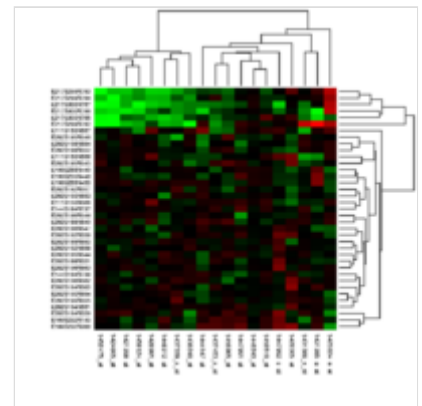


# Perfil de expresión génica

En el campo de la biología molecular, el **perfil de expresión génica** es la medida de la actividad (de la expresión génica) de miles de genes simultáneamente, para crear una imagen global de la función celular. Estos perfiles pueden, por ejemplo, distinguir entre las células que se están dividiendo activamente, o mostrar cómo las células reaccionan a un tratamiento en particular. Muchos experimentos de este tipo analizan un genoma completo simultáneamente, es decir, cada gen presente en una célula en particular.

La tecnología de chip de ADN<sup>1</sup> mide la actividad relativa de los genes diana previamente identificados. También se utilizan para perfiles de expresión génica las técnicas basadas en la secuenciación, como el análisis en serie de la expresión génica SuperSAGE. La técnica SuperSage es especialmente precisa y puede medir cualquier gen activo y no solo un conjunto predeterminado. La llegada de tecnologías de nueva generación de secuenciación ha hecho del análisis de expresión basado en la secuencia una alternativa “digital” llamada tecnología RNA-Seq (secuenciación masiva de transcriptoma), cada vez más popular comparada con la tecnología de microarreglos. Sin embargo la tecnología de microarreglos es mucho más común, considerando los 17 000 artículos PubMed de 2006.<sup>2</sup>



Los mapas de calor de valores de expresión génica muestran cómo las condiciones experimentales influyen en la producción (expresión) de ARNm para un conjunto de genes. El color verde indica una expresión reducida. El análisis de clusters ha colocado un conjunto de genes con regulación negativa en la esquina superior izquierda.

## Antecedentes

El perfil de expresión es el siguiente paso lógico después de la secuenciación del genoma: la secuencia nos dice lo que posiblemente podría hacer la célula, mientras que el perfil de expresión nos dice lo que está haciendo realmente en un momento determinado. Los genes contienen las instrucciones para producir ARN mensajero (ARNm), pero en cualquier momento cada célula produce ARNm a partir de sólo una fracción de los genes que porta. Si un gen se utiliza para producir ARNm se considera activado (“on”), si no desactivado (“off”). Hay muchos factores que determinan si un gen está activado o desactivado tales como la hora del día, si la célula se está dividiendo activamente o no, su medio ambiente local y las señales químicas procedentes de otras células. Por ejemplo, las células de la piel, las células del hígado y las células nerviosas pueden activar (expresar) diferentes genes, que es en gran parte lo que las hace diferentes. Por lo tanto, un perfil de expresión nos permite inferir un tipo de célula, el estado, el entorno y así sucesivamente.

Los experimentos de perfil de expresión a menudo implican la medida de la cantidad relativa de ARNm expresado en dos o más condiciones experimentales. Esto es porque niveles alterados de una secuencia específica de ARNm sugieren una necesidad de cambio para la proteína codificada por el ARNm, tal vez indicando una respuesta homeostática o una condición patológica. Por ejemplo, niveles más altos de RNAm que codifica para la enzima alcohol deshidrogenasa sugieren que las células o tejidos en estudio están respondiendo al aumento de los niveles de etanol en su entorno. Del mismo modo, si las células de cáncer de mama expresan niveles más altos de ARNm asociados con un receptor transmembrana en particular que las células normales, puede ser que este receptor juegue un papel en el cáncer de mama. Un medicamento que interfiera con este receptor podría prevenir o tratar el cáncer de mama. En el desarrollo de un medicamento se pueden realizar experimentos de perfiles de expresión génica para ayudar a evaluar la

toxicidad de dicho medicamento, quizás mediante la búsqueda de cambios en los niveles de la expresión de genes del citocromo P450, que puede ser un biomarcador del metabolismo de los fármacos.<sup>3</sup> Los perfiles de expresión génica pueden convertirse en una prueba de diagnóstico importante.<sup>4 5</sup>

## Comparación con proteómica

---

El genoma humano contiene alrededor de 25 000 genes que trabajan en conjunto para producir alrededor de 1 000 000 proteínas distintas. Esto es debido al corte y splicing alternativo y también debido a que las células llevan a cabo importantes cambios en las proteínas a través de la modificación postraduccional después de haberlas sintetizado primero, por lo que un gen dado sirve como base para muchas versiones posibles de una proteína en particular. En cualquier caso, una única espectrometría de masas puede identificar alrededor de 2 000 proteínas<sup>6</sup> o el 0.2% del total. Si bien es más importante saber la proteína exacta que sintetiza la célula (proteómica) que saber cuánto ARN mensajero se fabrica de cada gen, los perfiles de expresión génica ofrecen una imagen global en un único ensayo.

