

Aplicaciones de la técnica de *microarrays* en ciencias biomédicas: presente y futuro

Dra. Paola Roxana Barrero

Investigadora Asistente CONICET
Laboratorio de Virología, Hospital de Niños "Dr. Ricardo Gutiérrez"
Gallo 1330 piso 2 C1425EDF Buenos Aires Argentina
paobarrero@fibertel.com.ar

Recibido: 28/10/2005

Aceptado: 2/12/2005

En esta era donde la globalización ocupa las primeras páginas de los medios de comunicación masivos, donde la tecnología y las comunicaciones son pilares de la economía mundial, la ciencia tiene sus propios desafíos. Tanto la globalización como la tecnificación han producido un cambio esencial en el método científico, la relación "un experimento, un dato" deja de ser obvia, y el resultado final de los nuevos experimentos es un caudal de datos que deben procesarse minuciosamente para extraer *la información* útil al científico. Desde que se completó el proyecto genoma humano, las preguntas se han sucedido en cuanto a identificar las funciones, regulación e interacción de genes en procesos celulares y cambios en la expresión génica debidos a enfermedad o tratamiento. Una década atrás pensar en un experimento que abarcara la detección de cada transcripto en una célula hubiera parecido extraído de una película de ciencia ficción. Hoy es posible gracias a que múltiples disciplinas aunaron esfuerzos para integrarse en el desarrollo, diseño y análisis de *microarrays*. Pero ¿qué es un *microarray*? ¿Qué preguntas pueden contestar estos experimentos?

Espero que luego de leer este artículo tenga el lector una idea más clara y una perspectiva de los alcances de esta nueva tecnología.

Introducción

Los genes consisten en porciones de ácido desoxirribonucleico (DNA) que codifican por una proteína en particular. Cada segmento de DNA genómico, organizado en cromosomas en el núcleo celular, está construido con cuatro bloques de nucleótidos constituido por un grupo fosfato, un azúcar desoxirribosa y una de las cuatro bases nitrogenadas: Adenina (A), guanina (G), citosina (C) y timina (T). Durante la síntesis de RNA (transcripción) la doble cadena de DNA se despliega y una de ellas es usada como molde para generar una copia complementaria de cadena simple con la base uracilo que reemplaza a la timina. El RNA es procesado luego de manera que a partir de un transcripto primario se generan diferentes clases, entre ellas el mRNA que en eucariotas lleva un cap 5', una cola de poli-A y no contiene intrones. En el proceso siguiente de traducción, se sintetizan proteínas a partir del mRNA en los ribosomas. La correspondencia entre las cuatro letras presentes en el DNA y los 20 aminoácidos que forman las proteínas está dada por el código genético, que vincula tres bases (codón) a un aminoácido. La proteína sufre diferentes modificaciones post-traducción hasta ser completamente funcional. De esta manera, las características estructurales y funcionales de células y tejidos están determinadas por la expresión simultánea, selectiva y diferencial de miles de genes (Figura 1).

Una propiedad importante del DNA es la complementariedad de las bases que lo componen, así la timina se complementa con la adenina y la guanina con la citosina por medio de puentes hidrógeno de las cadenas opuestas del DNA. Esta propiedad es la base molecular de la técnica de *microarrays*, donde hebras de DNA de diferente origen van a hibridizarse por complementariedad de las bases que las componen.

