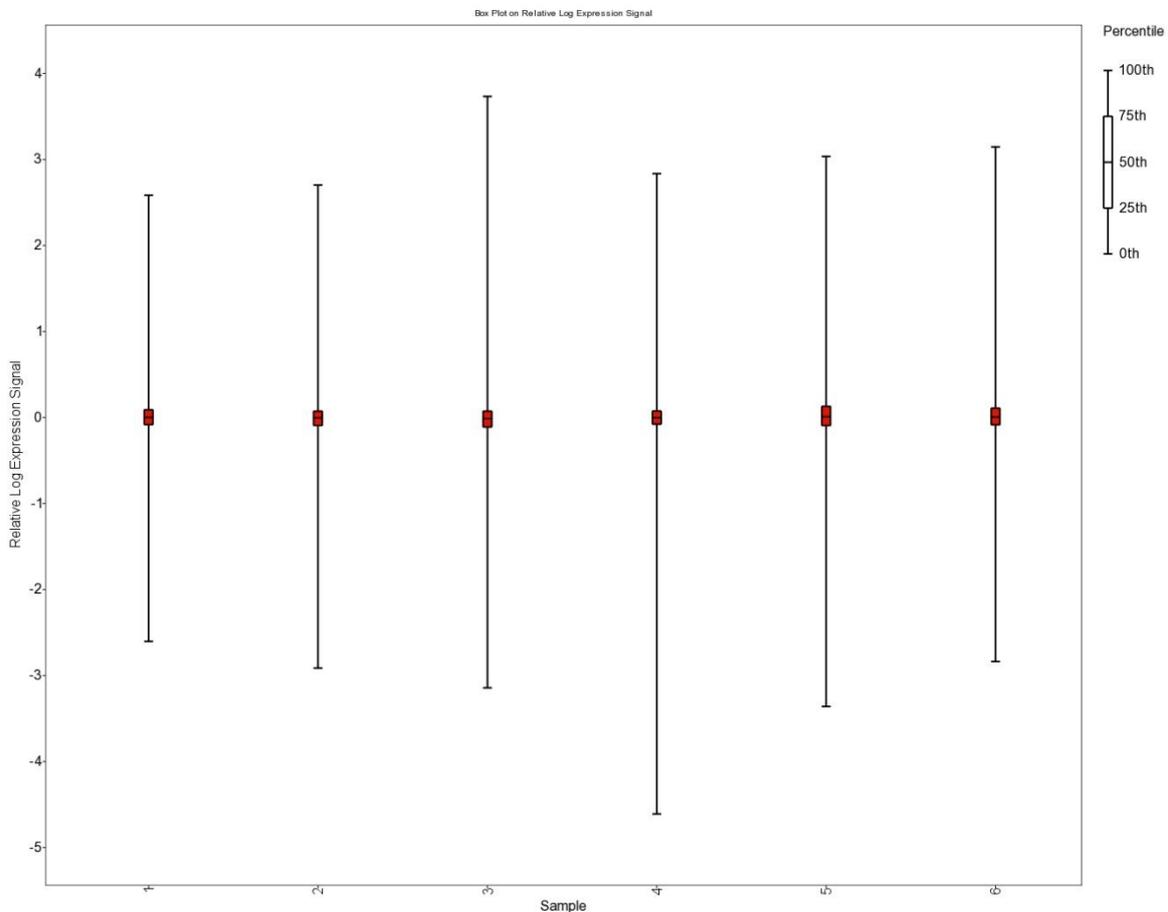
En la misma figura se observan valores muy bajo para el parámetro “*mad residual mean*” para todos los microarray, así como para la categoría que agrupa a todas las sondas del chip (“*all probeset*”), como para la categoría que agrupa a las sondas diseñadas para evaluar regiones exónicas de genes constitutivos (“*positive controls*”). El valor medio de los seis microarrays analizados en el presente trabajo de todas las sondas y de las regiones

exónicas de genes constitutivos es  $0,25 \pm 0,01$  y  $0,09 \pm 0,01$ , respectivamente. Por tanto, hay un buen ajuste al modelo de sonda robusto para las sondas contenidas en la respectiva categoría.

### -Control de sesgos y comparabilidad de las muestras

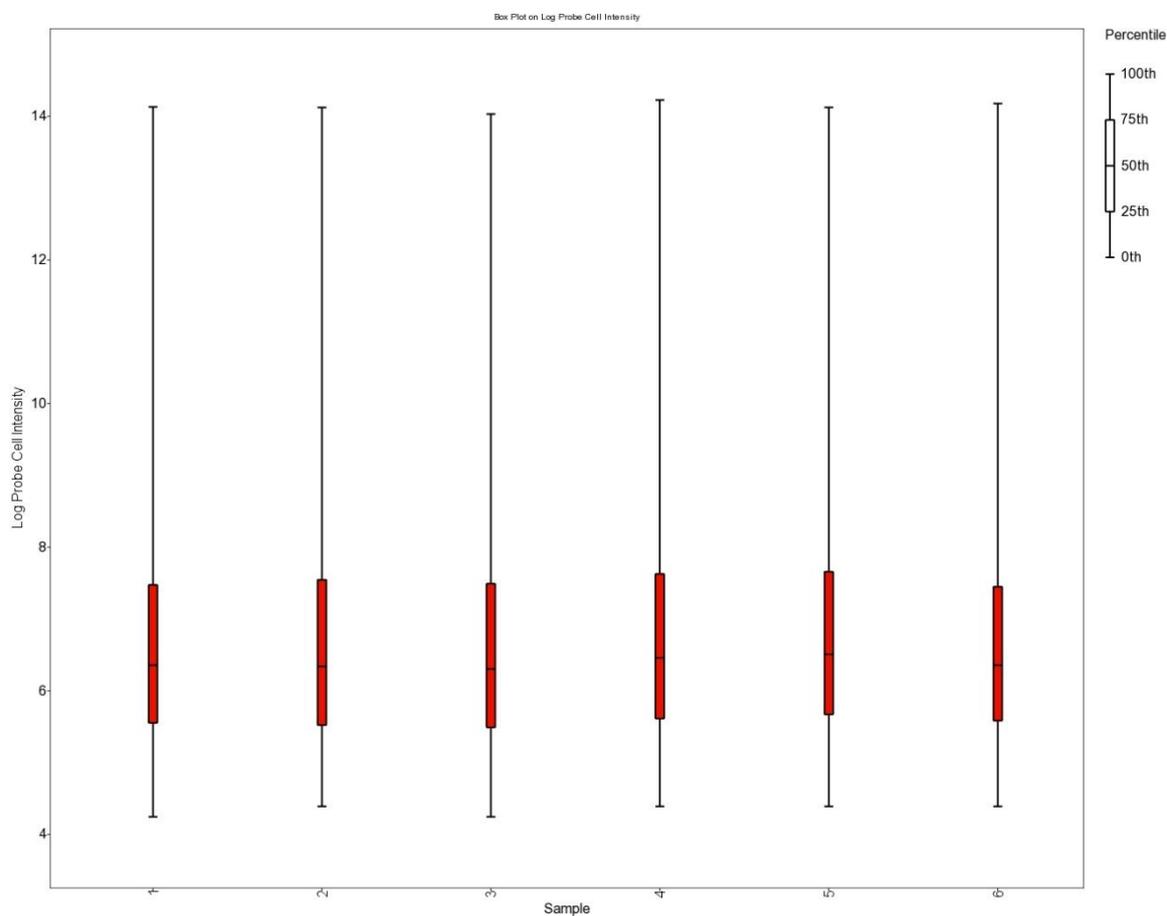
Para realizar un control de sesgos y comprobar la comparabilidad entre muestras se genera un diagrama de cajas con patillas de los valores de expresión logarítmica relativa. Este gráfico permite identificar diferencias en la distribución de los datos de intensidad entre microarrays. Por lo general, se espera que las cajas sean similares.