## Empleo para generación de hipótesis y experimentos

---

A veces un científico ya tiene una idea de lo que está pasando, una hipótesis, y lleva a cabo un experimento de perfiles de expresión con la idea de potencialmente refutar esta hipótesis. En otras palabras, el científico está haciendo una predicción específica sobre los niveles de expresión que podría resultar ser falsa.

Usualmente el perfil de expresión se lleva a cabo antes de saber lo suficiente acerca de cómo los genes interactúan en diferentes condiciones experimentales para que exista una hipótesis comprobable. Si no hay una hipótesis, no hay algo que refutar, pero el perfil de expresión puede ayudar a identificar una posible hipótesis para futuros experimentos. La mayoría de los primeros experimentos de determinación de perfiles de expresión, y muchos de los actuales, tienen esta forma<sup>7</sup> que se conoce como descubrimiento de clases. Un enfoque usual de descubrimiento de clase implica agrupar similares genes o muestras utilizando el método de agrupación K-means o métodos de agrupación jerárquica. Además de seleccionar un [algoritmo de agrupamiento], el usuario normalmente tiene que elegir una medida de proximidad apropiada (distancia o similitud) entre los objetos de datos.<sup>8</sup> La figura anterior representa la salida de un agrupamiento en dos dimensiones, en el que muestras similares (filas, arriba) y similares sondas de genes (columnas) se organizaron de tal forma que pudieran estar juntos. La forma más simple de descubrimiento de clase sería una lista de todos los genes que cambiaron en más de una cierta cantidad entre dos condiciones experimentales.

La predicción de clase es más difícil que el descubrimiento de clase, pero permite responder a las preguntas de relevancia clínica tales como, dado este perfil, ¿cuál es la probabilidad de que este paciente va a responder a este medicamento? Esto requiere de muchos ejemplos de perfiles que respondieron y no respondieron, así como de técnicas de validación cruzada para poder discriminar entre ellos.

## Limitaciones

---

En general, los estudios de perfiles de expresión informan acerca de aquellos genes que mostraron diferencias estadísticamente significativas bajo condiciones experimentales diferentes. Esto es típicamente una pequeña fracción del genoma por varias razones. En primer lugar, diferentes células y tejidos expresan un subconjunto de genes como una consecuencia directa de la diferenciación celular, por lo tanto, muchos genes se desactivan. En segundo lugar, muchos de los genes codifican para proteínas que son necesarias para la supervivencia en cantidades muy específicas, por lo tanto, muchos genes no cambian. En tercer lugar, las células utilizan muchos otros mecanismos para regular las proteínas además de alterar la cantidad de ARNm, por lo tanto, estos genes pueden permanecer expresados constantemente incluso cuando las

concentraciones de las proteínas están subiendo y bajando. En cuarto lugar, las limitaciones financieras restringen los experimentos de perfiles de expresión a un pequeño número de observaciones del mismo gen en condiciones idénticas, lo que reduce el poder estadístico del experimento, lo que hace imposible para dicho experimento el poder identificar cambios importantes pero sutiles. Por último, se necesita mucho esfuerzo para poder estudiar la importancia biológica de cada gen regulado, por lo que los científicos suelen limitar el estudio a un subconjunto. Las últimas técnicas de análisis de Microarreglos automatizan ciertos aspectos de la vinculación entre los resultados de perfiles de expresión y la importancia biológica de los mismos, sin embargo éste sigue siendo un problema muy complicado.

Las listas de genes, de longitud relativamente cortas publicadas a partir de experimentos de perfiles de expresión, limitan el grado de concordancia de los experimentos realizados en diferentes laboratorios. La colocación de los resultados de perfiles de expresión en una base de datos de microarreglos de acceso público, permite a los investigadores evaluar los patrones de expresión más allá del alcance de los resultados publicados, pudiendo así identificar alguna similitud con su propio trabajo.

