



UNIVERSIDAD DE JAÉN  
*Facultad de Ciencias Experimentales*

Trabajo Fin de Grado

# **Análisis de datos de expresión génica en autismo mediante teoría de juegos cooperativos**

**Alumna: Eva Vargas Liébanas**

**Junio, 2014**



UNIVERSIDAD DE JAÉN

FACULTAD DE CIENCIAS EXPERIMENTALES

GRADO EN BIOLOGÍA

Trabajo Fin de Grado

# Análisis de datos de expresión génica en autismo mediante teoría de juegos cooperativos

Eva Vargas Liébanas

Jaén y Junio, 2014

ÍNDICE

ABSTRACT .....	3
1. INTRODUCCIÓN .....	4
2. MATERIAL Y MÉTODOS .....	12
2.1. Datos .....	12
2.2. Análisis de datos de microarrays por métodos convencionales .....	13
2.3. Análisis de datos de microarrays por Teoría de Juegos Cooperativos .....	14
2.4. Análisis funcional .....	16
3. RESULTADOS.....	17
3.1. Análisis convencional .....	17
3.2. Análisis por Teoría de Juegos .....	17
3.2.1. <i>Autistas frente a controles hijos de padres jóvenes.....</i>	<i>17</i>
3.2.2. <i>Autistas frente a controles hijos de padres mayores.....</i>	<i>18</i>
3.2.3. <i>Controles hijos de padres jóvenes frente a controles hijos de padres mayores.....</i>	<i>19</i>
3.3. Análisis funcional .....	23
4. DISCUSIÓN .....	25
5. BIBLIOGRAFÍA.....	28

## **ABSTRACT**

Microarray-based gene expression technology, which allows simultaneous monitoring of the expression of thousand of genes, is the most widely used platform for transcriptome analysis and it is an useful tool to detect variations in the human genome in health and disease. The big amount of data generated by this approach require an adequate computational analysis, always crucial for providing conclusive results. There are conventional and well established microarray bioinformatical techniques, but Game Theory (GT) has recently

arisen as a powerful analytical method of application in this field. This work aims to evaluate the effectiveness of GT in microarray gene expression data analysis when compared with conventional approaches and its application in autism, a disease in which the complex molecular mechanisms have not yet been clarified. In this work, the obtained results showed that GT improves the detection of significant differentially expressed genes in autism (according to previous studies). Thus, Game Theory may be proposed as a useful approach to be applied in microarray gene expression analysis in health and disease.

## 1. INTRODUCCIÓN

La expresión génica es el proceso mediante el cual se lleva a cabo la producción por parte de una célula de un producto génico a partir de la secuencia de nucleótidos existente en el ADN. Se trata de un proceso necesario para la vida que y que cuenta con mecanismos complejos para su regulación (Watson et al., 2014).

El estudio y análisis de la expresión génica supuso una verdadera revolución en el campo de la ciencia a finales del siglo XX y comienzos del XXI, fundamentalmente a partir de la aparición de proyectos como el proyecto Genoma Humano (años 90) o el Proyecto NCODE (2003) (Ding et al., 2014). El conocimiento de la secuencia de nucleótidos presente en nuestro genoma permitió el diseño de técnicas basadas en la hibridación, entre las cuales una de las más destacables es la reacción en cadena de la polimerasa (PCR), desarrollada en 1986 por Kary Mullis. A partir del desarrollo de la PCR surgieron otras múltiples herramientas y técnicas que permiten analizar la expresión génica como la RT-PCR o qPCR, sin dejar de lado técnicas clásicas basadas en la hibridación como el *northern blot* o la hibridación *in situ* fluorescente (FISH) (Dale y Schantz, 2002). Llegados a este punto y con la finalidad de poder evaluar a gran escala la expresión génica surgió la tecnología de microarrays.

Un microarray es un chip de pequeño tamaño que permite el análisis simultáneo de todos los genes del genoma con una gran precisión, en muy poco tiempo y en un solo experimento (Lee et al., 2007). Se trata de una herramienta de amplio uso en biomedicina basada en la hibridación de sondas específicas de ADN, pues un único microarray puede contener de cientos a miles de secuencias cortas de ADN hibridadas sobre una superficie sólida. Así pues, y en comparación con otras técnicas moleculares reduccionistas, un experimento con microarrays implica el análisis simultáneo de varios cientos o miles de genes, convirtiéndolos en una herramienta clave de la llamada era genómica (Rogers et al., 2007).

La aplicación de la tecnología de microarrays ha demostrado que es una herramienta útil para el diagnóstico de enfermedades en las que se ven afectados los niveles de expresión génica (Quackenbush et al., 2006). A pesar de presentar multitud de ventajas desde el punto de vista experimental y de la aplicación biomédica, se trata de un sistema de análisis que genera un gran volumen de información que debe ser procesada y analizada cuidadosamente mediante potentes aplicaciones bioinformáticas, lo cual dificulta la obtención e interpretación de los resultados. Para una buena comprensión de los distintos conceptos que se mencionarán a lo largo de este Trabajo Fin de Grado, a continuación comentamos brevemente una serie de términos relacionados con la tecnología de microarrays.

En primer lugar, existen diferentes tipos de arrays, que varían en función de las necesidades del experimento: los más destacables son los arrays de oligonucleótidos y los arrays de cDNA. En nuestro caso, los arrays analizados pertenecen al primer grupo. En ellos, la sonda empleada para la hibridación es una secuencia corta de unos 25 nucleótidos (si bien esta longitud puede variar en función de la casa comercial responsable de la fabricación). El diseño experimental de los arrays de oligonucleótidos implica la existencia de sondas que aparean completamente con la secuencia problema presente en la muestra (*perfect match*) y sondas que aparean en todas las posiciones excepto una (*mismatch*). Esta combinación de sondas es empleada a modo de control para asegurar en todo momento la fiabilidad del proceso y la calidad de las sondas empleadas.

Por su parte, los arrays de cDNA utilizan sondas de ADN complementario (cDNA) obtenidas de bibliotecas génicas (mediante clonación), así como fluoróforos de distintos colores para marcar cDNA obtenido a partir del ARN de las muestras de la situación control y la situación experimental con lo cual, al producirse la hibridación simultánea de ambos tipos de muestras, se obtendrá una fluorescencia determinada que dependerá del nivel de expresión de las mismas (Moreno y Solé, 2004).

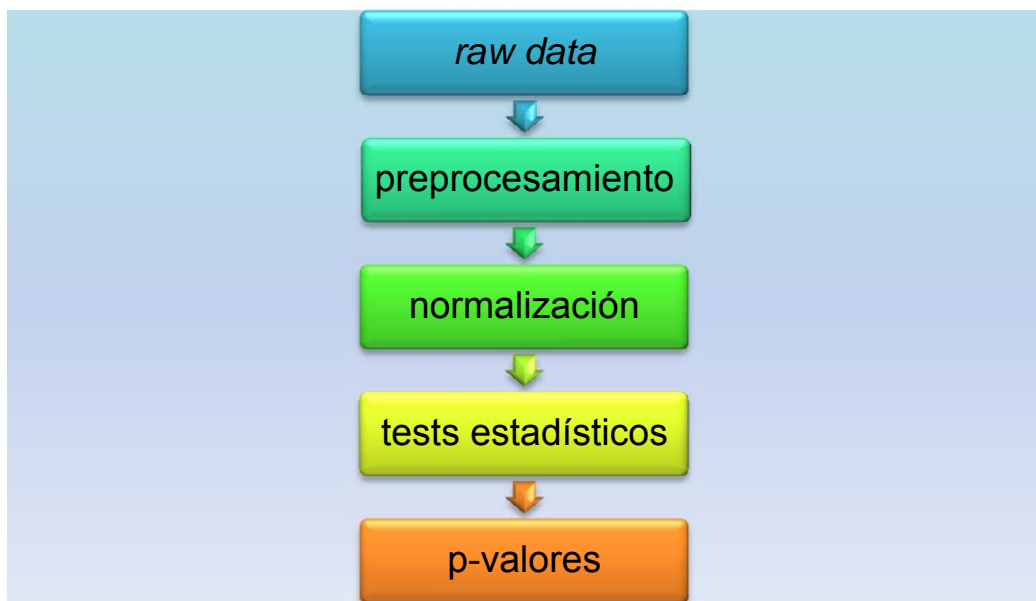
De igual forma, los arrays también pueden ser clasificados en función de los fluoróforos que se utilicen para marcar las sondas específicas necesarias. Así, podemos encontrar arrays que emplean un único fluoróforo (arrays de un color o *single - channel*), o bien aquellos que emplean dos fluoróforos diferentes (arrays de dos colores o *two - channel*) (Shalon et al., 1996). En el Trabajo Fin de Grado que nos ocupa se trabaja con arrays de oligonucleótidos y de un solo color (tipo *Affymetrix*).