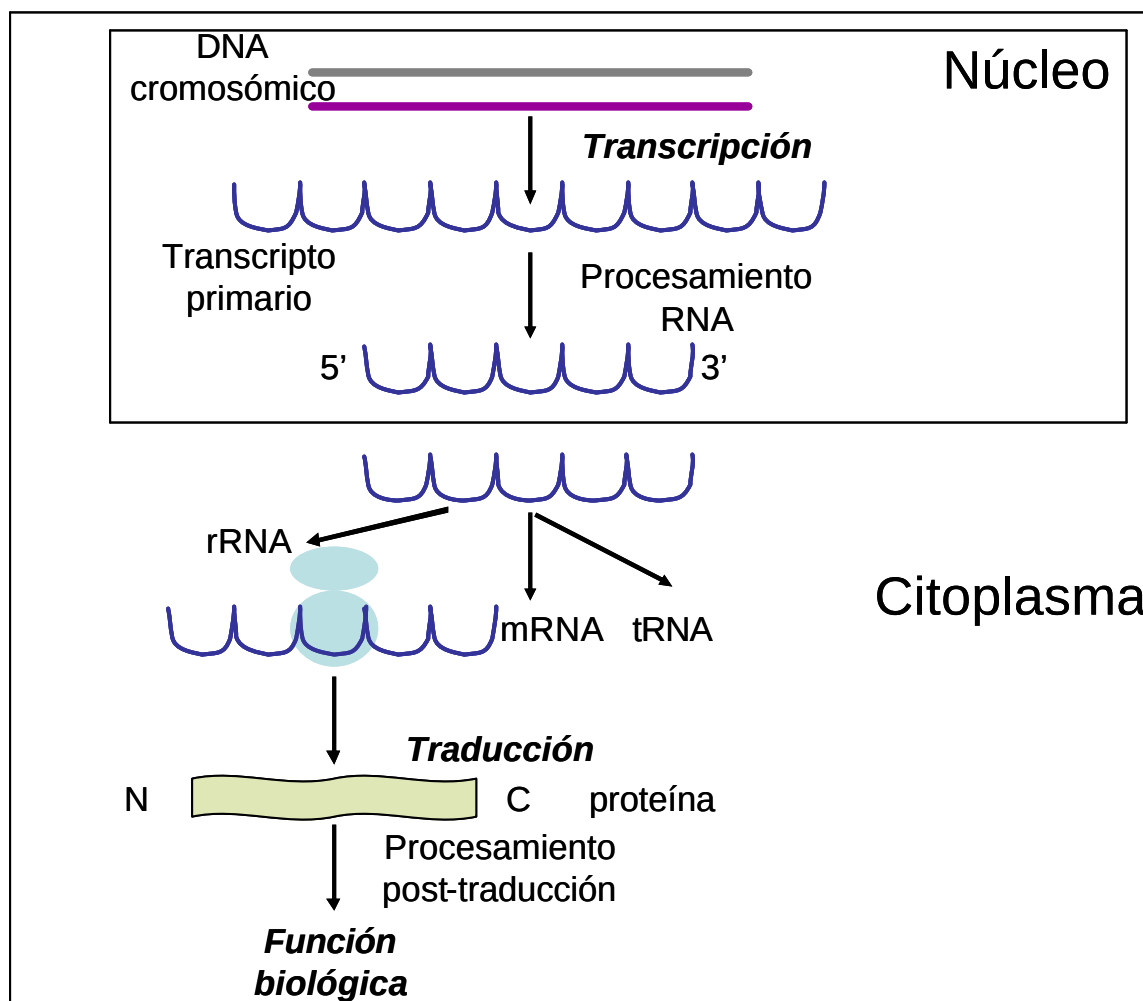


Figura 1: Dogma central de la biología molecular: el material genético consiste de DNA, que se transcribe a mRNA que sirve como molde para la síntesis de proteínas. mRNA: RNA mensajero; rRNA: RNA ribosomal; tRNA: RNA de transferencia.

El fundamento de la tecnología empleada en *microarrays*

La metodología clásica utilizada para la detección y cuantificación de mRNA presente en una célula es el análisis por *Northern blot* donde una sonda radioactiva en solución se une con un mRNA inmovilizado en un soporte. Otro método usado es la RT-PCR, donde el mRNA es copiado a cDNA (DNA complementario) y se genera un intermediario de doble cadena por la reacción de transcripción reversa (RT) y posterior amplificación de la cadena por la reacción de la polimerasa (PCR). En estos ensayos el principio fundamental es la hibridación, que se basa en la complementariedad de las bases, pero su debilidad es la detección. En el primer caso la detección de emisión de radioactividad requiere pasos adicionales como la autorradiografía y en el mejor de los casos el escaneo con detectores de radiaciones β y γ que aumentan hasta 100 veces la sensibilidad de la técnica. En el caso de la RT-PCR los niveles de detección han aumentado significativamente mostrando en tiempo real la duplicación exacta del material inicial en la reacción de PCR por la incorporación de fluoróforos

a la doble cadena naciente. Esto permite detectar cambios de dos órdenes de magnitud comparado con los geles de bromuro de etidio que manifiestan cambios de diez órdenes de magnitud.