## **Validación de las mediciones de alto rendimiento**

---

Tanto el Chip de ADN y la PCR en tiempo real de ADN aprovechan la unión preferencial o "apareamiento de bases" de secuencias de ácidos nucleicos complementarios, y ambos se utilizan en perfiles de expresión génica a menudo de manera seriada. Mientras que los microarreglos de ADN de alto rendimiento carecen de la precisión cuantitativa de la qPCR (PCR cuantitativa), se tarda aproximadamente el mismo tiempo en medir la expresión génica de unas pocas docenas de genes vía qPCR, que el que emplearía la técnica de microarreglos de ADN en medir todo un genoma. Por lo tanto, a menudo tiene sentido realizar experimentos de análisis de microarreglos de ADN semicuantitativos para identificar los [genes candidatos] y a continuación realizar una qPCR, en algunos de los genes candidatos más interesantes, para validar los resultados de microarreglos. Otros experimentos tales como una Western blot o de algunos de los productos proteicos de los genes diferencialmente expresados, sacan algunas conclusiones basadas en el perfil de expresión más convincente, ya que los niveles de ARNm no se correlacionan necesariamente con la cantidad de proteína expresada.

## **Análisis estadístico**

---

El análisis de datos de microarreglos se ha convertido en un área de intensa investigación.<sup>9</sup> Indica simplemente que un grupo de genes fueron regulados por lo menos dos veces, una sola vez carece de una base estadística sólida. Con cinco repeticiones o menos en cada grupo, típico de microarreglos, una sola observación de un valor atípico puede crear una diferencia aparente mayor del doble. Además, establecer arbitrariamente el margen a una revisión de expresión doble no es biológicamente acertado, ya que no son considerados muchos genes de importancia biológica obvia.

En lugar de identificar los genes expresados diferencialmente utilizando un corte en el cambio en la proporción, se puede utilizar una variedad de pruebas estadísticas o pruebas ómnibus como ANOVA, todos los cuales consideran el "fold change" y la variabilidad para crear un valor p, una estimación de la frecuencia con que se podría observar los datos por pura casualidad. La aplicación de los valores p para microarreglos es complicado por el gran número de comparaciones múltiples (genes) involucrados. Por ejemplo, un valor de p de 0.05 se considera típico para indicar significación, debido a que estima un 5% de probabilidad de que los resultados observados se deban al azar. Sin embargo con 10 000 genes en un microarreglo, 500 genes serían identificados como significativos a valores de  $p < 0.05$ , aunque no hubiera diferencia entre los grupos experimentales. Una solución obvia es considerar importantes solo aquellos genes identificados con criterio más riguroso de valor de p, por ejemplo, se podría realizar una corrección de Bonferroni en los valores de p, o utilizar un cálculo de la tasa de descubrimiento falso para ajustar los valores p en proporción al número de pruebas paralelas implicadas. Desafortunadamente, estas

aproximaciones pueden reducir el número de genes importantes a cero, incluso cuando los genes son de hecho expresados de forma diferencial. Las estadísticas actuales tales como el método no paramétrico Rank Products, tienen como objetivo lograr un equilibrio entre tasa de falsos descubrimientos de genes debido a la variación aleatoria y tasa de no descubrimiento de genes expresados en forma diferencial. Los métodos comúnmente citados incluyen, el análisis de la Importancia de microarreglos (SAM en inglés),<sup>10</sup> una amplia variedad de métodos que están disponibles en el proyecto Bioconductor y una variedad de paquetes de análisis procedentes de empresas de bioinformática.

Al seleccionar una prueba distinta usualmente se identifica una lista diferente de genes importantes,<sup>11</sup> ya que cada prueba opera bajo un conjunto específico de supuestos y pone diferente énfasis en ciertas características de los datos. Muchos ensayos comienzan con la suposición de una distribución normal de los datos, ya que parece un punto de partida razonable y con frecuencia genera resultados que parecen más significativos. Algunas pruebas consideran la distribución conjunta de todas las observaciones de genes para estimar la variabilidad general en las mediciones,<sup>12</sup> mientras que otros miran cada gen de manera aislada. Muchas de las técnicas de análisis de microarreglos modernas implican el procedimiento estadístico bootstrapping (estadística), las técnicas de aprendizaje de máquinas o el método de Montecarlo.<sup>13</sup>

Conforme aumenta el número de mediciones repetidas en un experimento de microarreglos, diversas propuestas estadísticas arrojan resultados cada vez más similares, pero la falta de concordancia entre los diferentes métodos estadísticos hace que los resultados del experimento de microarreglos parezcan menos dignos de confianza. El Proyecto MAQC<sup>14</sup> da recomendaciones para guiar a los investigadores en la selección de los métodos más estándar (por ejemplo, el uso conjunto del valor p y el fold-change para la selección de los genes expresados diferencialmente), para que los experimentos realizados en diferentes laboratorios sean concordantes.