Los valores de expresión logarítmica relativa (RLE acrónimo del inglés *Relative Log Expression*) son obtenidos calculando para cada set de sondas el ratio entre la expresión de un set de sondas y la mediana de la expresión de ese set a lo largo de todos los microarrays del experimento. Estos ratios deberían ser cercanos a cero en escala logarítmica. Por tanto, los diagramas de cajas que representan la distribución de ratios deberían estar centrados alrededor de cero y tener similar distribución. Cualquier otro comportamiento sería señal de un microarray de baja calidad. Como se observa en la figura 27, todas las distribuciones para los diferentes biochips se encuentran centradas en cero, con lo cual se verifica que las muestras no reflejan un sesgo en los valores de expresión obtenidos y que, por tanto, todas son comparables entre sí. El valor medio de RLE para las dos categorías comentadas anteriormente (all probeset” y “positive controls”) presentan valores cercanos a cero ( $0,15 \pm 0,02$  y  $0,19 \pm 0,04$ , respectivamente), mostrando la alta calidad de los datos de expresión obtenidos.

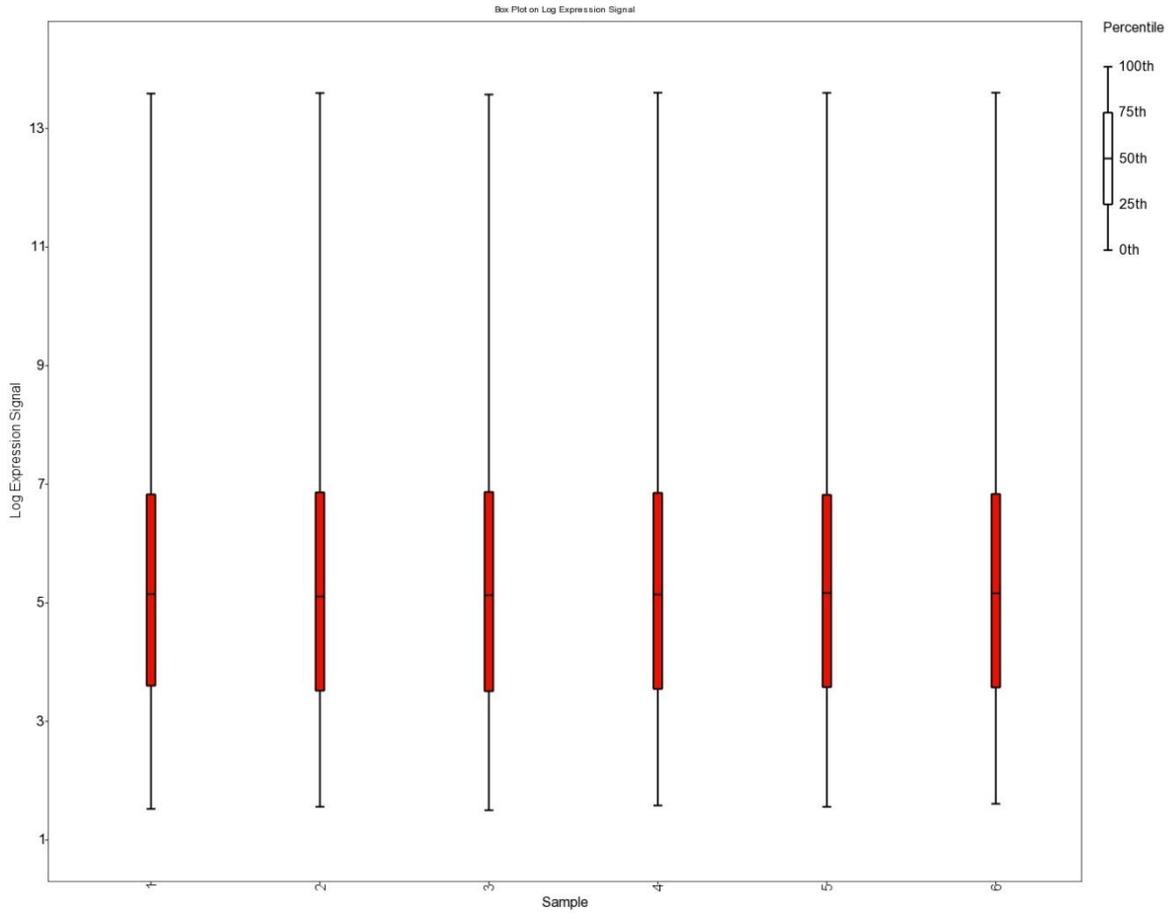


**Figura 27:** Diagrama de cajas de los valores de expresión logarítmica relativa.

Además, realizar los diagramas de la distribución de las intensidades logarítmicas de intensidad antes (Figura 28) y después (Figura 29) de la normalización permite verificar el paso de normalización y comprobar que, tras dicho paso, las distribuciones son más comparables aún, incluso idénticas.