Estas técnicas suplieron las necesidades durante muchos años pero sufren la limitación del número de genes a reportar. Actualmente se requieren de reportes globales o de miles de genes a la vez, imposibles de detectar por radioactividad, debido a que estos ensayos soportan una cierta densidad de puntos en simultáneo, o por PCR en tiempo real, dadas las limitaciones de realizar los ensayos de 96 muestras cada vez (una placa).

La palabra *microarray* deriva del griego *mikro* (pequeño) y del inglés *array* (distribución ordenada). Podríamos decir que las micromatrices o microarreglos permiten el depósito de miles de puntos conteniendo genes o parte de genes sobre un portaobjetos para su estudio en paralelo. De esta manera es posible tener una visión instantánea de actividad de genomas completos o de un grupo selecto de genes.

En los estudios de *microarrays* se combinan las técnicas de hibridización de ácidos nucleicos y detección por fluorescencia. De esta manera, sólo en los puntos del portaobjeto donde haya ocurrido hibridización habrá fluorescencia y la intensidad de la fluorescencia detectada será proporcional al nivel de expresión del gen en estudio.

Una condición indispensable es que cada uno de los genes que esté representado sea fácilmente distinguible de otros. En otras palabras la porción del gen inmovilizada en el portaobjeto debe llevar consigo, independientemente de su tamaño, su cédula de identidad. Este punto es de especial importancia en el diseño de los *microarrays* y se basa en búsquedas exhaustivas en las bases de datos públicas (como *Unigene* y *GenBank*) y selección por ensayos de prueba y error.

Componentes de un experimento de *microarrays*

Dada una pregunta biológica o un planteo de sondeo global se obtiene una muestra que será marcada para su posterior detección. Esta muestra marcada es enfrentada al *microarray* donde ocurre la hibridización. Un seguimiento exhaustivo por análisis de imágenes detectará los puntos marcados y los remitirá al correspondiente gen. Luego, por uso de algoritmos de interpretación se extraerán los datos más relevantes según la hipótesis o se describirán los aspectos que marcan diferencias significativas en experimentos de exploración (Figura 2). Este esquema básico refleja, en cada uno de sus componentes, decisiones importantes que el investigador debe tomar y que no limitan el método a una hazaña biotecnológica sino al resultado de un arduo entrenamiento técnico y del trabajo en equipo. Comentaremos sus componentes a continuación.

Muestras biológicas

El primer componente que introduce variación al experimento es la muestra en sí. El éxito de un ensayo de *microarrays* depende en gran parte de la calidad de la muestra obtenida. El RNA total debe estar intacto y puro, no debe contener proteínas ni DNA contaminante. Para ello se exige que sea evaluada su cantidad y calidad por espectrofotometría y su integridad por electroforesis (deben observarse las bandas de rRNA 28S y 18S en proporción y peso molecular adecuado).

Sin hacer mención de los aspectos éticos subyacentes, definir patrones universales con experimentos costosos como los de *microarrays* es una cuestión que no tiene solución aún. La heterogeneidad que presentan líneas celulares inmortalizadas es mínima ante la que pueda presentar una biopsia tomada de un tejido tumoral. Es imposible tener dos muestras biológicas del mismo individuo que den resultados similares por el solo hecho de haberlas manipulado en eventos independientes. A su vez, las comparaciones suelen hacerse contra tejidos normales, pero no existen consensos para las definiciones de normal y afectado, dado que factores como los antecedentes genéticos, sexo, edad, etapa de desarrollo, diferenciación y crecimiento dirigen los patrones de expresión. Tampoco debemos olvidarnos que los tejidos son de diversa composición celular y que cada una de estas células activará diferentes programas de expresión ante diferentes estímulos.

Se ha planteado el uso de mRNA "universales" o de especies diferentes que funcionen como controles internos del ensayo. Estos controles se procesan en conjunto con la muestra y contienen en el *microarray* sondas específicas que monitorean la calidad del ensayo. De esta manera, se comprobó por diluciones seriadas que los más modernos *microarrays* pueden detectar hasta un transcripto (1 molécula aislada de mRNA). Este es un paso importante que da confiabilidad a la técnica, probando que hasta en el nivel más bajo de transcripción, los transcriptos más raros y menos abundantes serán tomados en cuenta para el análisis.