A diferencia del análisis de genes individuales expresados diferencialmente, otro tipo de análisis se centra en la expresión diferencial o perturbación de conjuntos de genes predefinidos y se denomina análisis del conjunto de genes.<sup>15 16</sup> Este análisis ha demostrado tener varias ventajas importantes con respecto al análisis de genes individuales expresados diferencialmente. Conjuntos de genes son grupos de genes que están funcionalmente relacionados según los conocimientos actuales. Por consiguiente, el análisis de conjunto de genes se considera un enfoque de análisis basado en el conocimiento.<sup>15</sup> Los conjuntos de genes utilizados comúnmente incluyen aquellos pertenecientes a las vías KEGG (Enciclopedia de Genes y Genomas de Kyoto), términos de ontología génica, grupos de genes que comparten algunas otras anotaciones funcionales, tales como reguladores de la transcripción de genes comunes, etc. Métodos de análisis representativos de conjuntos de genes incluyen el GSEA<sup>15</sup> (método de análisis de enriquecimiento de conjuntos de genes) que estima la significación de conjuntos de genes basado en la permutación de etiquetas de la muestra, y GAGE<sup>16</sup> que pone a prueba la significación de conjuntos de genes basado en la permutación de etiquetas de genes o una distribución paramétrica.

## **Anotación de genes**

---

Si bien la estadística puede identificar qué productos génicos cambian según las condiciones experimentales, el dar un sentido biológico a los perfiles de expresión se basa en saber qué proteína produce cada producto génico y qué función desempeña dicha proteína. La anotación de genes proporciona información funcional y más información, por ejemplo, la ubicación de cada gen en un cromosoma en particular. Algunas anotaciones de genes son más fidedignas que otras, algunas no están presentes. Las bases de datos de la anotación de genes cambian regularmente y diversas bases de datos se refieren a la misma proteína con nombres diferentes, reflejando alteraciones en la comprensión de la función de la

proteína. El uso de la nomenclatura genética estandarizada ayuda a abordar el aspecto de la nomenclatura del problema en cuestión, aunque la coincidencia exacta de las transcripciones de genes<sup>17 18</sup> sigue siendo un importante aspecto a considerar.

## **Categorización de genes regulados**

---

Una vez identificado un conjunto de genes regulados, el siguiente paso es buscar en el perfil de expresión génica patrones dentro del conjunto regulado. ¿Las proteínas sintetizadas a partir de estos genes realizan funciones similares? ¿Son químicamente similares? ¿Residen en partes similares de la célula? El análisis de ontología génica proporciona una forma estándar de determinar estas relaciones. La ontología génica comienza con categorías muy amplias, por ejemplo “el proceso metabólico”, y las descompone en categorías más pequeñas, por ejemplo “el proceso metabólico de carbohidratos”, y finalmente en categorías muy restrictivas como “inositol y fosforilación de derivados”.

Los genes tienen otros atributos además de la función biológica, las propiedades químicas y la localización celular. Uno puede componer conjuntos de genes basados en la proximidad a otros genes, asociación con una enfermedad y relaciones con drogas y toxinas. La base de datos de firmas moleculares (MSigDB siglas en inglés)<sup>19</sup> y la base de datos comparativa toxicogenómica (CTD)<sup>20</sup> son ejemplos de recursos para clasificar genes de muchas maneras.

## **Búsqueda de patrones entre los genes regulados**

---

Los genes regulados se clasifican en función de lo que son y lo que hacen; pueden surgir relaciones importantes entre los genes.<sup>21</sup> Por ejemplo, podríamos ver la evidencia de que un determinado gen produce una proteína para producir una enzima que activa una proteína para encender un segundo gen en nuestra lista. Este segundo gen puede ser un factor de transcripción que regula aún otro gen de nuestra lista. Si observamos estas relaciones podemos empezar a sospechar que representan mucho más que asociaciones casuales en los resultados y que se encuentran en nuestra lista debido a un proceso biológico subyacente. Por otro lado, podría ser que si uno seleccionara genes al azar, podría encontrar muchos que parecen tener algo en común. En este sentido necesitamos procedimientos rigurosos de estadística para probar que los resultados biológicos obtenidos tienen o no significancia. Aquí es donde el análisis de conjunto de genes<sup>15 16</sup> entra en acción.