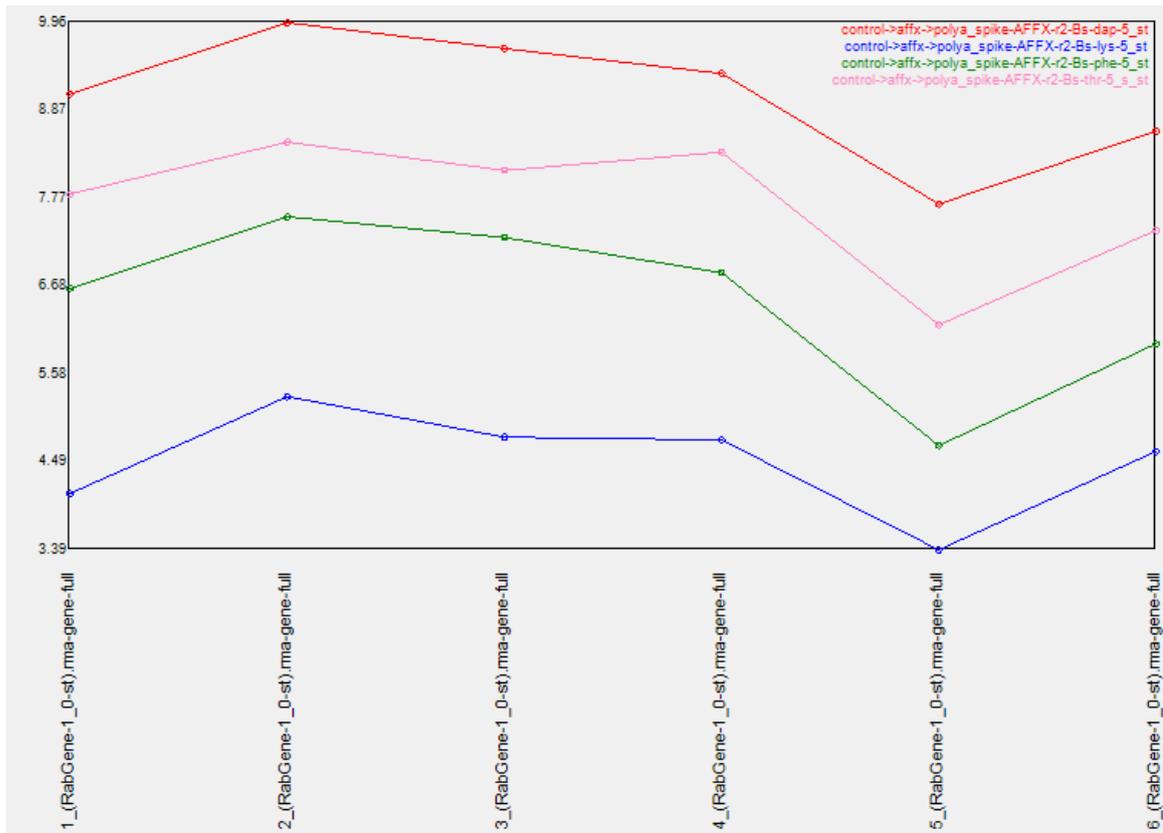
**Figura 28:** Diagrama de cajas de los valores de expresión logarítmicas antes de la normalización.



**Figura 29:** Diagrama de cajas de las señales de expresión logarítmicas normalizadas.

### -Control de calidad del marcaje de las muestras

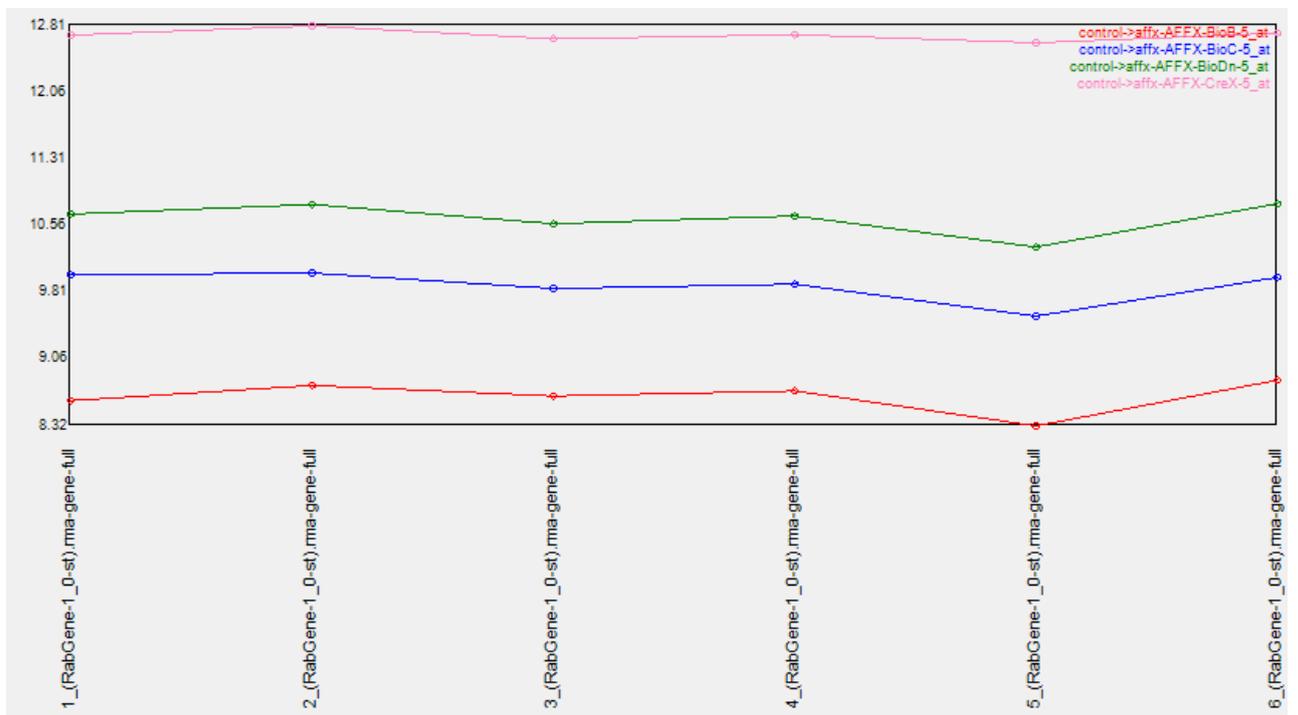
Las concentraciones finales de estos controles de marcaje (dap, thr, phe y Lys), también denominados controles poli-A relativos al total de RNA, son 1:100.000, 1:50.000, 1:25.000, 1:7500, respectivamente. La figura 30 muestra que el patrón de intensidad se cumple perfectamente para todas las muestras analizadas en el presente trabajo dado que  $Lys < Phe < Thr < Dap$  por lo que no se ha producido sesgo durante la reacción de retro-transcripción entre los genes altamente expresados, y los genes expresados en menor medida.