Para la aplicación de las técnicas de análisis de microarrays objeto de este trabajo y que serán descritas en apartados posteriores, hemos escogido como modelo el análisis de la expresión génica en autismo. El autismo es un desorden multigénico complejo con un amplio rango de fenotipos (espectro autista), que suelen incluir deficiencias en el lenguaje o comportamientos estereotipados y repetitivos, entre otros (Buxbaum et al., 2012). La incidencia creciente del autismo en la última década ha incrementado el interés científico por su investigación (Pop-Jordanova et al., 2014). Estudios previos han demostrado la existencia de numerosos genes implicados en distintas funciones biológicas e interacciones moleculares en este trastorno neurológico (Wall et al., 2009).

Debido a su componente multigénico, parece lógico pensar que puede resultar útil analizar los niveles de expresión génica mediante la utilización de microarrays en individuos con autismo y compararlos con individuos sanos, con el fin de detectar las posibles diferencias existentes entre ambos estados. Es entonces cuando entra en juego la necesidad de llevar a cabo análisis estadísticos sobre el amplio volumen de datos generado en los experimentos con microarrays. Para poder interpretar de forma correcta la información

relativa a la expresión génica, deben realizarse cuidadosos análisis que proporcionen un significado biológico coherente.

En los arrays de oligonucleótidos, marcados con un solo color (tipo *Affymetrix*, como los utilizados en este trabajo) y siguiendo una estrategia de análisis convencional (Figura 1), el primer paso tras la obtención de los “datos iniciales generados tras el escaneo de las intensidades de color debidas a la hibridación diferencial de las sondas en el array” (*raw data*) es el preprocesamiento de los mismos. Existen diferentes metodologías asociadas al preprocesamiento de los datos, entre ellas la transformación logarítmica y la corrección del fondo (*background*). Este preprocesamiento de los datos facilita el manejo de los mismos en posteriores operaciones computacionales.



**Figura 1.** Esquema general en el análisis convencional de datos de expresión génica procedentes de microarrays.

La transformación logarítmica resulta de gran utilidad cuando nos encontramos ante datos entre los cuales existe una gran variación (como ocurre con los niveles de expresión génica cuando se analizan simultáneamente todos los genes del genoma). Gracias a ella se consigue que

los valores más bajos se expandan y los más altos se compriman, con lo cual se obtiene una mayor homogeneidad de los datos.

La corrección del fondo (*background*) trata de eliminar el llamado “ruido de fondo” y las posibles fluctuaciones en la señal general debido a cambios locales en determinadas posiciones del array. En este tipo de arrays es particularmente eficaz el empleo de los algoritmos RMA (*Robust Multi – array Average*) y MAS.5 (*Microarray analysis suite*) para esta etapa del preprocesamiento, algoritmos que también incluyen funciones de normalización, tal y como se comenta a continuación (Boldstad et al., 2003).

El siguiente paso a realizar es la normalización de los datos. Se trata de un paso crítico para la obtención de datos utilizables para análisis posteriores tales como la expresión diferencial de los genes o el agrupamiento de los datos (Do et al., 2006). Este procedimiento tiene el objetivo de eliminar todas aquellas causas de origen no biológico que puedan enmascarar los datos originales y que se deban a distintas variables no controladas que puedan influir sobre los resultados del experimento (fallos esporádicos en la hibridación, artefactos y variaciones aparentemente no apreciables en la preparación de las muestras o en la calibración del escáner de un día a otro).

Finalmente, el último paso de los análisis estadísticos supone la selección de aquellos genes cuya expresión se encuentra modificada con respecto a una situación control. Se han descrito diferentes métodos estadísticos que pueden ser aplicados al análisis de microarrays y que proporcionarían información acerca de aquellos genes que se muestran diferencialmente expresados (revisado en Draghici, 2003). Dichos métodos pueden estar basados, por ejemplo, en la selección de genes atendiendo a sus *p-valores*, obtenidos mediante la aplicación de tests paramétricos o no paramétricos en función de las necesidades del experimento. Yendo un paso más allá, el análisis se puede implementar llevando a cabo, en los tests de hipótesis múltiple, un procesamiento de los *p-valores* obtenidos (Dudoit et al., 2003). Debido a la gran variabilidad existente en la actualidad en las técnicas estadísticas aplicables a los análisis de microarrays, es necesario tener en



cuenta el hecho de que los resultados pueden variar en función de las aproximaciones empleadas.

Gracias al análisis de microarrays, en los últimos años se han propuesto multitud de genes candidatos que podrían estar implicados, como causa o consecuencia, en el autismo. De un modo muy resumido, podemos decir que la mayor parte de ellos intervienen en rutas metabólicas relacionadas con el desarrollo neurológico durante las etapas embrionarias, así como en mecanismos de transmisión del impulso nervioso, entre otros (Lambert et al., 2014; Lugo et al., 2014; Warrier et al., 2013).

Otro tipo de variaciones en la secuencia génica tales como las variaciones en el número de copias (*copy number variants* - CNVs) o variaciones de un único nucleótido (*single nucleotide variants* - SNVs) también han sido asociadas en los últimos años a la aparición del autismo (Jiang et al., 2014; Anitha et al., 2014; Correia et al., 2014). Incluso se ha llegado a demostrar una influencia de la herencia epigenética en la incidencia de la enfermedad (Wang et al., 2014; Mbadiwe y Millis, 2013).

Sin embargo, al tratarse de una enfermedad de base genética compleja (esto es, que no responde únicamente a causas genéticas sino que en su aparición cobran trascendencia otros factores como el ambiente), a día de hoy no han podido concluirse las causas exactas que acaban por determinar el desarrollo del espectro autista (Berko et al., 2014). Continuamente se están detectando nuevos genes candidatos a formar parte del complejo mecanismo molecular responsable del autismo.

A pesar de que estos métodos convencionales han supuesto un gran avance en lo que a caracterización de los factores genéticos responsables de enfermedades se refiere, en ocasiones los *p-valores* no siempre muestran la significación biológica que debieran, de modo que aparecen discrepancias que pueden afectar a la interpretación de los resultados (Shi et al., 2008).

Recordemos que se consideran estadísticamente significativos aquellos *p-valores* que se encuentren por debajo de una cifra determinada (generalmente 0,01 o 0,05). En ciertas ocasiones se da la circunstancia de que

un *p*-valor puede no llevar asociada inicialmente una significación biológica, cuando la realidad es que el gen que lleva asociado ese *p*-valor puede estar desempeñando un papel importante en algún proceso biológico. Así, encontramos *p*-valores bajos que no tienen relación alguna con la condición objeto de estudio y, en contraposición, *p*-valores altos (y que por este motivo se descartan en los análisis) que juegan importantes papeles en mecanismos biológicos de interés relacionados con la situación experimental estudiada (Esteban y Wall, 2011). Por este motivo y dado el importante papel que juegan los microarrays en la actualidad en el análisis de la expresión génica, en los últimos años se ha propuesto un método alternativo para el análisis de expresión génica de microarrays basado en la Teoría de Juegos Cooperativos (Moretti et al., 2007).

La Teoría de Juegos es una rama de las matemáticas basada en el diseño de modelos para la toma de decisiones a través de una serie de incentivos (juegos). John von Neumann y Oskar Morgenstern describieron en su obra "Teoría de Juegos y Comportamiento Económico" (1944) que la Teoría de Juegos podía subdividirse en dos grupos en función del tipo de interacción (o *juego*). Según este criterio, distinguimos entre juegos cooperativos y no cooperativos: ambos se diferencian en el hecho de que en la modalidad de juegos no cooperativos, los jugadores -o elementos que participan en el juego-, no tienen la capacidad de tomar acuerdos vinculantes, mientras que esta capacidad sí que la presentan los jugadores de los juegos cooperativos.

De forma general podemos decir que la Teoría de Juegos Cooperativos es un método matemático que describe el fenómeno de conflicto y cooperación entre los encargados de tomar decisiones racionales inteligentes (Shi et al., 2012). Esta teoría puede aplicarse a múltiples ámbitos de estudio; en nuestro caso, los jugadores serían los genes.