## **Relaciones de causa y efecto**

Métodos estadísticos sencillos proporcionan estimaciones para saber si las asociaciones entre los genes de las listas son mayores de lo que cabría esperar por casualidad. Estas estadísticas son de interés incluso si representan una simplificación sustancial de lo que realmente está pasando. He aquí un ejemplo. Supongamos que hay 10 000 genes en un experimento, solamente 50 (0.5%) de los cuales desempeñan un papel conocido en la fabricación de colesterol. El experimento identifica 200 genes regulados. De ellos, 40 (20%) resultan estar también en una lista de genes del colesterol. Sobre la base de la prevalencia global de los genes del colesterol (0.5%), se espera un promedio de 1 gen del colesterol por cada 200 genes regulados, es decir, 0.005 veces 200. Esta expectativa es un promedio, por lo que algunas veces se espera ver más de 1. La pregunta es, con qué frecuencia veríamos 40 en lugar de 1, por pura casualidad.

De acuerdo con la distribución hipergeométrica, se tendría que tratar alrededor de  $10^{57}$  veces (10 seguido de 56 ceros) antes de escoger 39 o más de los genes del colesterol de una piscina de 10 000 sacando 200 genes al azar. Si se presta mucha atención a lo infinitesimalmente pequeña que es la probabilidad de

observar esto por casualidad, se podría concluir que la lista de genes regulados se enriquece en genes<sup>22</sup> con una asociación al colesterol conocida.

Se podría plantear además la hipótesis, que el tratamiento experimental regula el colesterol, ya que el tratamiento parece regular selectivamente los genes asociados con el colesterol. Si bien esto puede ser cierto, existen varias razones por las que, una conclusión firme basada solo en el enriquecimiento, sería un acto de fe injustificado. Una de las cuestiones mencionadas anteriormente tiene que ver con la observación de que la regulación de genes puede no tener un impacto directo en la regulación de proteínas, incluso si las proteínas codificadas por estos genes no hacen otra cosa que no sea recoger el colesterol, la indicación de que su ARNm está alterado no nos dice directamente lo que está sucediendo a nivel de proteínas. Es muy posible que la cantidad de estas proteínas relacionadas con el colesterol se mantenga constante bajo las condiciones experimentales. En segundo lugar, incluso si los niveles de proteína cambian, tal vez hay siempre bastantes alrededor para recoger el colesterol lo más rápido posible, es decir, otra proteína, que no es de nuestra lista, es la reacción limitante del proceso de toma de colesterol. Por último, las proteínas normalmente desempeñan muchos papeles, por lo que estos genes pueden estar regulados no por su asociación compartida con la recogida de colesterol sino por compartir un papel en un proceso totalmente independiente.

Teniendo en mente las advertencias anteriores, si bien los perfiles de genes por sí mismos no prueban relaciones causales entre los tratamientos y los efectos biológicos, sí ofrecen conocimientos biológicos únicos a los que resulta difícil llegar por otros caminos.

## Utilización de patrones para encontrar los genes regulados

Como se describió anteriormente, primero se pueden identificar de modo significativo los genes regulados y luego encontrar patrones comparando la lista de genes significativos con grupos de genes que se sabe comparten ciertas asociaciones. También se puede trabajar en orden inverso. He aquí un ejemplo muy simple. Supongamos que hay 40 genes asociados con un proceso conocido, por ejemplo, una predisposición a la diabetes. Mirando a dos grupos de perfiles de expresión, uno para ratones alimentados con una dieta alta en hidratos de carbono y otro para ratones alimentados con una dieta baja en hidratos de carbono, se observa que los 40 genes de la diabetes se expresan a un nivel más alto en el grupo de dieta alta en carbohidratos que en el grupo de dieta baja en hidratos de carbono. Independientemente de si alguno de estos genes pertenece a una lista de genes alterados significativamente, observar los 40 genes activados y ninguno desactivado, no parece ser un resultado aleatorio: la probabilidad de lanzar una moneda al aire y que salga cara 40 veces seguidas es una en un trillón de intentos.