**Figura 30:** Media de los valores de intensidad de los sets de sondas pertenecientes a la categoría de controles de marcaje obtenidos en cada chip.

### -Control de calidad de la hibridación de las muestras

En la figura 31 se observa el patrón de intensidades 4 controles de hibridación (BioB, BioC, BioD y Cre) añadidos en el cóctel de hibridación a diferentes concentraciones. Dado que en todas las muestras analizadas en el presente trabajo se obtiene que  $\text{BioB} < \text{BioC} < \text{BioD} < \text{CreX}$ . Se concluye que la hibridación de las muestras ha sido correcta. Patrones diferentes indicarían una posible incorrecta hibridación.



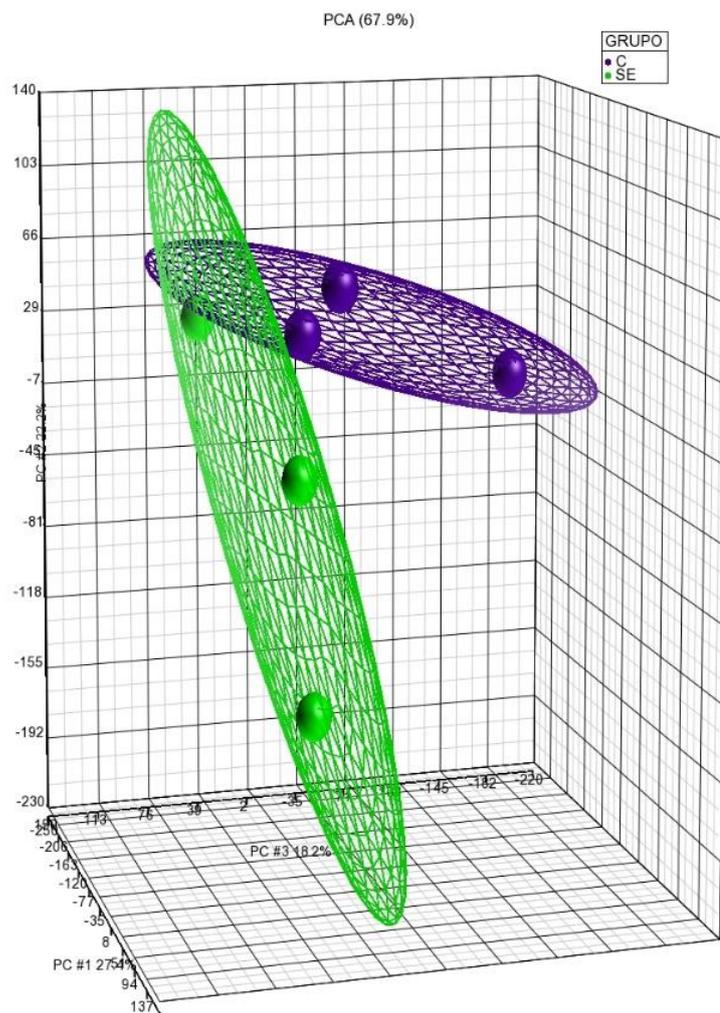
**Figura 31:** Media de los valores de intensidad de los sets de sondas pertenecientes a la categoría de controles hibridación obtenidos en cada chip.

En resumen, los resultados obtenidos de los seis microarrays pertenecientes a conejas de un experimento de selección por tasa de ovulación durante 10 generaciones (tres conejas de la población base y tres conejas de la generación 10) pueden utilizarse para realizar el análisis transcripcional de tejido ovárico.

## 4. Análisis transcripcional

### 4.1. Análisis de componentes principales (ACP)

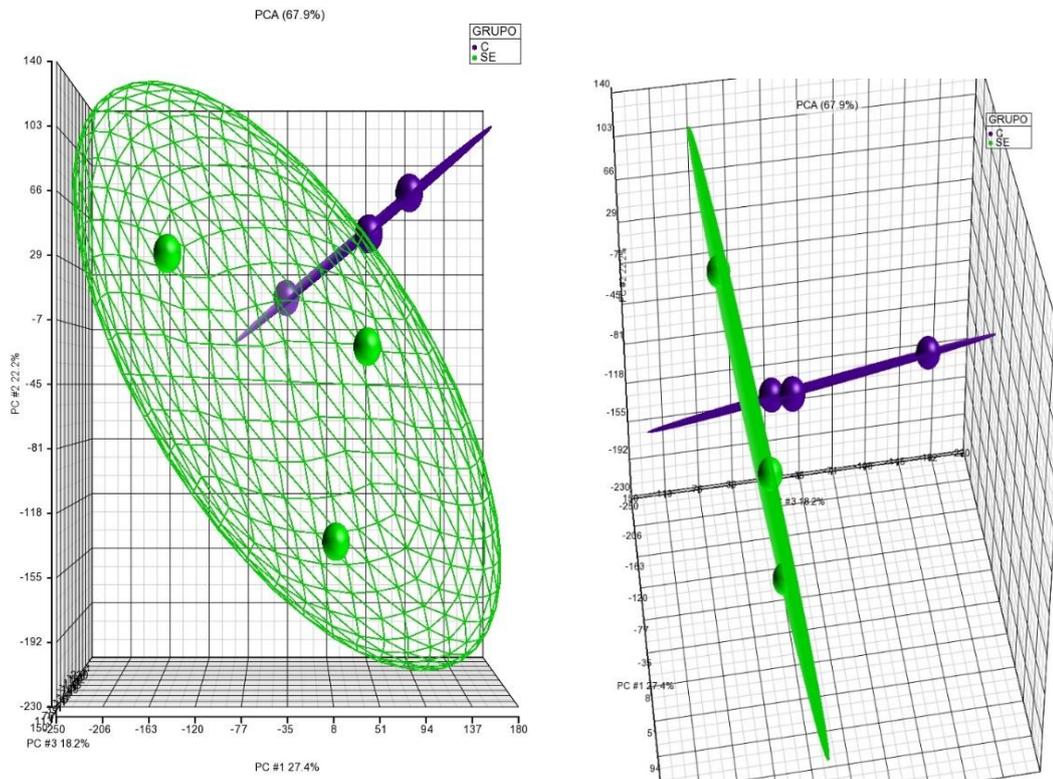
El ACP no supervisado permite observar la distribución en el espacio de las muestras de tejido ovárico teniendo en cuenta las tres componentes principales, que explican el 67,9 % de la variabilidad de las muestras. En el análisis ACP de las seis muestras analizadas (tres muestras pertenecientes a las hembras de la población control (C) y tres muestras pertenecientes a las hembras de la población seleccionada (SE) se observa que las muestras se diferencian desde el punto de vista transcriptómico (Figura 32).