La principal ventaja de este método es la asignación a cada gen de un índice numérico, el denominado valor de *Shapley*, que representa la relevancia de cada gen bajo una cierta condición teniendo en cuenta simultáneamente el comportamiento del resto de genes bajo la misma condición. Este método, en combinación con la estadística, ha demostrado ser útil en el análisis diferencial

de datos de expresión génica, (Albino et al., 2008; Esteban y Wall, 2011; Moretti et al., 2008) de modo que incluso aumenta la sensibilidad en la detección de genes expresados diferencialmente con respecto a las técnicas de análisis convencionales (Esteban y Wall, 2011).

El valor de *Shapley*, elemento clave cuando se habla de Teoría de Juegos Cooperativos, podría definirse como "la contribución marginal esperada de cada jugador (cada gen) al entrar en una coalición al azar" (Pérez, Jimeno y Cerdá, 2010).

Para llevar a cabo un análisis de datos de expresión génica obtenidos mediante microarrays a través de Teoría de Juegos Cooperativos, es necesario partir de una matriz de datos normalizados sobre la cual se realizan los análisis pertinentes. Posteriormente, mediante la utilización de software computacional, se aplican una serie de algoritmos sobre los datos con la finalidad de que éstos sean sometidos a rigurosos análisis estadísticos que permitirán discernir la expresión diferencial de determinados genes en diferentes situaciones.

Finalmente y con el objetivo de categorizar de alguna forma aquellos genes que muestran una expresión diferencial (obtenidos por diferentes aproximaciones), las herramientas bioinformáticas actuales permiten confeccionar análisis funcionales que proporcionan interesantes resultados acerca de los mecanismos, rutas y procesos en que se encuentran implicados dichos genes, posibilitando así una interpretación coherente y que pueda ser justificada desde el punto de vista de la Biología.

El objetivo de este Trabajo Fin de Grado es aplicar la Teoría de Juegos Cooperativos al análisis de la expresión de genes en autismo, comparando los resultados obtenidos con aquellos generados utilizando procedimientos de análisis convencionales. Para llevar a cabo dicho objetivo, se trató de reproducir los resultados obtenidos a partir de un estudio publicado recientemente. De modo general, este Trabajo Fin de Grado me ha permitido profundizar en los conocimientos relacionados (biológicos, experimentales y estadísticos) adquiridos durante el Grado y conocer en detalle su aplicación real.

## 2. MATERIAL Y MÉTODOS

### 2.1. Datos

Se analizaron los datos del experimento GSE25507, obtenido a partir de la base de datos de acceso público *Gene Expression Omnibus* (GEO) (<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE25507>). Se trata de datos procedentes de un *array* de un color (*Affymetrix Human Genome Array*) que permite analizar la expresión de más de 47000 transcritos y sus variantes, para lo cual cuenta con 54675 sondas (Alter et al., 2011).

El grupo de datos objeto de análisis consta de 146 muestras, 82 de las cuales pertenecen a pacientes con autismo y 64 a individuos sanos. La información acerca de la edad de los padres solo estaba disponible para 135 de los 146 individuos totales, por lo que únicamente se incluyeron en el análisis estadístico aquellos individuos de los cuales se conocía la edad paterna. Así pues, el conjunto de datos a incluir en este trabajo constaba de 78 individuos con autismo y 57 individuos control.

Dentro del grupo de individuos control, se estableció además una subdivisión en función de la edad de los padres, de modo que encontramos un subgrupo de 30 individuos control hijos de padres jóvenes (de edades inferiores a 31 años) y un subgrupo de 27 individuos control hijos de padres mayores (de edades iguales o superiores a 31 años).

Para el análisis estadístico se utilizaron paquetes de Bioconductor (<http://www.bioconductor.org/>) ejecutables en R (<http://www.r-project.org/>), que incluyen multitud de funciones estadísticas que permiten una completa evaluación de los datos. Se trata de un software de acceso libre empleado en estadística computacional que proporciona prácticamente toda la variedad de técnicas estadísticas de uso general y muchas recientes y de uso particular. Es una herramienta de bajo coste computacional, de uso en cualquier sistema operativo y ampliamente extensible, ya que existen multitud de paquetes de funciones que posibilitan la realización de diferentes tipos de tareas estadísticas.

Como se ha mencionado en la introducción, el estudio de las variaciones en la expresión génica en este conjunto de datos se abordó siguiendo dos estrategias: en primer lugar, un análisis de los datos por métodos convencionales y, por otro lado, un análisis mediante Teoría de Juegos Cooperativos.

## **2.2 Análisis de datos de microarrays por métodos convencionales**

Para el preprocesamiento de los datos, éstos fueron transformados logarítmicamente. A continuación se llevó a cabo la normalización de los datos preprocesados; con la finalidad de eliminar toda la variación que no tuviera un origen biológico, se utilizó la función de normalización RMA, incluida dentro del pack *affy* de Bioconductor para R ([bioconductor.org/packages/release/bioc/html/affy.html](http://bioconductor.org/packages/release/bioc/html/affy.html)). La función RMA realiza una corrección de fondo y la normalización mediante cuantiles, que tiene el potencial para eliminar diferencias a nivel de grupo en las distribuciones de la expresión génica. Debido a las numerosas ventajas que presenta su uso, la función RMA se utilizó en el análisis convencional (Alter et al., 2011).

El preprocesamiento y la normalización de los datos generaron un set de expresión sobre el cual se continuó trabajando. Dicho set de expresión fue sometido a la función PANP (*Presence/ Absence calls from Negative Probes*), que permite establecer la ausencia/presencia de determinados genes en un microarray (*Presence/ Absence Calls*). Asimismo, esta herramienta genera un set de *p* – valores relacionados con el grado de significación de la detección de cada valor de expresión (Warren et al., 2007).

Esta función generó un conjunto de datos en el cual se detallaba en cuántos de los individuos se encontraban presentes o ausentes las diferentes sondas. El análisis continuó con todas aquellas sondas que se encontraban presentes en 50 individuos o más, reduciéndose considerablemente el número de sondas con las que se trabajó a partir de ese momento (15040, de un número inicial de 54675 sondas).

A partir de los *p-valores* generados con la función PANP, se trató de seleccionar los genes significativos diferencialmente expresados entre los distintos grupos que componían el estudio; para ello, se establecieron tres tipos de comparaciones:

- Individuos autistas frente a individuos control hijos de padres jóvenes (menores de 31 años). El primer grupo estaba compuesto por 78 individuos y el segundo, por 30.
- Individuos autistas frente a individuos control hijos de padres mayores. El primer grupo estaba integrado por 78 individuos y el segundo, por 27.
- Individuos control hijos de padres mayores (de 31 años o más) frente a individuos control hijos de padres jóvenes (menores de 31 años). El primer grupo estaba compuesto por 27 individuos y el segundo, por 30.

### **2.3. Análisis de datos de microarrays por Teoría de Juegos Cooperativos**

Para analizar los datos mediante Teoría de Juegos, es necesario partir de la matriz de datos normalizados obtenida gracias a los análisis mediante técnicas convencionales. El análisis realizado sigue pasos similares a los establecidos por Moretti y sus colaboradores (Moretti et al., 2007; Moretti et al., 2008) que se presentan resumidos en la Figura 2.

A partir de los datos de expresión normalizados y con la finalidad de eliminar aquellos genes con niveles de expresión por encima de los valores medios en controles, se genera una matriz Booleana  $B^+$  (matriz de genes pre-seleccionados como sobreexpresados) en la cual los valores iguales o superiores a la media más la desviación estándar son denotados como 1 y aquellos que no cumplen esta condición, como 0. Por otro lado y con el objetivo en este caso de eliminar aquellos genes con niveles de expresión por debajo

de los medidos en controles, se genera una matriz Booleana  $B^-$  (matriz de genes pre-seleccionados con un nivel de expresión inferior a los controles) en la cual los valores de expresión iguales o inferiores a la media menos la desviación estándar son codificados como 1 y aquellos que no cumplen esta condición, como 0.