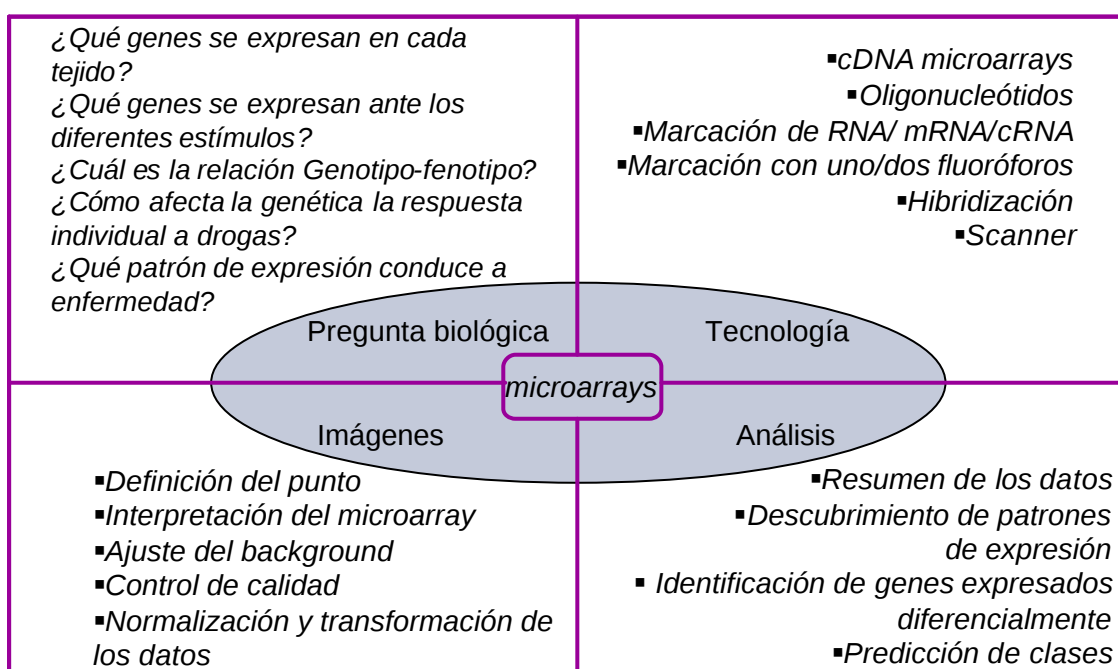


Tabla 2 :Puntos importantes en cada paso de un experimento de *microarrays*

Diseño experimental

El diseño experimental es esencial en los experimentos de *microarrays* y es crucial desde la recolección de muestras y criterios de inclusión, la elección del *microarray* a ensayar y el método para inferir los resultados. En general las preguntas biológicas conducen a probar hipótesis o a producir datos nuevos mediante asociación con parámetros estadísticos. En el primer caso los resultados

obtenidos corroboran la hipótesis planteada y en el segundo los datos hablan por sí mismos. Ambas estrategias son válidas si las herramientas usadas para convertir los datos en información y la información en conocimiento son robustas.

En muestras patológicas se ha propuesto trabajar con mezclas provenientes de diferentes pacientes afectados para reducir el efecto "personal" en el patrón de expresión. También es indispensable realizar duplicaciones biológicas y técnicas del experimento. Un duplicado biológico es aquel donde se obtiene RNA en dos eventos diferentes o RNA de individuos diferentes que se integran en el análisis. Los duplicados técnicos son las hibridaciones de un mismo RNA en diferentes *microarrays* para evaluar la reproducibilidad de los resultados obtenidos. De todas maneras, estas repeticiones, aunque mejoran la certeza de los datos obtenidos, no solucionan el problema del tamaño de la muestra en *microarrays*.

Tipos de *microarrays*

El concepto básico en *microarrays* es el posicionamiento preciso en un soporte sólido de elementos que funcionen como detectores moleculares en altas densidades. En la práctica, los *microarrays* abarcan una amplia gama que puede tener diferentes soportes (membranas o vidrio) y diferentes moléculas que interaccionan en este medio.

Para poder clasificar los *microarrays* debemos dejar claro qué se entiende por *target* (blanco) y que se define como *probe* (sonda). Existen controversias al respecto pero tomaremos por blanco a la molécula libre y como sonda a la inmovilizada.