Para un tipo de célula, el grupo de genes cuyo patrón de expresión combinada es característico únicamente para una determinada condición, constituye la firma genética de esta condición. Idealmente, la firma genética se podría utilizar, en un grupo de pacientes en los que se determina con exactitud el estadio específico de una enfermedad, para facilitar la selección de los tratamientos.<sup>23 24</sup> El análisis de enriquecimiento en grupos de genes (GSEA siglas en inglés)<sup>15</sup> y métodos similares<sup>16</sup> toman ventaja de este tipo de lógica pero utilizan estadísticas más sofisticadas, porque los genes constitutivos en procesos reales muestran un comportamiento más complejo que simplemente un desplazamiento corriente arriba o corriente abajo, y la cantidad de genes moviéndose corriente arriba o corriente abajo es importante, no sólo el sentido. En cualquier caso, estas estadísticas miden qué tan diferente es el comportamiento de un pequeño conjunto de genes comparado con genes que no pertenecen a ese pequeño grupo.

El método GSEA utiliza una prueba de Kolmogórov-Smirnov para ver si alguno de los conjuntos de genes previamente definidos exhibió un comportamiento inusual en el perfil de expresión común. Esto conduce al reto de tener que hacer múltiples pruebas de hipótesis, sin embargo disponemos de métodos para hacerle frente.<sup>25</sup>

## Conclusiones

---

Los perfiles de expresión proporcionan nueva información acerca de lo que hacen los genes bajo diversas condiciones. En general la tecnología de microarreglos produce perfiles de expresión confiables.<sup>26</sup> A partir de esta información se pueden generar nuevas hipótesis sobre biología o probar las ya existentes. Sin embargo, la magnitud y complejidad de estos experimentos a menudo dan lugar a una amplia variedad de posibles interpretaciones. En muchos casos, el análisis de los resultados de los perfiles de expresión requiere de mucho más esfuerzo que la realización de los experimentos.

La mayoría de los investigadores utilizan múltiples métodos estadísticos y análisis exploratorio de datos antes de publicar sus resultados de perfiles de expresión, coordinando sus esfuerzos con expertos en bioinformática o en la tecnología de microarreglos. Un buen diseño experimental, la respuesta biológica adecuada y el seguimiento de los experimentos desempeñan un papel clave en los experimentos de perfiles de expresión exitosos.

## Referencias

---

1. «Microarrays Factsheet» (<https://web.archive.org/web/20071029145451/http://www.ncbi.nlm.nih.gov/About/primer/microarrays.html>). Archivado desde el original (<http://www.ncbi.nlm.nih.gov/About/primer/microarrays.html>) el 29 de octubre de 2007. Consultado el 28 de diciembre de 2007.
2. Kawasaki ES (julio de 2006). «The end of the microarray Tower of Babel: will universal standards lead the way?» (<http://jbt.abrf.org/cgi/pmidlookup?view=long&pmid=16870711>) ( \_ Scholar search ([http://scholar.google.co.uk/scholar?hl=en&lr=&q=intitle%3AThe+end+of+the+microarray+Tower+of+Babel%3A+will+universal+standards+lead+the+way%3F&as\\_publication=J+Biomol+Tech&as\\_ylo=2006&as\\_yhi=2006&btnG=Search](http://scholar.google.co.uk/scholar?hl=en&lr=&q=intitle%3AThe+end+of+the+microarray+Tower+of+Babel%3A+will+universal+standards+lead+the+way%3F&as_publication=J+Biomol+Tech&as_ylo=2006&as_yhi=2006&btnG=Search))). *J Biomol Tech* **17** (3): 200-6. PMC 2291790 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2291790>). PMID 16870711 (<https://www.ncbi.nlm.nih.gov/pubmed/16870711>). (enlace roto disponible en Internet Archive; véase el historial ([https://web.archive.org/web/\\*/http://jbt.abrf.org/cgi/pmidlookup?view=long&pmid=16870711](https://web.archive.org/web/*/http://jbt.abrf.org/cgi/pmidlookup?view=long&pmid=16870711)), la primera versión (<https://web.archive.org/web/1/http://jbt.abrf.org/cgi/pmidlookup?view=long&pmid=16870711>) y la última (<https://web.archive.org/web/2/http://jbt.abrf.org/cgi/pmidlookup?view=long&pmid=16870711>)).
3. Suter L, Babiss LE, Wheeldon EB (2004). «Toxicogenomics in predictive toxicology in drug development». *Chem. Biol.* **11** (2): 161-71. PMID 15123278 (<https://www.ncbi.nlm.nih.gov/pubmed/15123278>). doi:10.1016/j.chembiol.2004.02.003 (<https://dx.doi.org/10.1016%2Fj.chembiol.2004.02.003>).
4. Magic Z, Radulovic S, Brankovic-Magic M (2007). «cDNA microarrays: identification of gene signatures and their application in clinical practice». *J BUON.* 12 Suppl 1: S39-44. PMID 17935276 (<https://www.ncbi.nlm.nih.gov/pubmed/17935276>).
5. Cheung AN (2007). «Molecular targets in gynaecological cancers». *Pathology* **39** (1): 26-45. PMID 17365821 (<https://www.ncbi.nlm.nih.gov/pubmed/17365821>). doi:10.1080/00313020601153273 (<https://dx.doi.org/10.1080%2F00313020601153273>).
6. Mirza SP, Olivier M (2007). «Methods and approaches for the comprehensive characterization and quantification of cellular proteomes using mass spectrometry» (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2771641>). *Physiol Genomics* **33** (1): 3-11. PMC 2771641 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2771641>). PMID 18162499 (<https://www.ncbi.nlm.nih.gov/pubmed/18162499>).