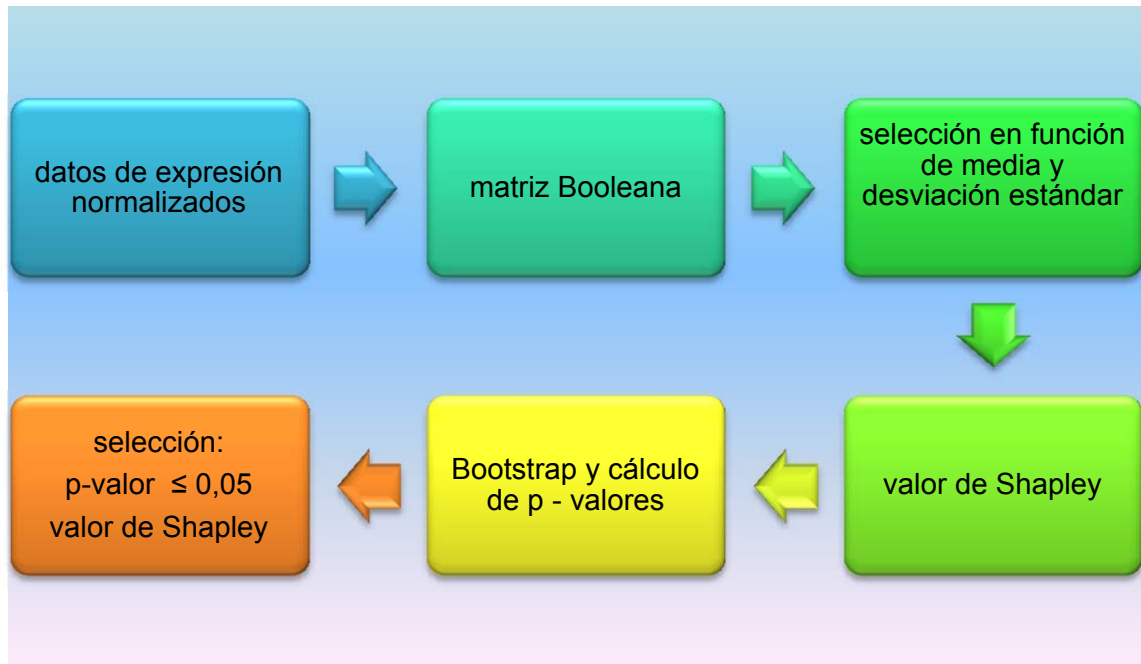
Una vez realizado este paso, la siguiente etapa consiste en realizar una separación de las matrices Booleanas  $B^+$  y  $B^-$  en los distintos grupos experimentales, obteniendo, en cada caso, una matriz Booleana  $B^+$  y  $B^-$  tanto para los individuos control como para los sujetos experimentales.

Tras la separación en grupos de los distintos datos, es necesario proceder al cálculo del valor de *Shapley* para éstos. Posteriormente y con el objetivo de eliminar aquellos genes cuyos valores de *Shapley* puedan ser elevados debido al azar, el siguiente paso consistiría en aplicar un test estadístico tipo Bootstrap (1000 iteraciones) de los valores de *Shapley* (CASH o *Comparative Analysis of Shapley value*), y el posterior cálculo de los *p-valores* no corregidos.

Finalmente, tras este procesamiento de los datos, se seleccionan aquellos genes que cumplan simultáneamente dos condiciones:

- 1.- *p - valor* inferior o igual a 0,05.
- 2.- Diferencia absoluta del valor de *Shapley* entre controles y experimentales superior a la media más la desviación estándar para cada grupo.

De estos genes seleccionados se realiza un análisis de enriquecimiento funcional y deben ser los utilizados para validaciones experimentales posteriores que confirmen su implicación funcional.



**Figura 2.** Esquema general del análisis de microarrays mediante Teoría de Juegos Cooperativos.

#### 2.4. Análisis funcional

El análisis funcional de los resultados obtenidos permite detectar aquellos procesos biológicos y mecanismos moleculares en los cuales se encuentren implicados los genes detectados como diferencialmente expresados en las distintas situaciones. Para llevar a cabo dicho análisis funcional, se emplearon herramientas bioinformáticas de acceso público disponibles en la red como DAVID (*Database for Annotation, Visualization and Integrated Discovery*; <http://david.abcc.ncifcrf.gov/>). Con este tipo de análisis se pretende encontrar grupos de genes que guarden relación con procesos biológicos que puedan verse afectados en el espectro autista y tratar de proponer una explicación biológica coherente al respecto.



### **3. RESULTADOS**

#### **3.1. Análisis convencional**

El análisis de los datos por métodos convencionales proporcionó los resultados que a continuación se indican.

En la comparación establecida entre los individuos con autismo y aquellos controles de padres jóvenes, no se obtuvieron genes expresados diferencialmente. Lo mismo ocurrió al comparar las dos subpoblaciones de individuos control entre sí (padres mayores frente a padres jóvenes). Por otro lado, se obtuvo un gen diferencialmente expresado (RAB8A) al comparar individuos con autismo frente a individuos control de padres mayores. Así pues, la falta de resultados con significación estadística mediante el uso de métodos convencionales justifica la aplicación del procedimiento de Teoría de Juegos previamente descrito para el análisis de estos microarrays.

#### **3.2. Análisis por Teoría de Juegos**

En el análisis por Teoría de Juegos, para las mismas comparaciones realizadas en los análisis convencionales, los resultados obtenidos han puesto de manifiesto la presencia de genes diferencialmente expresados. En todos los casos se ha mantenido la nomenclatura original de los genes.

##### *3.2.1. Autistas frente a controles hijos de padres jóvenes*

Se detectaron 21 genes que muestran expresión diferencial, concretamente un gen con sobreexpresión (*up*; tasa de cambio  $> 1$ ) y 21 con un nivel de expresión por debajo de lo normal (*down*; tasa de cambio  $< -1$ ). El listado de genes diferencialmente expresados aparece en la Tabla 1.

Símbolo	Nombre	Tasa de cambio
KCNAB2	<i>potassium voltage-gated channel, shaker-related subfamily, beta member 2</i>	10,796
FAM120A	<i>family with sequence similarity 120A</i>	-12,906
ROCK2	<i>Rho-associated, coiled-coil containing protein kinase 2</i>	-10,843
GCLC	<i>glutamate-cysteine ligase, catalytic subunit</i>	-10,892
ATP2B1	<i>ATPase, Ca<sup>2+</sup> transporting, plasma membrane 1</i>	-12,383
EDEM3	<i>ER degradation enhancer, mannosidase alpha-like 3</i>	-12,077
USP38	<i>ubiquitin specific peptidase 38</i>	-11,636
C6orf89	<i>chromosome 6 open reading frame 89</i>	-11,002
AGPS	<i>alkylglycerone phosphate synthase</i>	-10,843
GTF3C4	<i>general transcription factor III C, polypeptide 4, 90kDa</i>	-10,915
LOC100133398 , SELT	<i>selenoprotein T; similar to Selenoprotein T</i>	-11,296
RSPRY1	<i>ring finger and SPRY domain containing 1</i>	-12,169
SNTB2	<i>syntrophin, beta 2 (dystrophin-associated protein A1, 59kDa, basic component 2)</i>	-11,328
SLC35A3	<i>solute carrier family 35 (UDP-N-acetylglucosamine (UDP-GlcNAc) transporter), member A3</i>	-11,515
EAF1	<i>ELL associated factor 1</i>	-11,728
ZADH2	<i>zinc binding alcohol dehydrogenase domain containing 2</i>	-11,745
CCNYL1	<i>cyclin Y-like 1</i>	-11,209
22833_at	no identificado	-11,766
RAB11FIP1	<i>RAB11 family interacting protein 1 (class I)</i>	-11,111
NEDD1	<i>neural precursor cell expressed, developmentally down-regulated 1</i>	-1,112
MKLN1	<i>muskelin 1, intracellular mediator containing kelch motifs</i>	-12,332
PAK2	<i>p21 protein (Cdc42/Rac)-activated kinase 2</i>	-11,443

**Tabla 1.** Genes expresados diferencialmente en la comparación establecida entre individuos autistas e individuos control hijos de padres jóvenes.

### 3.2.2. Autistas frente a controles hijos de padres mayores

Se detectaron 19 genes con expresión diferencial; 2 de ellos se encuentran sobreexpresados (*up*; tasa de cambio > 1) y 17 muestran un nivel de expresión inferior a lo normal (*down*; tasa de cambio < -1) . El listado de genes diferencialmente expresados aparece en la Tabla 2.