Existen diversos tipos de *microarrays* según las sondas utilizadas que abarcan metodologías muy variadas que van desde lo más casero hasta lo más sofisticado. Los tipos más comunes son:

Microarrays de ácidos nucleicos

Microarrays de cDNA

Las sondas son producidas en laboratorios mediante la amplificación selectiva de cDNAs (100-3000 nucleótidos) por PCR en placas de 96 pocillos. Estos amplicones se purifican, se verifica su calidad y cantidad y se depositan por capilaridad en portaobjetos de vidrio mediante costosos robots que requieren un ambiente libre de partículas.

Microarrays de oligonucleótidos

Las sondas son porciones de DNA sintético de cadena simple que pueden ser cortas (15-25 nucleótidos) o largas (50-120 nucleótidos). Estos fragmentos pueden ser presintetizados y depositados en portaobjetos por robots o sintetizados *in situ* y depositados por *ink jet* o fotolitografía (DNAchips).

Los *microarrays* que contienen fragmentos presintetizados (cDNA u oligos) pueden ser fabricados en laboratorios con infraestructura adecuada, pero los sintetizados *in situ* o los que vienen con genes preseleccionados prearreglados (*bioarrays*) deben ser adquiridos a diferentes proveedores que poseen plantas con un nivel más alto de complejidad y delicados controles de calidad. Los *arrays* de cDNA son los más flexibles y usados en investigación porque permiten depositar genes o fragmentos de genes amplificados de cualquier especie y así diseñar y generar de manera sencilla y

menos costosa el grupo de sondas. Por otro lado, requiere de réplicas tanto en el mismo soporte como duplicados técnicos, dado que su punto débil es el depósito del amplicón y la reproducibilidad de sus características físicas (dimensiones, área, límites). La tendencia actual es usar oligonucleótidos de longitud corta, aunque todavía existen quienes cuestionan la especificidad dada por el diseño de sondas tan pequeñas, evidentemente, años de experiencia en diseño de PCRs avalan las ventajas de esta opción.

Microarrays de proteínas

Las sondas son anticuerpos fijados a portaobjetos de vidrio y los blancos son muestras de suero o tejido. Esta técnica se ve por el momento restringida por varios puntos que requieren de tiempo para esclarecerse. Entre ellos podemos mencionar la dificultad de fabricar e inmovilizar estructuras 3D como son las proteínas y detectar interacciones de proteínas plegadas, sin olvidar mencionar que no se dispone aún de colorantes fluorescentes que permitan cuantificar eficientemente a estas moléculas.

Microarrays de tejidos (TMA)

Esta técnica trata de resolver uno de los problemas principales y limitantes en análisis moleculares de tejidos: el tamaño limitado de la muestra. Se utiliza una aguja hueca par tomar muestras milimétricas de las regiones de interés de tejidos embebidos en parafina, en especial biopsias. Luego se depositan de manera ordenada en un nuevo bloque de parafina y se cortan con un micrótopo entre 100-500 veces y se reordenan sobre portaobjetos de vidrio donde se realizarán pruebas múltiples a nivel DNA, RNA y proteínas (inmunohistoquímica, hibridización *in situ*).

Comparative Genomic Hybridization (CGH)

Es un método citogenético molecular que permite monitorear anomalías cromosómicas. Las alteraciones se clasifican en pérdidas, ganancias y amplificaciones de DNA, incluyendo mutaciones a nivel de cromosomas completos y por *loci*. Permite monitorear tumores y defectos congénitos a partir de cromosomas en metafase o DNA genómico. La técnica se basa en la hibridización competitiva donde se colorea el DNA tumoral con un marcador fluorescente y el DNA blanco con otro. Permite el estudio de material de archivo como muestras congeladas o embebidas en parafina con el fin de correlacionar la evolución clínica con aberraciones cromosómicas.