**Figura 32:** Análisis de componentes principales de las seis muestras analizadas. Muestras de ovario pertenecientes a tres hembras de la población control representadas en esferas de color morado, y a tres hembras de la población seleccionada en esferas color verde.

En la figura 33 se observan otras imágenes del mismo ACP, pero en otro plano tridimensional.

Así, se concluye que existen diferencias en el transcriptoma entre los dos grupos estudiados.



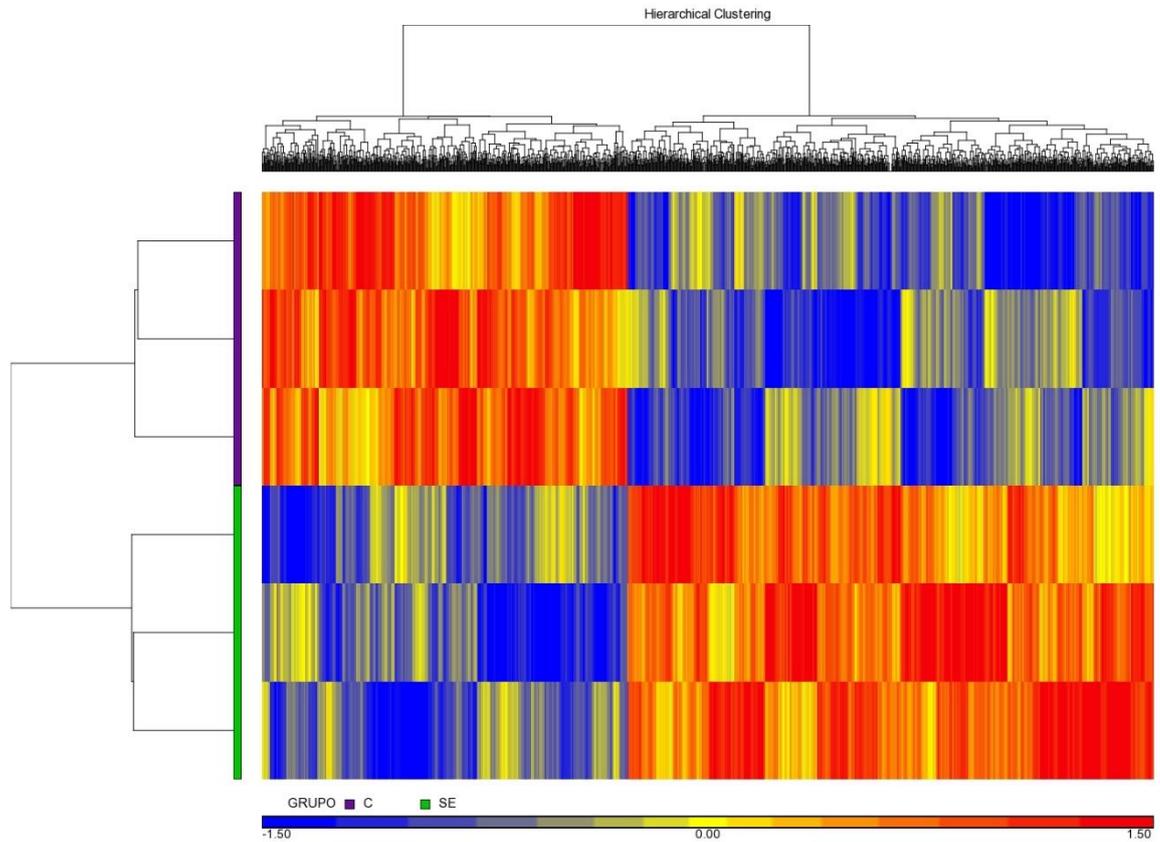
**Figura 33:** Planos tridimensionales del análisis de componentes principales de las seis muestras analizadas. Muestras de ovario pertenecientes a tres hembras de la población control representadas en esferas de color morado y a tres hembras de la población seleccionada en esferas de color verde.

## 4.2. Análisis diferencial entre los 2 grupos estudiados

Posteriormente, se realizó el análisis de la varianza (ANOVA) de la expresión de los genes que se expresaban de manera diferencial entre la población seleccionada y el control mediante el uso del software Partek Genomics Suite 6.6.

El número de genes diferencialmente expresados obtenidos ha sido de 1.357. La representación de estos genes se ha realizado mediante mapas de calor (*heatmap* en inglés) en el que se representa de una manera gráfica los genes con diferente expresión ( $p$  valores  $\leq 0,05$ ) para cada una de las muestras estudiadas. Esto nos permite estudiar si las muestras, intragrupo y/o comparadas con el otro grupo, se parecen o se diferencian específicamente para esos genes filtrados mediante su significancia (Figura 34).

En el análisis de las muestras realizado, se observa que las muestras para cada grupo responden de manera diferente en cuanto a la expresión de esos 1.357 genes, y que no hay cambios dentro del mismo grupo, y, por tanto, no se detectan muestras anómalas (*outliers en inglés*).



**Figura 34:** *Heatmap* de los 1.357 genes diferencialmente expresados entre las tres hembras de la población control y las tres hembras de la población seleccionada. Cada columna representa un gen y cada línea una muestra. Los genes sobre expresados se representan en rojo y los infra expresados en azul. La barra morada representa a las muestras control y la verde a las muestras de las hembras seleccionadas por tasa de ovulación.