Símbolo	Nombre	Tasa de cambio
<b>SPON2</b>	<i>spondin 2, extracellular matrix protein</i>	1,2738
<b>TUBB1</b>	<i>tubulin, beta 1</i>	1,1918
<b>CNOT3</b>	<i>CCR4-NOT transcription complex, subunit 3</i>	-1,2406
<b>PRAF2</b>	<i>PRA1 domain family, member 2</i>	-1,1343
<b>RNF24</b>	<i>ring finger protein 24</i>	-1,2439
<b>BTBD2</b>	<i>BTB (POZ) domain containing 2</i>	-1,158
<b>NAPA</b>	<i>N-ethylmaleimide-sensitive factor attachment protein, alpha</i>	-1,2552
<b>PSME3</b>	<i>proteasome (prosome, macropain) activator subunit 3 (PA28 gamma; Ki)</i>	-1,2214
<b>PML, LOC653246</b>	<i>promyelocytic leukemia; similar to promyelocytic leukemia protein isoform 1</i>	-1,1739
<b>CYTH4</b>	<i>cytohesin 4</i>	-1,2373
<b>ARF3</b>	<i>ADP-ribosylation factor 3</i>	-1,2279
<b>MYOF</b>	<i>myoferlin</i>	-1,2894
<b>SMARCB1</b>	<i>SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily b, member 1</i>	-1,122
<b>NADK</b>	<i>NAD kinase</i>	-1,3043
<b>PSD4</b>	<i>pleckstrin and Sec7 domain containing 4</i>	-1,229
<b>G6PC3</b>	<i>glucose 6 phosphatase, catalytic, 3</i>	-1,1462
<b>RHBDD2</b>	<i>rhomboid domain containing 2</i>	-1,2386
<b>POU2F2</b>	<i>POU class 2 homeobox 2</i>	-1,0865
<b>ACOT9</b>	<i>acyl-coA thioesterase 9</i>	-1,149

**Tabla 2.** Genes expresados diferencialmente en la comparación establecida entre individuos autistas e individuos control hijos de padres mayores.

### 3.2.3. Controles hijos de padres jóvenes frente a controles hijos de padres mayores

Se han detectado 67 genes con expresión diferencial, 28 de los cuales se encuentran sobreexpresados (*up*; tasa de cambio > 1) y 39 muestran un nivel de expresión inferior a lo normal (*down*; tasa de cambio < -1) . El listado de genes diferencialmente expresados aparece en la Tabla 3.

<b>Símbolo</b>	<b>Nombre</b>	<b>Tasa de cambio</b>
<b>COASY</b>	<i>Coenzyme A synthase</i>	11,218
<b>TUBQCP2</b>	<i>tubulin, gamma complex associated protein 2</i>	11,029
<b>FUCA1</b>	<i>fucosidase, alpha-L- 1, tissue</i>	11,239
<b>PNPLA6</b>	<i>patatin-like phospholipase domain containing 6</i>	11,478
<b>RPS6KA4</b>	<i>ribosomal protein S6 kinase, 90kDa, polypeptide 4</i>	11,009
<b>SCO2</b>	<i>SCO cytochrome oxidase deficient homolog 2 (levadura)</i>	12,809
<b>SH2B2</b>	<i>SH2B adaptor protein 2</i>	11,907
<b>CSNK1D</b>	<i>casein kinase 1, delta</i>	11,497
<b>CNP</b>	<i>2',3'-cyclic nucleotide 3' phosphodiesterase</i>	11,122
<b>PAPSS1</b>	<i>3'-phosphoadenosine 5'-phosphosulfate synthase 1</i>	1,182
<b>AATF</b>	<i>apoptosis antagonizing transcription factor</i>	10,651
<b>D4S234E</b>	<i>DNA segment on chromosome 4 (unique) 234 expressed sequence</i>	1,143
<b>IFFO1</b>	<i>intermediate filament family orphan 1</i>	11,542
<b>HGS</b>	<i>hepatocyte growth factor-regulated tyrosine kinase substrate</i>	11,598
<b>OGFR</b>	<i>opioid growth factor receptor</i>	11,393
<b>ATP13A1</b>	<i>ATPase type 13A1</i>	11,377
<b>B4GALT5</b>	<i>UDP-Gal:betaGlcNAc beta 1,4- galactosyltransferase, polypeptide 5</i>	12,619
<b>SLC27A3</b>	<i>solute carrier family 27 (fatty acid transporter), member 3</i>	11,354
<b>TSPAN14</b>	<i>tetraspanin 14</i>	10,857
<b>SDF4</b>	<i>stromal cell derived factor 4</i>	11,283
<b>KIAA2013, LOC728138</b>	<i>KIAA2013; hypothetical LOC728138</i>	11,003
<b>ZNFX1</b>	<i>zinc finger, NFX1-type containing 1</i>	11,271
<b>ZNF641</b>	<i>zinc finger protein 641</i>	11,473
<b>IKBKG</b>	<i>inhibitor of kappa light polypeptide gene enhancer in B-cells, kinase gamma</i>	11,519
<b>DNPEP</b>	<i>aspartyl aminopeptidase</i>	10,809
<b>E2F4</b>	<i>E2F transcription factor 4, p107/p130-binding</i>	11,088
<b>SBF1, SBF1P1</b>	<i>SET binding factor 1; SET binding factor 1 pseudogene 1</i>	11,616
<b>N4BP1</b>	<i>NEDD4 binding protein 1</i>	11,103
<b>SDCCAG1</b>	<i>serologically defined colon cancer antigen 1</i>	-11,651
<b>RPLP2, RPLP2P3</b>	<i>ribosomal protein, large, P2 pseudogene 3; ribosomal protein, large, P2</i>	-13,919
<b>ATP6V0E1</b>	<i>ATPase, H+ transporting, lysosomal 9kDa, V0 subunit e1</i>	-13,514
<b>TOPORS</b>	<i>topoisomerase I binding, arginine/serine-rich</i>	-12,741
<b>FLJ10038</b>	<i>hypothetical protein FLJ10038</i>	-11,166
<b>PCSK5</b>	<i>proprotein convertase subtilisin/kexin type 5</i>	-11,614
<b>CASP7</b>	<i>caspase 7, apoptosis-related cysteine peptidase</i>	-13,047

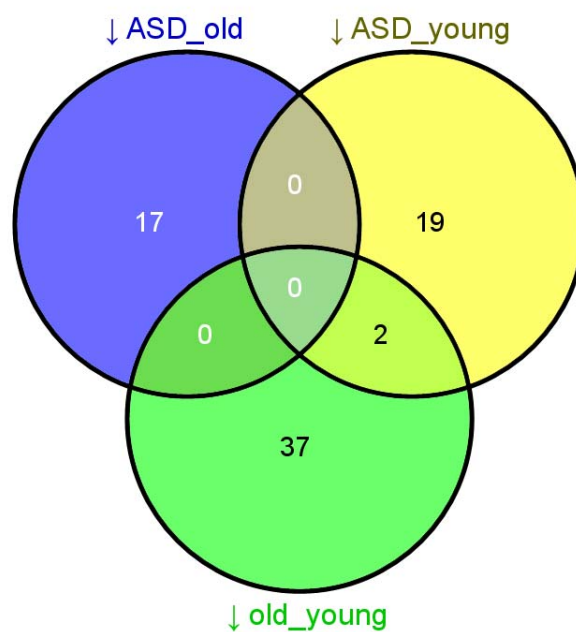
<b>INHBC</b>	<i>inhibin, beta C</i>	-15,007
<b>PTCH1</b>	<i>patched homolog 1 (Drosophila)</i>	-12,012
<b>MAGOH</b>	<i>mago-nashi homolog, proliferation-associated (Drosophila)</i>	-11,983
<b>ZNF292</b>	<i>zinc finger protein 292</i>	-11,948
<b>CHD9</b>	<i>chromodomain helicase DNA binding protein 9</i>	-12,222
<b>TBL1X</b>	<i>transducin (beta)-like 1X-linked</i>	-13,362
<b>MED4</b>	<i>mediator complex subunit 4</i>	-12,236
<b>DERL1</b>	<i>Der1-like domain family, member 1</i>	-12,818
<b>CDK12</b>	<i>Cdc2-related kinase, arginine/serine-rich</i>	-11,188
<b>ERMP1</b>	<i>endoplasmic reticulum metallopeptidase 1</i>	-13,203
<b>COMMD10</b>	<i>COMM domain containing 10</i>	-12,033
<b>FAM122B</b>	<i>family with sequence similarity 122B</i>	-12,948
<b>WDR75</b>	<i>WD repeat domain 75</i>	-11,208
<b>225055_AT</b>	no identificado	-10,887
<b>KIAA1715</b>	KIAA1715	-11,123
<b>JMY</b>	<i>junction mediating and regulatory protein, p53 cofactor</i>	-12,309
<b>KCTD18</b>	<i>potassium channel tetramerisation domain containing 18</i>	-12,518
<b>HEATR5A</b>	<i>HEAT repeat containing 5A</i>	-13,248
<b>SNTB2</b>	<i>syntrophin, beta 2 (dystrophin-associated protein A1, 59kDa, basic component 2)</i>	-11,959
<b>GXYLT1</b>	<i>glycosyltransferase 8 domain containing 3</i>	-1,091
<b>SERBP1</b>	<i>SERPINE1 mRNA binding protein 1</i>	-11,689
<b>STYX</b>	<i>serine/threonine/tyrosine interacting protein</i>	-13,184
<b>IMMP1L</b>	<i>IMP1 inner mitochondrial membrane peptidase-like (S. cerevisiae)</i>	-12,758
<b>MAGEE1</b>	<i>melanoma antigen family E, 1</i>	-11,249
<b>229710_at</b>	no identificado	-11,625
<b>229861_at</b>	no identificado	-13,142
<b>MGA</b>	<i>MAX gene associated</i>	-11,445
<b>MKLN1</b>	<i>muskelin 1, intracellular mediator containing kelch motifs</i>	-12,208
<b>DCP2</b>	<i>DCP2 decapping enzyme homolog (S. cerevisiae)</i>	-11,715
<b>PHAX</b>	<i>phosphorylated adaptor for RNA export</i>	-13,251
<b>RGS3</b>	<i>regulator of G-protein signaling 3</i>	-11,443
<b>SATB1</b>	<i>SATB homeobox 1</i>	-12,987