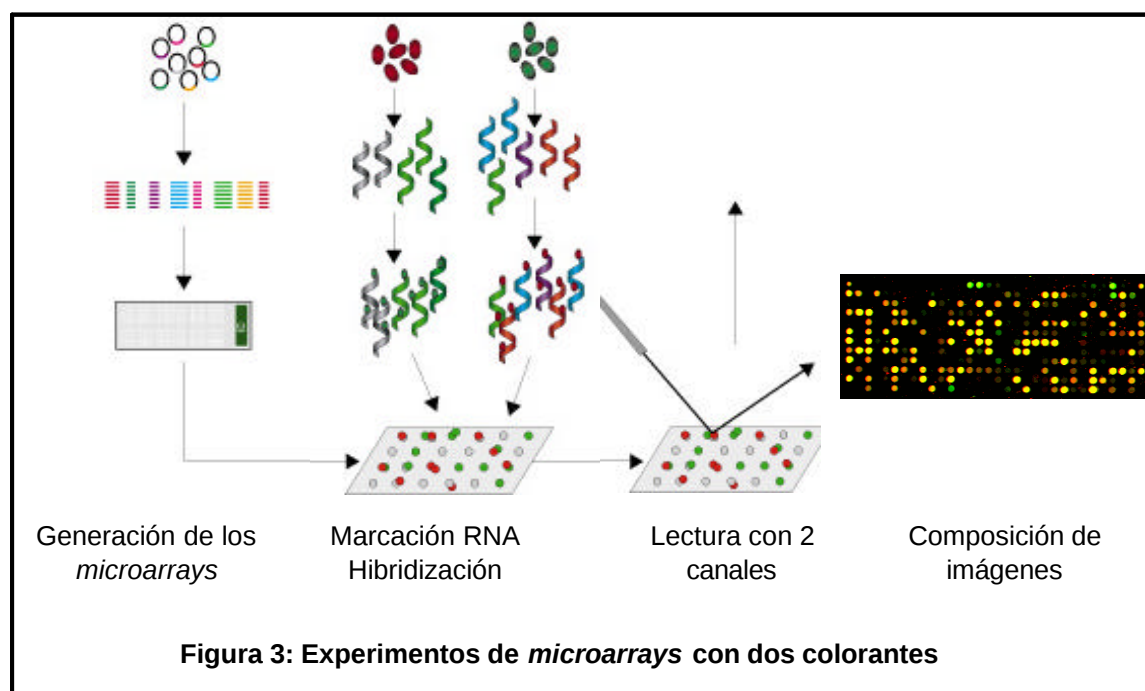
La descripción técnica de los aspectos analíticos de estos últimos tres puntos excede el objetivo de esta revisión y en adelante nos centraremos en los arreglos más usados que son los microarreglos de DNA.

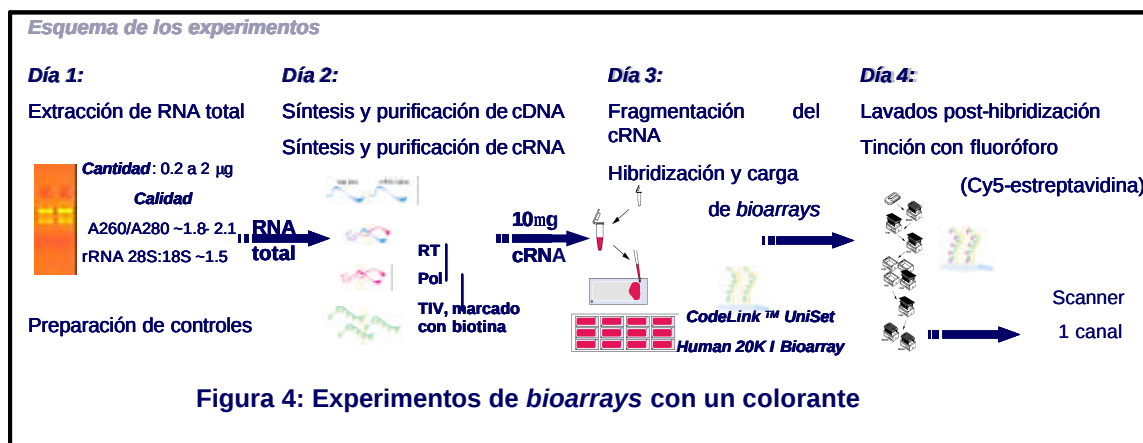
Consideraciones técnicas

Marcación e hibridización del blanco

En los experimentos de expresión génica el RNA total es obtenido de la muestra biológica y marcado con un colorante fluorescente para su posterior detección. Las primeras propuestas fueron marcar con dos colorantes derivados de cianinas (Cy5 rojo y Cy3 verde) las muestras de RNA total durante el proceso de transcripción reversa o modificar químicamente los nucleótidos *a posteriori*. Estas muestras provenientes de dos condiciones a comparar se mezclan en cantidades iguales y se