- doi:10.1152/physiolgenomics.00292.2007 (<https://dx.doi.org/10.1152%2Fphysiolgenomics.00292.2007>).
7. Chen JJ (2007). «Key aspects of analyzing microarray gene-expression data». *Pharmacogenomics* **8** (5): 473-82. PMID 17465711 (<https://www.ncbi.nlm.nih.gov/pubmed/17465711>). doi:10.2217/14622416.8.5.473 (<https://dx.doi.org/10.2217%2F14622416.8.5.473>).
  8. Jaskowiak, Pablo A; Campello, Ricardo JGB; Costa, Ivan G (24 de enero de 2014). «On the selection of appropriate distances for gene expression data clustering». *BMC Bioinformatics* **15** (Suppl 2): S2. doi:10.1186/1471-2105-15-S2-S2 (<https://dx.doi.org/10.1186%2F1471-2105-15-S2-S2>).
  9. Vardhanabhuti S, Blakemore SJ, Clark SM, Ghosh S, Stephens RJ, Rajagopalan D (2006). «A comparison of statistical tests for detecting differential expression using Affymetrix oligonucleotide microarrays» ([https://archive.org/details/sim\\_scandinavian-journal-of-economics\\_2006-12\\_108\\_4/page/555](https://archive.org/details/sim_scandinavian-journal-of-economics_2006-12_108_4/page/555)). *OMICS* **10** (4): 555-66. PMID 17233564 (<https://www.ncbi.nlm.nih.gov/pubmed/17233564>). doi:10.1089/omi.2006.10.555 (<https://dx.doi.org/10.1089%2Fomi.2006.10.555>).
  10. «Significance Analysis of Microarrays» (<http://www-stat.stanford.edu/~tibs/SAM/>). Consultado el 27 de diciembre de 2007.
  11. Yauk CL, Berndt ML (2007). «Review of the literature examining the correlation among DNA microarray technologies» (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2682332>). *Environ. Mol. Mutagen.* **48** (5): 380-94. PMC 2682332 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2682332>). PMID 17370338 (<https://www.ncbi.nlm.nih.gov/pubmed/17370338>). doi:10.1002/em.20290 (<https://dx.doi.org/10.1002%2Fem.20290>).
  12. Breitling R (2006). «Biological microarray interpretation: the rules of engagement». *Biochim. Biophys. Acta* **1759** (7): 319-27. PMID 16904203 (<https://www.ncbi.nlm.nih.gov/pubmed/16904203>). doi:10.1016/j.bbaexp.2006.06.003 (<https://dx.doi.org/10.1016%2Fj.bbaexp.2006.06.003>).
  13. Draminski M, Rada-Iglesias A, Enroth S, Wadelius C, Koronacki J, Komorowski J (2008). «Monte Carlo feature selection for supervised classification». *Bioinformatics* **24** (1): 110-7. PMID 18048398 (<https://www.ncbi.nlm.nih.gov/pubmed/18048398>). doi:10.1093/bioinformatics/btm486 (<https://dx.doi.org/10.1093%2Fbioinformatics%2Fbtm486>).
  14. Dr. Leming Shi, National Center for Toxicological Research. «MicroArray Quality Control (MAQC) Project» (<http://www.fda.gov/nctr/science/centers/toxicoinformatics/maqc/>). U.S. Food and Drug Administration. Consultado el 26 de diciembre de 2007.
  15. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP (2005). «Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles» (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1239896>). *Proc. Natl. Acad. Sci. U.S.A.* **102** (43): 15545-50. PMC 1239896 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1239896>). PMID 16199517 (<https://www.ncbi.nlm.nih.gov/pubmed/16199517>). doi:10.1073/pnas.0506580102 (<https://dx.doi.org/10.1073%2Fpnas.0506580102>).
  16. Luo W, Friedman M, Shedden K, Hankenson KD, Woolf JP (2009). «GAGE: generally applicable gene set enrichment for pathway analysis» (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2696452>). *BMC Bioinformatics* **10**: 161. PMC 2696452 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2696452>). PMID 19473525 (<https://www.ncbi.nlm.nih.gov/pubmed/19473525>). doi:10.1186/1471-2105-10-161 (<https://dx.doi.org/10.1186%2F1471-2105-10-161>).
  17. Dai M, Wang P, Boyd AD, et al. (2005). «Evolving gene/transcript definitions significantly alter the interpretation of GeneChip data» (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1283542>). *Nucleic Acids Res.* **33** (20): e175. PMC 1283542 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1283542>). PMID 16284200 (<https://www.ncbi.nlm.nih.gov/pubmed/16284200>). doi:10.1093/nar/gni179 (<https://dx.doi.org/10.1093%2Fnar%2Fgni179>).
  18. Alberts R, Terpstra P, Hardonk M, et al. (2007). «A verification protocol for the probe sequences of Affymetrix genome arrays reveals high probe accuracy for studies in mouse, human and rat» (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1865557>). *BMC Bioinformatics* **8**: 132. PMC 1865557 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1865557>). PMID 17448222 (<https://www.ncbi.nlm.nih.gov/pubmed/17448222>). doi:10.1186/1471-2105-8-132 (<https://dx.doi.org/10.1186%2F1471-2105-8-132>).
  19. «GSEA – MSigDB» (<http://www.broad.mit.edu/gsea/msigdb/index.jsp>). Consultado el 3