**Tabla 3.** Genes expresados diferencialmente en la comparación establecida entre individuos control hijos de padres jóvenes e individuos control hijos de padres mayores.

Para comprobar si existía solapamiento entre genes en alguna de las situaciones, se elaboró un diagrama de Venn. Los resultados (Figura 3)

muestran la existencia de dos genes que se encuentran expresados a la baja tanto en la comparación establecida entre individuos autistas e individuos control de padres jóvenes como en la comparación realizada entre individuos control de padres jóvenes frente a individuos control de padres mayores. Se trata de los genes MKLN1 y SNTB2. En la actualidad, ninguno de ellos ha sido demostrado como implicado en autismo. Asimismo, se llevó a cabo el análisis funcional de estos dos genes con el objetivo de determinar si participan en algún proceso biológico común, no obteniéndose resultados significativos.

En lo que respecta a los genes que muestran sobreexpresión, no se encontraron en ningún caso coincidencias entre los diferentes grupos experimentales.



**Figura 3.** Diagrama de Venn que muestra la existencia de genes comunes con expresión diferencial en los casos experimentales analizados. ASD: individuos autistas; young: individuos control hijos de padres jóvenes; old: individuos control hijos de padres mayores.

### 3.3 Análisis funcional

El análisis funcional de los distintos grupos experimentales permitió identificar las rutas metabólicas en las que se ven implicados distintos grupos de genes que muestran una expresión diferencial.

Los genes con expresión diferencial (tanto por encima como por debajo de los niveles normales) en el caso de la comparación establecida entre individuos autistas e individuos control de padres mayores comparten de modo estadísticamente significativo una serie de procesos biológicos, relacionados fundamentalmente con metabolismo proteico y ciclo celular. Los genes diferencialmente expresados en esta comparación y que se encuentran involucrados en estos procesos biológicos figuran en la Tabla 4.

10 de los genes diferencialmente expresados no parecían intervenir en ninguna de estas rutas ni parecían compartir procesos biológicos comunes.

proceso biológico	<i>p</i> - valor	genes implicados	FDR
transporte de proteínas	0,041	ARF3, NAPA, PRAF2, PML	43
establecimiento de localización de proteínas	0,042	ARF3, NAPA, PRAF2, PML	44
ciclo celular	0,043	SMARCB1, PML, PSME3, TUBB1	44
regulación de la transducción de señales de la proteína ARF	0,045	CYTH4, PSD4	46
localización proteica	0,059	ARF3, NAPA, PRAF2, PML	56

**Tabla 4.** Procesos biológicos en los que aparecen implicados los genes con expresión diferencial en la comparación entre individuos con autismo e individuos controles hijos de padres mayores. FDR: *false discovery rate*.

En la comparación establecida entre los individuos con autismo y aquellos controles hijos de padres jóvenes, únicamente se detectó un proceso biológico en el que participan conjuntamente algunos de los genes expresados diferencialmente: se trata de la homeostasis redox celular y en ella se encuentran involucrados los genes GCLC y SELT (ambos con unos niveles de expresión por debajo de lo normal) (Tabla 5) . Por otra parte, 19 de los genes diferencialmente expresados no estaban vinculados a ningún proceso biológico común.

proceso biológico	p - valor	genes implicados	FDR
homeostasis redox celular	0,063	GCLC, SELT	56

**Tabla 5.** Proceso biológico en el que aparecen implicados los genes con expresión diferencial en la comparación entre individuos con autismo e individuos controles hijos de padres jóvenes. FDR: *false discovery rate*.

En la comparación establecida entre individuos control hijos de padres mayores e individuos control hijos de padres jóvenes, aparecen diferentes procesos biológicos, fundamentalmente relacionados con linfocitos y desarrollo, en los que intervienen algunos de los genes que muestran expresión diferencial (Tabla 6). Además, 36 de los genes diferencialmente expresados no participan en ningún proceso biológico común.

proceso biológico	p - valor	genes implicados	FDR
procesos metabólicos de nucleósidos bisfosfato	0,038	PAPSS1, COASY	43
homeostasis de linfocitos B	0,038	SH2B2, IKBKG	43
procesos homeostáticos	0,044	ATP6V0E1, DERL1, E2F4, SCO2, SH2B2, IKBKG, PTCH1	49
morfogénesis del apéndice	0,046	225718, 209815, 205559	50
morfogénesis de las extremidades	0,046	KIAA1715, PTCH1, PCSK5	50
desarrollo de las extremidades	0,050	KIAA1715, PTCH1, PCSK5	53
desarrollo del apéndice	0,050	KIAA1715, PTCH1, PCSK5	53
proteolisis	0,070	DERL1, IMMP1L, DNPEP, CASP7, ERMP1, PCSK5, TOPORS, TBL1X	65
regulación de procesos metabólicos del ARN	0,090	E2F4, MGA, SATB1, SERBP1, AATF, JMY, MED4, RPS6KA4, TBL1X, ZNF641, ZNFX1	75
homeostasis de linfocitos	0,093	SH2B2, IKBKG	76
procesos catabólicos del ARNm transcrito en el núcleo, degradación del ARNm mediada por mutación terminadora	0,099	DCP2, MAGOH	78

**Tabla 6.** Procesos biológicos en los que aparecen implicados los genes con expresión diferencial en la comparación entre individuos controles hijos de padres mayores e individuos controles hijos de padres jóvenes. FDR: *false discovery rate*.



#### 4. DISCUSIÓN

Los análisis por métodos convencionales evidencian la existencia de un gen que muestra una expresión diferencial entre individuos autistas e individuos control de padres mayores, lo cual sugiere que a medida que se incrementa la edad paterna es mayor la incidencia del autismo. Se trata del gen RAB8A, miembro de la familia de oncogenes RAS. La proteína codificada por este gen jugaría un importante papel en el transporte de proteínas desde el retículo endoplasmático hasta el aparato de Golgi y la membrana plasmática (<http://www.genecards.org/cgi-bin/carddisp.pl?gene=RAB8A>). Sin embargo, a día de hoy no existe ninguna evidencia de que este gen se encuentre relacionado con la aparición del autismo.