hibridizan competitivamente en el mismo *microarray*. La lectura se hace con detectores que permitan detectar los espectros de emisión de los dos colorantes en canales diferentes y generar imágenes separadas para cada uno de ellos. Los inconvenientes que plantea este procedimiento son las masas abundantes necesarias de RNA (20-75 μg) y la afinidad diferente de los distintos colorantes por el blanco. Esto hace imperativo que en el diseño se incluyan experimentos donde las muestras sean marcadas adicionalmente con el colorante opuesto para obtener resultados más confiables. Una complejidad adicional se plantea en el análisis de las imágenes, ya que para medir expresión diferencial deben componerse en un solo archivo de imágenes la superposición resultante del archivo generado con el colorante verde y el archivo generado con el colorante rojo (Figura 3). De esta manera los puntos que resulten verdes estarán expresados diferencialmente en una condición y los rojos en otra, mientras que los amarillos estarán expresados en ambas (Figura 3). Estos inconvenientes son superados cuando se trabaja con protocolos más modernos con RNA copia (cRNA) que resulta de la transcripción *in vitro* del RNA original. Las cantidades requeridas para realizar el experimento son ínfimas (0.2-2 μg) y se marca con un solo colorante. La desventaja de este método es que se hibridiza una muestra por *microarray*. En cuanto a la hibridización, las técnicas y protocolos han mejorado radicalmente con procedimientos que tienden a reemplazar los hornos de hibridización y posteriores lavados manuales por condiciones más estandarizadas y cámaras de hibridización flexibles incluidas sobre el portaobjeto de vidrio (Figura 4)





Análisis de los datos

Una vez obtenido el/los archivos de imágenes, hay que transformar las intensidades de las señales obtenidas en datos numéricos, discriminando la señal informativa del ruido que pudiera haber en segundo plano. En este proceso hay que considerar las dimensiones y forma de cada punto analizado de la imagen, su localización y los parámetros estadísticos que pueden asociarse a ellos. Los controles de calidad son realizados para cada imagen.

Es en este punto donde el trabajo multidisciplinario se vuelve indispensable. El manejo de una cantidad extensa de datos en simultáneo requiere el uso de algoritmos de computación para obtener, manejar, procesar y almacenar la información en paralelo. En este escenario los bioinformáticos aportan su conocimiento a los experimentos de *microarrays*.

De esta manera se obtiene una matriz de expresión donde las filas serán genes y las columnas experimentos. En los diseños con dos colorantes se trabaja con proporciones entre las intensidades medidas con cada uno de ellos y en los realizados con un solo colorante estos datos son absolutos. En ambos casos los datos se normalizan y transforman para disminuir las variaciones y hacer los cálculos posteriores más sencillos.

En este punto agregamos al grupo de trabajo un experto en bioestadística que pueda aplicar los supuestos correspondientes para realizar comparaciones que sean válidas estadísticamente. Para un gen dado (fila) pueden compararse las intensidades entre muestras y generar un reporte que exprese encendido y apagado de genes o de cuántas veces más o menos expresado se encuentra en las diferentes condiciones ensayadas. No deben compararse intensidades entre genes (filas) del mismo experimento ya que el nivel de expresión es una propiedad de cada gen, se puede modificar por la expresión de otro gen presente en altos niveles en el mismo experimento y está ligado a complejas vías de control.

Métodos de Agrupamiento y Visualización de los datos

El análisis de agrupamiento o *Clustering* de la matriz de expresión consiste en reunir genes basándose en la similitud de su perfil de expresión

Existen métodos no supervisados y supervisados basándose en datos previos para concentrar los patrones de expresión relacionados. Entre los métodos no supervisados el más empleado es "k-media" y es apto para organizar datos exploratorios exhaustivos. "K-media" es un algoritmo de partición que divide los ítems en k-grupos de manera que la suma de las distancias al

centro del grupo sea mínima. Por otro lado los métodos supervisados requieren un grupo de experimentos que los entrene (*training set*) para generar reglas que puedan hacer predicciones o clasificar datos a testear (*testing set*). Entre ellos podemos mencionar a los basados en redes neuronales de aprendizaje. Los esquemas resultantes son de fácil visualización y altamente informativos.