- de enero de 2008.
20. «CTD: The Comparative Toxicogenomics Database» (<http://ctdbase.org/>). Consultado el 3 de enero de 2008.
  21. Alekseev OM, Richardson RT, Alekseev O, O'Rand MG (2009). «Analysis of gene expression profiles in HeLa cells in response to overexpression or siRNA-mediated depletion of NASP» (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2686705>). *Reprod. Biol. Endocrinol.* **7**: 45. PMC 2686705 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2686705>). PMID 19439102 (<https://www.ncbi.nlm.nih.gov/pubmed/19439102>). doi:10.1186/1477-7827-7-45 (<https://dx.doi.org/10.1186%2F1477-7827-7-45>).
  22. Curtis RK, Oresic M, Vidal-Puig A (2005). «Pathways to the analysis of microarray data» ([https://archive.org/details/sim\\_trends-in-biotechnology\\_2005-08\\_23\\_8/page/429](https://archive.org/details/sim_trends-in-biotechnology_2005-08_23_8/page/429)). *Trends Biotechnol.* **23** (8): 429-35. PMID 15950303 (<https://www.ncbi.nlm.nih.gov/pubmed/15950303>). doi:10.1016/j.tibtech.2005.05.011 (<https://dx.doi.org/10.1016%2Fj.tibtech.2005.05.011>).
  23. Mook S, Van't Veer LJ, Rutgers EJ, Piccart-Gebhart MJ, Cardoso F (2007). «Individualization of therapy using Mammaprint: from development to the MINDACT Trial». *Cancer Genomics Proteomics* **4** (3): 147-55. PMID 17878518 (<https://www.ncbi.nlm.nih.gov/pubmed/17878518>).
  24. Corsello SM, Roti G, Ross KN, Chow KT, Galinsky I, DeAngelo DJ, Stone RM, Kung AL, Golub TR, Stegmaier K (junio de 2009). «Identification of AML1-ETO modulators by chemical genomics» (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2699238>). *Blood* **113** (24): 6193-205. PMC 2699238 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2699238>). PMID 19377049 (<https://www.ncbi.nlm.nih.gov/pubmed/19377049>). doi:10.1182/blood-2008-07-166090 (<https://dx.doi.org/10.1182%2Fblood-2008-07-166090>).
  25. «GSEA» (<http://www.broad.mit.edu/gsea/migdb/index.jsp>). Consultado el 9 de enero de 2008.
  26. Couzin J (2006). «Genomics. Microarray data reproduced, but some concerns remain». *Science* **313** (5793): 1559. PMID 16973852 (<https://www.ncbi.nlm.nih.gov/pubmed/16973852>). doi:10.1126/science.313.5793.1559a (<https://dx.doi.org/10.1126%2Fscience.313.5793.1559a>).

---

Obtenido de «[https://es.wikipedia.org/w/index.php?title=Perfil de expresión génica&oldid=147393922](https://es.wikipedia.org/w/index.php?title=Perfil_de_expresi%C3%B3n_g%C3%A9nica&oldid=147393922)»

■