Los análisis por Teoría de Juegos aumentan la sensibilidad en el estudio de la expresión génica en lo que a detección de genes diferencialmente expresados se refiere. Algunos de los genes diferencialmente expresados en nuestras comparaciones ya habían sido previamente descritos como implicados en autismo (<http://genotator.hms.harvard.edu/geno/disorder/autism/>). Es el caso del gen PAK2 (*p21 protein (Cdc42/Rac)-activated kinase 2*), relacionado con procesos de apoptosis. Su expresión se encuentra disminuida en la comparación establecida entre individuos autistas e individuos control hijos de padres jóvenes. Dicho gen se encuentra ubicado en el cromosoma 3, en una región que previamente había sido descrita como implicada en autismo y otros desórdenes intelectuales como la esquizofrenia (Ballif et al., 2008; Carroll et al., 2011). El hecho de que su expresión disminuya en los pacientes control hijos de padres jóvenes reforzaría la implicación del gen en la aparición del autismo, cuya expresión iría incrementándose a medida que aumenta la edad paterna.

De igual forma, el gen TBL1X (*transducin (beta)-like 1X-linked*), cuya expresión se encuentra por debajo de los niveles normales en la comparación establecida entre individuos control hijos de padres jóvenes e individuos control hijos de padres mayores, también había sido previamente implicado en autismo. Este gen se encuentra asociado a la ruta de señalización de *Wnt*, que juega un importante papel en la señalización celular durante la embriogénesis,

siendo de gran importancia en la adecuada formación del embrión (Chung et al., 2011).

Además, como se destaca en el apartado de resultados, existía un solapamiento entre genes diferencialmente expresados en las comparaciones entre el grupo de individuos autistas y controles hijos de padres jóvenes y el grupo de individuos control hijos de padres jóvenes y controles hijos de padres mayores. En concreto, los genes MKLN1 y SNTB2 mostraban una expresión diferencial en estos casos, presentando un nivel de expresión por debajo de lo habitual. El primero de ellos, MKLN1, es el gen de la muskelina 1; la proteína codificada por este gen actúa como mediador de la propagación celular y en procesos de respuesta del citoesqueleto frente a un componente de la matriz extracelular, la trombospondina (Adams et al., 1998). En la actualidad no se ha establecido relación alguna entre el gen MKLN1 y el autismo, aunque sí se han podido reportar evidencias de su implicación en procesos de regulación del transporte neuronal debido a su interacción con receptores GABAérgicos (Heisler et al., 2011).

En lo que respecta al gen de la sintrofina  $\beta_2$  (SNTB2), no se le ha atribuido ninguna función relacionada con el sistema nervioso, aunque sí se han encontrado evidencias de su participación en procesos relacionados con el sistema locomotor. Algunas de las enfermedades asociadas con deficiencias en este gen son la distrofia muscular de Duchenne o la de Becker (<http://www.genecards.org/cgi-bin/carddisp.pl?gene=SNTB2>). La proteína codificada por este gen se perfila como una importante mediadora de la transducción de señales en diversos procesos del metabolismo (Bhat et al., 2013).

Otros genes que hemos podido identificar como diferencialmente expresados no han sido aún catalogados como candidatos a autismo como tal, a pesar de existir literatura en la cual se evidencia que pueden llegar a desempeñar un papel importante en algunos de los aspectos moleculares y fenotípicos de la enfermedad. Algunos ejemplos se describen a continuación.

La subunidad catalítica de la glutamato cisteína ligasa (GCLC), implicada en la síntesis de la enzima glutatión transferasa, muestra un nivel de

expresión por debajo de lo normal en individuos controles hijos de padres jóvenes, fue recientemente descrita como elevada en individuos autistas (Gu et al., 2013). Este hecho afianzaría aún más la evidencia de la asociación entre el aumento de la edad paterna y la incidencia del autismo.

Asimismo, se han identificado mutaciones en el gen SLC35A3 (identificado como expresado por debajo de lo normal en individuos controles hijos de padres jóvenes) que provocan la aparición de epilepsia y del espectro autista. Se trata de un gen que codifica el principal transportador de UDP - N - acetilglucosamina del aparato de Golgi. Las mutaciones en este transportador provocarían la alteración de las glicoproteínas en la superficie celular, reduciendo su número y por tanto modificando su función habitual (Edvardson et al., 2013).

Por otro lado, el gen de la caspasa 7 (CASP7), una cisteína peptidasa relacionada con procesos de apoptosis, muestra un nivel de expresión bajo en individuos controles hijos de padres jóvenes cuando se compara con individuos controles hijos de padres mayores. Al parecer, este gen podría estar detrás de la etiología del autismo y representa por tanto un importante biomarcador sobre el cual actuar en el tratamiento de la enfermedad (El-Ansary y Al-Ayadhi, 2012).

Recientemente se reportó una pequeña delección en el gen PTCH1 que estaría relacionada con el desarrollo de autismo y el síndrome de Gorlin (Delbroek et al., 2010). Este gen aparece como expresado por debajo de lo normal en individuos control hijos de padres jóvenes cuando se comparan con individuos control hijos de padres mayores, lo cual entra en conflicto con la asociación entre el aumento de la edad paterna y la incidencia del autismo. Por ello, serían necesarios estudios posteriores para poder emitir conclusiones al respecto.

Además, existe una duplicación en la cual se ve implicado otro de los genes detectados como diferencialmente expresados, concretamente el gen KCTD18, que contribuiría al desarrollo del fenotipo autista (Usui et al., 2013). Dicho gen aparece con un nivel de expresión inferior al normal en los individuos control hijos de padres jóvenes al compararlos con controles hijos de padres mayores.

En lo que respecta al trabajo de Alter y sus colaboradores a partir del cual se plantea el nuestro, existen notables diferencias en los resultados, debidas fundamentalmente al hecho de la utilización por su parte de *p-valores* no corregidos (en nuestro caso, se trabaja con *p-valores* corregidos), con lo cual el incremento en los falsos positivos es patente. Por otra parte, el trabajo de Alter, a diferencia del nuestro, trata de buscar variaciones en los niveles de expresión génica global en lugar de genes diferencialmente expresados, con lo cual los resultados que se obtienen difieren ampliamente de los nuestros en determinados aspectos. Además, en nuestro trabajo se llevó a cabo un paso adicional, el análisis con Teoría de Juegos.

En definitiva y como conclusión, la Teoría de Juegos incrementa ampliamente el poder de detección de genes con expresión diferencial. Así se ha demostrado también en otros estudios (Moretti et al., 2007; Esteban y Wall, 2011).

Debido a sus múltiples ventajas en el manejo del gran volumen de datos procedentes de microarrays en autismo, su uso podría extenderse para analizar otros datos de expresión génica con la finalidad de caracterizar el perfil genético responsable de la aparición de enfermedades complejas y abordar así el diseño de posibles dianas terapéuticas para restablecer los niveles funcionales adecuados de las alteraciones moleculares detectadas.

## 5. BIBLIOGRAFÍA

Adams JC, Seed B, Lawler J. 1998. Muskelin, a novel intracellular mediator of cell adhesive and cytoskeletal responses to thrombospondin-1. *EMBO J.* 17: 4964-4974.

Albino D, Scaruffi P, Moretti S, Coco S, Truini M, Di Cristofano C, Cavazzana A, Stigliani S, Bonassi S and Tonini GP. 2008. Identification of low intratumoral gene expression heterogeneity in neuroblastic tumors by genome-wide expression analysis and game theory. *Cancer.* 113: 1412-1422.

- Alter MD, Kharkar R, Ramsey KE, Craig DW, Melmed RD, et al. 2011. Autism and increased paternal age related changes in global levels of gene expression regulation. *PLoS One*. 6: e16715.
- Anitha A, Thanseem I, Nakamura K, Vasu MM, Yamada K, Ueki T, Iwayama Y, Toyota T, Tsuchiya KJ, Iwata Y, Suzuki K, Sugiyama T, Tsujii M, Yoshikawa T, Mori N. 2014. Zinc finger protein 804A (ZNF804A) and verbal deficits in individuals with autism. *J Psychiatry Neuro sci*. (in press).
- Ballif BC, Theisen A, Coppinger J, Gowans GC, Hersh JH, Madan-Khetarpal S, Schmidt KR, Tervo R, Escobar LF, Friedrich CA, McDonald M, Campbell L, Ming JE, Zackai EH, Bejjani BA, Shaffer LG. 2008. Expanding the clinical phenotype of the 3q29 microdeletion syndrome and characterization of the reciprocal microduplication. *Mol Cytogen*. 1: 8.
- Bhat HF, Adams ME, Khanday FA. 2013. Syntrophin proteins as Santa Claus: role(s) in cell signal transduction. *Cell Mol Life Sci*. 70: 2533-2554.
- Berko ER, Suzuki M, Beren F, Lemetre C, Alaimo CM, et al. 2014. Mosaic Epigenetic Dysregulation of Ectodermal Cells in Autism Spectrum Disorder. *PLoS Genet*. 10: e1004402.
- Bolstad BM, Irizarry RA, Astrand M, Speed TP. 2003. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics*. 19: 185-193.
- Buxbaum JD, Hof PR. 2012. *The Neuroscience of Autism Spectrum Disorders*, 1st Edition. Academic Press.
- Carroll LS, Williams HJ, Walters J, Kirov G, O'Donovan MC, Owen MJ. 2011. Mutation Screening of the 3q29 Microdeletion Syndrome Candidate Genes DLG1 and PAK2 in Schizophrenia. *Am J Med Gen. (Part B)* 156: 844–849.
- Chung RH, Ma D, Wang K, Hedges DJ, Jaworski JM, Gilbert JR, Cuccaro ML, Wright HH, Abramson RK, Konidari I, Whitehead PL, Schellenberg