Interpretación de los datos: *data mining*

Existen diferentes algoritmos que permiten extraer de un grupo selecto de genes su ontología (GO). La ontología de los genes proporciona un vocabulario controlado para describir características de genes y productos génicos en términos de los procesos biológicos asociados, los componentes celulares y la función molecular de manera independiente de la especie en cuestión. Los principios de organización del GO son que un gen tiene una o más funciones moleculares, usadas en uno o más procesos biológicos y puede asociarse a uno o más componentes de la célula.

A partir del conocimiento de los términos de GO y de atribución de éstos a los datos de *microarrays* se pueden visualizar las vías metabólicas que están siendo modificadas.

Reporte de los datos: se creó un consorcio que colecta los datos mínimos acerca de un experimento de *microarrays* (MIAME) de manera que el formato para importar y exportar datos entre laboratorios sea compatible. Este campo es crítico y las publicaciones requieren que se cumplan estos puntos.

Las herramientas contenidas el GEO (*Gene Expression Omnibus*) permiten visualizar, buscar y obtener datos sobre expresión génica.

Reflexión final

Cada célula contiene una dotación completa de cromosomas. Sin embargo, la expresión diferencial de los genes es la que dará a esa célula su función biológica. Este proceso de expresión génica es muy complejo en cuanto a regulación y permite a la célula responder de manera dinámica ante cambios instantáneos. Los experimentos de *microarrays* permiten monitorear en cada momento la expresión génica durante una enfermedad, los cambios en tejidos tumorales, la expresión de marcadores relacionados con factores pronósticos de enfermedad o toxicidad, y los polimorfismos de genes, entre otras tantas aplicaciones.

La magnitud creciente de información y el desarrollo de técnicas innovadoras provee hoy al investigador de herramientas poderosas, flexibles, no tóxicas y de alto rendimiento. Esta escala ampliada, tanto en la cantidad como en la calidad de los datos obtenidos en experimentos de *microarrays* plantea una cuestión adicional no sólo porque el volumen de datos que deben ser procesados es muy elevado, sino también porque muchos de dichos datos son obtenidos de manera colateral, sin una hipótesis previa que guíe el experimento, o involucran genes que hoy no tienen función asignada aún para organismos modelo. La información global no es ni buena ni mala en sí misma, es un instrumento que, adecuadamente utilizado, permitirá alcanzar mayores niveles de conocimiento.

La evolución de esta tecnología de avanzada en el tiempo llevó a una disminución de los costos asociados y de los requerimientos de infraestructura, haciendo hoy factible su aplicación en empresas farmacéuticas y en un futuro no muy lejano en centros biomédicos especializados. La implementación de estudios de expresión génica en la población marcará el camino hacia una medicina personalizada donde las estrategias de diagnóstico y monitoreo del tratamiento se basarán en la evidencia aportada por experimentos de *microarrays*.

Responsabilidad: los datos y opiniones presentados en esta revisión reflejan mi punto de vista que ha sido elaborado durante cinco años de lectura de bibliografía pertinente, discusión con expertos y entrenamiento en los diferentes aspectos de los ensayos de *microarrays*.

Referencias y Sitios de interés en Internet

Nature Genetics (1999), Volume 21 No 1s: The chipping forecast I

Nature Genetics (2002), Volume 32 No 4s: The chipping forecast II.

Nature Medicine (2003) Volume 9 No 1: 140-145

Nature Genetics (2005) Volume 37 No 6s: The chipping forecast III

The Gene Expression Omnibus (GEO): A Gene Expression and Hybridization Repository.

<http://www.ncbi.nlm.nih.gov/geo/>

Microarrays: chipping away at the mysteries of science and medicine.

<http://www.ncbi.nlm.nih.gov/About/primer/microarrays.html#microarrays>

The 7 keys to successful microarray data analysis. <http://www.microarraysuccess.com/web/info.html>



ISSN 1666-7948
www.quimicaviva.qb.fcen.uba.ar

Revista **QuímicaViva**
Número 3, año 4, diciembre 2005
quimicaviva@qb.fcen.uba.ar