De estos 1.357 genes, 800 genes tienen unos valores de cambio de expresión (*fold change*) positivos en la población seleccionada, de los cuales 59 eran superiores a un valor de *fold change* de  $|1,5|$ , y 557 genes infraexpresados, de los cuales 44 eran menores a  $|1,5|$ .

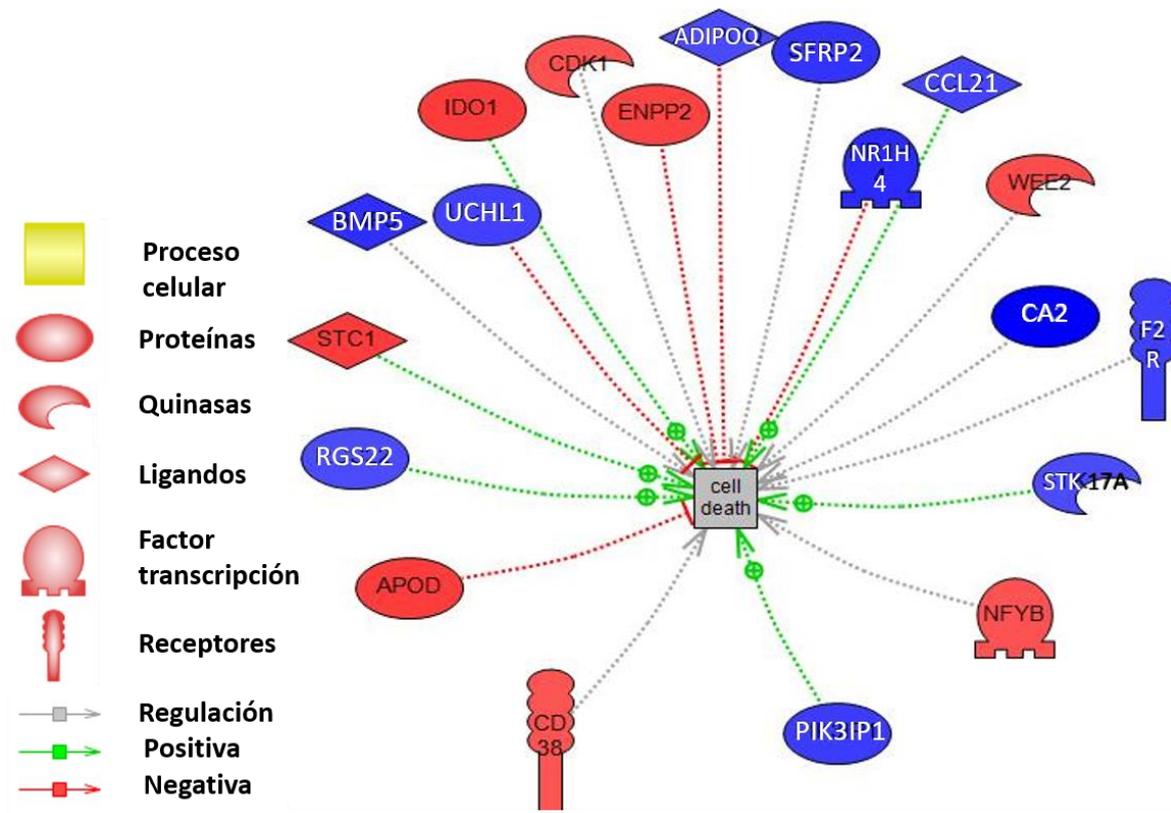
Para generar el listado definitivo de nuestros genes, se han despreciado aquellos genes que no tienen una anotación funcional hasta el momento y el resultado son 922 genes anotados, es decir, casi un 68 %. De los 922 genes, 529 estaban sobreexpresados y 393 genes estaban infraexpresados. De estos, 14 genes presentan valores superiores a un *fold change* de 1,5 y 19 inferiores a un este cambio de expresión (tabla 5).

**Tabla 5:** Genes cuya expresión entre las hembras de la población seleccionada y control presentan un p-valor <0,05 y *fold change* | 1,5| .

<b>Gen</b>	<b>Asignación de genes</b>	<b>P-valor</b>	<b>Fold change</b>
C16H1orf131	<i>chromosome 16 open reading frame, human C1orf131</i>	1,39E-02	2,94
STC1	<i>stanniocalcin 1</i>	1,11E-02	1,90
IDO1	<i>indoleamine 2,3-dioxygenase 1</i>	8,58E-03	1,89
APOD	<i>apolipoprotein D</i>	4,76E-02	1,88
ENPP2	<i>ectonucleotide pyrophosphatase/phosphodiesterase 2</i>	2,76E-02	1,78
LOC100356950	<i>placenta-specific protein 1-like</i>	3,88E-02	1,66
WEE2	<i>WEE1 homolog 2 (S. pombe)</i>	4,26E-02	1,62
LOC100337961	<i>sulfotransferase 1C4-like</i>	2,34E-02	1,59
CDK1	<i>cyclin-dependent kinase 1</i>	4,09E-02	1,55
CBFB	<i>core-binding factor, beta subunit</i>	4,20E-02	1,53
ZNF367	<i>zinc finger protein 367</i>	2,48E-02	1,53
CD38	<i>CD38 molecule</i>	3,75E-02	1,52
LOC100339957	<i>uncharacterized LOC100339957</i>	9,30E-03	1,52
GNA14	<i>guanine nucleotide binding protein (G protein), alpha 14</i>	4,37E-03	1,50
SLC51A	<i>solute carrier family 51, alpha subunit</i>	2,69E-02	-1,50
CNTN3	<i>contactin 3 (plasmacytoma associated)</i>	2,93E-02	-1,51
SYT12	<i>synaptotagmin XII</i>	2,61E-02	-1,52
RGS22	<i>regulator of G-protein signaling 22</i>	2,69E-02	-1,53
STK17A	<i>serine/threonine kinase 17a</i>	1,77E-02	-1,54
NR6A1	<i>nuclear receptor subfamily 6, group A, member 1</i>	2,05E-02	-1,54
ADIPOQ	<i>adiponectin, C1Q and collagen domain containing</i>	1,77E-02	-1,57
CCL21	<i>chemokine (C-C motif) ligand 21</i>	3,56E-02	-1,64
ACSL4	<i>acyl-CoA synthetase long-chain family member 4</i>	2,98E-02	-1,66
UCHL1	<i>ubiquitin carboxyl-terminal esterase L1</i>	2,84E-02	-1,69
F2R	<i>coagulation factor II (thrombin) receptor</i>	1,80E-02	-1,70
PIK3IP1	<i>phosphoinositide-3-kinase interacting protein 1</i>	4,65E-02	-1,77
FABP2	<i>fatty acid binding protein 2, intestinal</i>	1,33E-02	-1,83
SFRP2	<i>secreted frizzled-related protein 2</i>	2,20E-02	-1,88
BMP5	<i>bone morphogenetic protein 5</i>	3,72E-02	-1,97