- GD, Hakonarson H, Haines JL, Pericak-Vance MA, Martin E.R. 2011. An X chromosome-wide association study in autism families identifies TBL1X as a novel autism spectrum disorder candidate gene in males. *Mol Autism*. 2: 18.
- Correia CT, Conceição IC, Oliveira B, Coelho J, Sousa I, Sequeira AF, Almeida J, Café C, Duque F, Mouga S, Roberts W, Gao K, Lowe JK, Thiruvahindrapuram B, Walker S, Marshall CR, Pinto D, Nurnberger JI, Scherer SW, Geschwind DH, Oliveira G, Vicente AM. 2014. Recurrent duplications of the annexin A1 gene (ANXA1) in autism spectrum disorders. *Mol Autism*. (in press).
- Dale JW, von Schantz M. 2002. From genes to genomes. Concepts and applications of DNA technology. John Wiley & Sons. pp 227-244.
- Delbroek H, Steyaert J, Legius E. 2010. An 8.9 year old girl with autism and Gorlin syndrome. *Eur J Paediatr Neurol*. 15: 268-270.
- Ding N, Qu H, Fang X. 2014. The ENCODE project and functional genomics studies. *Yi Chuan*. 36: 237-247.
- Do JH, Choi DK. 2006. Normalization of microarray data: single-labeled and dual-labeled arrays. *Mol Cells*. 22: 254-261.
- Draghici S. 2003. Data analysis tools for DNA microarrays. Chapman & Hall/CRC Press, UK.
- Dudoit S, Popper-Shaffer J, Boldrick JC. 2003. Multiple hypothesis testing in microarray experiments. *Stat Sci*. 18: 71-103.
- Edvardson S, Ashikov A, Jalas C, Sturiale L, Shaag A, Fedick A, Treff NR, Garozzo D, Gerardy-Schahn R, Elpeleg O. 2013. Mutations in SLC35A3 cause autism spectrum disorder, epilepsy and arthrogyriposis. *J Med Genet*. 50: 733-739.
- El-Ansary A, Al-Ayadhi L. 2012. Neuroinflammation in autism spectrum disorders. *J Neuroinflammation*. 11: 265.

- Esteban FJ, Wall DP. 2011. Using game theory to detect genes involved in Autism Spectrum Disorder. *TOP*. 19: 121-129.
- Gu F, Chauhan V, Chauhan A. 2013. Impaired synthesis and antioxidant defense of glutathione in the cerebellum of autistic subjects: alterations in the activities and protein expression of glutathione-related enzymes. *Free Radic Biol Med*. 65: 488-496.
- Heisler FF, Loebrich S, Pechmann Y, Maier N, Zivkovic AR, Tokito M, Hausrat TJ, Schweizer M, Bähring R, Holzbaur ELF, Schmitz D, Kneussel M. 2011. Muskelin Regulates Actin Filament- and Microtubule-Based GABA<sub>A</sub> Receptor Transport in Neurons. *Neuron*. 70: 66–81.
- Jiang YH, Wang Y, Xiu X, Choy KW, Pursley AN, Cheung SW. 2014. Genetic diagnosis of autism spectrum disorders: The opportunity and challenge in the genomics era. *Critical Reviews in Clinical Laboratory Sciences*. 30: 1-14.
- Lambert N, Wermenbol V, Pichon B, Acosta S, van den Aamele J, Perazzolo C, Messina D, Musumeci MF, Dessars B, De Leener A, Abramowicz M, Vilain C. 2014. A Familial Heterozygous Null Mutation of MET in Autism Spectrum Disorder. *Autism Research*. (in press).
- Lee NH, Saeed AI. 2007. Microarrays: an overview. *Methods Mol Biol*. 353: 265-300.
- Lugo JN, Smith GD, Arbuckle EP, White J, Holley AJ, Floruta CM, Ahmed N, Gomez MC, Okonkwo O. 2014. Deletion of PTEN produces autism-like behavioral deficits and alterations in synaptic proteins. *Front Mol Neurosci*. 16: 27.
- Mbadiwe T, Millis RM. 2013. Epigenetics and Autism. *Autism Res Treat*. (in press).
- Moreno V, Solé X. 2004. Uso de chips de ADN (microarrays) en medicina: Fundamentos técnicos y procedimientos básicos para el análisis estadístico de resultados. *Medicina Clínica*. 122 (Supl 1), 73–79.

- Moretti S, van Leeuwen D, Gmuender H, Bonassi S, van Delft J, Kleinjans J et al. 2008. Combining Shapley value and statistics to the analysis of gene expression data in children exposed to air pollution. *BMC Bioinformatics*. 9: 361.
- Moretti S, Patrone F, Bonassi S. 2007. The class of Microarray games and the relevance index for genes. *TOP*. 15: 265-280.
- Neumann J, Morgenstern O. 1944. *Theory of Games and Economic Behavior*. Princeton University Press.
- Pérez J, Jimeno JL, Cerdá E. 2010. *Teoría de Juegos*. Madrid. Pearson Prentice Hall.
- Pop-Jordanova N, Plasevska-Karanfilska D. 2014. Autism - genetics, electrophysiology and clinical syndromes. *Prilozi*. 35: 133-146.
- Quackenbush, J. 2006. Microarray Analysis and Tumor Classification. *N Engl J Med*. 354: 2463-2472.
- Rogers S, Cambrosio A. 2007. Making a new technology work: the standardization and regulation of microarrays. *Yale Journal of Biology and Medicine*. 80: 165–178.
- Shalon D, Smith SJ, Brown PO. 1996. A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization. *Genome Res*. 6: 639–645.
- Shi L, Perkins RG, Fang H, Tong W. 2008. Reproducible and reliable microarray results through quality control: good laboratory proficiency and appropriate data analysis practice are essential. *Current Opinion in Biotechnology*. 19: 10-18.
- Shi H, Wang W, Kwok NM, Chen SY. 2012. Game theory for wireless sensor networks: a survey. *Sensors*. 12: 9055–9097.
- Usui D, Shimada S, Shimojima K, Sugawara M, Kawasaki H, Shigematu H, Takahashi Y, Inoue Y, Imai K, Yamamoto T. 2013. Interstitial



duplication of 2q32.1-q33.3 in a patient with epilepsy, developmental delay, and autistic behavior. *Am J Med Genet.* 161: 1078-1084.

Wall DP, Esteban FJ, Deluca TF, Huyck M, Monaghan T, Velez de Mendizabal N, Goñi J, Kohane IS. 2009. Comparative analysis of neurological disorders focuses genome-wide search for autism genes. *Genomics.* 93: 120-129.

Wang Y, Fang Y, Zhang F, Xu M, Zhang J, Yan J, Ju W, Brown WT, Zhong N. 2014. Hypermethylation of the enolase gene (ENO2) in autism. *Eur J Pediatr.* (in press).

Warren P, Taylor D, Martini PGV, Jackson J, Bienkowska J. 2007. PANP - a new method of gene detection on oligonucleotide expression arrays. In 7th IEEE International Conference on Bio-informatics and Bioengineering. Boston, MA. pp. 108–115.

Warrier V, Baron-Cohen S, Chakrabarti B. 2013. Genetic variation in GABRB3 is associated with Asperger syndrome and multiple endophenotypes relevant to autism. *Mol Autism.* 4: 48.

Watson JD, Baker TA, Bell SP, Gann A, Levine M, Losik R; with Harrison SC. 2014. Molecular biology of the gene. Seventh Edition. Boston: Pearson/CSH Press.