<b>Tabla 5 (sigue):</b> Genes cuya expresión entre las hembras de la población seleccionada y control presentan un p-valor <0,05 y <i>fold change</i>   1,5 .			
NR1H4	<i>nuclear receptor subfamily 1, group H, member 4</i>	1,31E-02	-2,03
PLA2G2D	<i>phospholipase A2, group IID</i>	2,76E-03	-2,29
PTGFR	<i>prostaglandin F receptor (FP)</i>	1,51E-02	-2,65
CA2	<i>carbonic anhydrase II</i>	2,81E-02	-2,75

De los 33 genes expresados diferencialmente y filtrados por su cambio de expresión se encuentran 19 genes implicados directamente con la muerte celular: STC1, IDO1, CDK1, CD38, APOD, ENPP2, WEE2, NFYB, CCL21, SFRP2, ADIPOQ, UCHL1, BMP5, NR1H4, RGS22, F2R, STK17A, PIK3IP1, CA2 (Figura 35). Todos estos genes representan casi un 58 % del total de genes diferenciales filtrados.



**Figura 35:** Representación de 19 genes de los 1.357 genes diferencialmente expresados entre la población seleccionada y control estudiados que se relacionan con muerte celular (*cell death* en inglés). En azul se representan los genes infraexpresados y en rojo los sobreexpresados. La intensidad del color es directamente proporcional a la intensidad del cambio de expresión.

Concretamente, STC1 (en inglés *Stanniocalcin 1*) es una glicoproteína implicada en la homeostasis mineral. Su homólogo de mamífero parece tener un papel en la regulación de multitud de procesos biológicos, incluyendo la angiogénesis y la esteroidogénesis, ambos eventos fundamentales en la función ovárica (Gerritsen ME, 2005). La expresión del gen STC1 sería la primera vez que se detecta en tejido ovárico de conejos. Estudios en porcino, muestran que STC1 tiene un papel fisiológico y modulador en la esteroidogénesis en el folículo ovárico. Además, STC1 tiene un efecto negativo sobre la producción de estradiol 17B y un efecto positivo en la producción de progesterona sobre las células granulosas (Baioni L, 2009). Por tanto, STC1 parece tener una función importante en la maduración del folículo.

En otros estudios, en porcino también, se demuestra que STC1 es producido en las células de la granulosa, tanto en condiciones basales como en respuesta a la falta de oxígeno. De ahí, que se le pueda atribuir, un papel modulador en condiciones hipoxias impulsada por angiogénesis (Basini G, 2010; Meidan R, 2013). Recientemente, experimentos en humano atribuyen a la STC1 un potente regulador de la actividad proteolítica de la proteína A plasmática asociada al embarazo durante el desarrollo del folículo ovárico humano (Jepsen MR, 2016).

Además, se ha observado que STC1 ejerce una regulación positiva para que se de mortalidad celular (Figura 35).

Un estudio de Basini y sus colaboradores (Basini G, 2009) en células de la granulosa en porcino, observaron que STC1 tenía un papel muy importante inhibiendo la angiogénesis, proceso tan importante y necesario para la foliculogénesis.

Las causas de la mortalidad prenatal es un tema de gran interés no solo en especies de interés productivo como el conejo sino también en reproducción asistida en humanos. Los resultados de este trabajo han puesto de relieve un conjunto de genes que podrían estar implicados en el aumento de la mortalidad prenatal observada en la línea seleccionada por tasa de ovulación y que pueden ayudar a conocer mejor los mecanismos de la ovulación y su relación con el aumento de la mortalidad prenatal en el conejo y otras especies.



## V. Conclusiones

---

- Las hembras de la población seleccionada por tasa de ovulación durante diez generaciones presentan un perfil transcripcional diferente a las hembras de la población control mediante el uso de microarrays de expresión génica de Affymetrix.
- Se han detectado 1.357 genes significativos expresados diferencialmente entre la línea seleccionada y la control. De los cuales, 33 genes presentan un valor de expresión superior a  $|1,5|$ .
- Entre los genes con diferente expresión se han caracterizado 19 genes implicados directamente con muerte celular.
- El gen *Stanniocalcin 1* (STC 1) es el más sobre expresado. Este gen es la primera vez que se ha detectado en tejido ovárico en conejos.



## VI. Referencias

---

- Adams, CE & Chang, MC. (1962). Capacitation of rabbit spermatozoa in the fallopian tube and in the uterus. *Journal of Experimental Zoology* 5, 151-159.
- Affymetrix. (2005). *GeneChip Microarrays: Student Manual*. Retrieved 11 1, 2012, from [http://www.affymetrix.com/about\\_affymetrix/outreach/educator/microarray\\_curricula.affx](http://www.affymetrix.com/about_affymetrix/outreach/educator/microarray_curricula.affx)
- Affymetrix. (2009-2010). *GeneChip WT Terminal Labeling and Hybridization User Manual*. Retrieved 11 2012, from [www.Affymetrix.com](http://www.affymetrix.com)
- Al-shorepy, SA., Clutter, AC., Blair, RM., Nielsen, MK. (1992). Effects of three methods of selection for litter size in mice on pre-implantation embryonic development. *Biology of Reproduction*,46, 958-963.
- Baioni, L., Basini, S., Bussolati, F., Grasselli, F. (2009). Stanniocalcin 1 is a potential physiological modulator of steroidogenesis in the swine ovarian follicle. *Veterinary Research Communications*,33 suppl 1 ,73-76.
- Baselga, M., Blasco, A. (1989). Mejora genética del conejo de producción de carne. *Mundi-Prensa*,110.
- Basini, G., Bussolati, S., Santini, SE., Grasselli, F. (2009). Stanniocalcin, a potential ovarian angiogenesis regulator, does not affect endothelial cell apoptosis. *Annals of the New York Academy of Sciences*,1171,94-99.
- Basini,G,Baioni,L,Bussolati,S,Grolli,S,Kramer,LH,Wagner,GF,Grasselli,F. (2010). Expression and localization of stanniocalcin 1 in swine ovary. *General and Comparative Endocrinology*,166(2),404-408.
- Bazer, FW., Terqui, M., Martinat-Botte, F., (1990). Physiological limits to reproduction. *Proceedings of the 4th World Congress on Genetics Applied to Livestock Production*, Edinburgh (UK) 16, 292-298.
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing. *Royal Statistical Society*, 289-300.
- Blasco, A., Bidanel, JP., Bolet, G., Haley, CS., Santacreu, MA. (1993). The genetics of prenatal survival of pigs and rabbits: a review. *Livestock Production Science*, 37, 1-21.
- Cartuche, L., Pascual, M., Gómez, EA., Blasco, A. (2014) . Economic Weights in Rabbit Meat Production. *World Rabbit Science*,22, 165-177.

- Chomczynski,P & Sacchi,N. (1987). Single-Step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Analytical Biochemistry*, 162,156-159.
- Crick, F. (1970). Central Dogma of Molecular Biology. *Nature*, 227, 561-563.
- Fodor, SP., Read, JL., Pirrung, MC., Stryer, L., Lu, AT., Solas, D. (1991). Light-directed, spatially addressable parallel chemical synthesis. *Science*, 251(4995),767-773.
- García, ML., Baselga, M., Viudes de castro, MP., Vicente, JS. (2000).Reconstitución de una línea de conejos a partir de embriones crioconservados. *Archivos de Zootecnia*,49, 81-86.
- García, ML & Baselga, M. (2002). Estimation of genetic response to selection in litter size of rabbits using a cryopreserved control population. *Livestock Production Science*,74,45-53.
- Geisert, RD., & Schmitt, RAM. (2002). Early embryonic survival in the pig: Can it be improved? *Journal of Animal Science.*, 80(E-Suppl.), E54-E65.
- Gerritsen, ME & Wanner, GF. (2005). Stanniocalcini:no longer just a fish tale. *Vitamins and Hormones*, 70,105-135.
- Jepsen, MR., Kloverpris, S., Botkjaar, JA., Wissing, Ml., Andersen, CY., Oxvig, C. (2016). The proteolytic activity of pregnancy associated plasma protein-A is potentially regulated by stanniocalcin-1 and -2 during human ovarian follicle development. *Human Reproduction*,31(4),866-874.
- Koenig, JLF., Zimmerman, DR., Eldridge, FE., Kopf, JD. (1986). The effect of superovulation and selection for high ovulation rate on chromosomal abnormalities in swine ova. *Journal of Animal Science* ,(Suppl 1) 63,202.
- Krisher, RL. (2004). The effect of oocyte quality on development. *Journal of Animal Science*, 82 (E. Suppl.), E14-E23.
- Laborda, P., Mocé, ML., Santacreu, MA., Blasco, A. (2011). Selection for ovulation rate in rabbits: Genetic parameters, direct response, and correlated response on litter size. *Journal of Animal Science*, 89,2981-2987.
- Laborda, P., Mocé, ML., Blasco, A., Santacreu, MA.(2012a) Selection for ovulation rate in rabbits: Genetic parameters and correlated responses on survival rates. *Journal of Animal Science*, 90,439-446.
- Laborda, P, Santacreu, MA,Blasco,A, Mocé, ML.(2012b). Selection for ovulation rate in rabbits: Direct and correlated responses estimated with a cryopreserved control population. *Journal of Animal Science*, 90,3392-3397.
- Leiser, R & Kaufmann,P. (1994). Placental structure: in a comparative aspect. *Experimental and Clinical Endocrinology*, 102,122-134.

- Mueller, O., Lightfoot, S., & Schroeder, A. (2004). RNA Integrity Number (RIN). Standardization of RNA Quality Control. *Agilent Application Note*, 1-8.
- Pope, WF. (1988). Uterine Asynchrony: A cause of embryonic loss. *Biology of Reproduction*, 39,999-1003.
- Sánchez, A., & Ruíz de Villa, M. C. (2008, Julio 7). *A Tutorial Review of Microarray Data Analysis*. Retrieved 12 20, 2012, from Universidad de Barcelona: <http://www.ub.edu/stat/docencia/bioinformatica/microarrays/ADM/>
- Santacreu, MA. (2006). La supervivencia prenatal en la coneja reproductora. XXXI *Symposium de cunicultura de ASESCU*, Lorca, Murcia, Spain. Vol. I, 229-236.
- Torres, S., Hulot F., Meunier, M. (1984). Étude comparée du développement et de la mortalité embryonnaire chez deux genotypes de lapines. *3rd World Rabbit Congress*. Rome-Italy, Vol. 2, 417-425.
- Torres, S., Hulot F., Meunier, M., Sevellec, C. (1987). Comparative study of preimplantation development and embryonic loss in two rabbit strains. *Reproduction Nutrition Development*. 27(3), 707-714
- Venter, JC., Adams, MD., Myers, EW., Li, PW., Mural, RJ., Sutton, GG., Smith, HO. (2001). The Sequence of the Human Genome. *Science Signalling*, 291(5507), 1304-1351.
- Wilde, M., Xie, HS., Day, ML., Pope, WF. (1988). Survival of small and large littermate blastocysts in swine after synchronous and asynchronous transfer procedures. *Theriogenology*, 30, 1069-1074.
- Wilmut, I., Sales, DI., Ashworth, CJ. (1986). Maternal and embryonic factors associated with prenatal loss in mammals. *Journal of Reproduction and Fertility*, 76, 851-864.
- Wintenberger-torres, S. (1974). Relation entre la taille des blastocystes de lapine a l'implantation et la survie embryonnaire. *Ann. Biol. Anim. Bioch. Biophys.*, 14(1), 41-52.
- Xie, S., Broermann, DM., Nephew, KP., Geisert, RD., Pope, WF. (1990). Ovulation and early embryogenesis in swine. *Biology of Reproduction*, 43, 236-2

