

Tecnología, análisis y aplicaciones de los Bio-arreglos

✉ Isabel Guillén,¹ Erik Torres,² Julio Raúl Fernández¹

Centro de Ingeniería Genética y Biotecnología. Subdirección de Investigaciones Biomédicas.

¹División de Farmacéuticos, ²División de Química-Física. E-mail: isabel.guillen@cigb.edu.cu

RESUMEN

Los bio-arreglos son una tecnología versátil y de amplias posibilidades para el avance científico en el siglo XXI. Su aplicación se ha acrecentado enormemente en los últimos años y con ello los análisis genéticos en las investigaciones científicas. Los bio-arreglos le han proporcionado velocidad al desarrollo de la biología molecular y la ciencia en general. Esta tecnología se basa en la deposición de miles de muestras de forma organizada sobre un soporte sólido y, debido a que tiende a ser utilizada de forma miniaturizada, será posible interrogar el genoma completo en un solo experimento de bio-arreglos. La hibridación de estos soportes permite monitorear la expresión y la secuencia de los genes en ellos organizados. Los resultados son procesados e interpretados a través de programas muy potentes, sensibles y de alta velocidad, diseñados para estos fines. La mayoría de las aplicaciones actuales de esta tecnología están relacionadas con la biomedicina y la agropecuaria y su uso futuro es de dimensiones inimaginables para la ciencia actual.

Palabras claves: microarreglos de ADN, bio-arreglos, expresión de genes, microarreglos de proteínas, microarreglos de tejidos, genómica, SNP

Biotecnología Aplicada 2003;20:77-84

ABSTRACT

Technology, analysis and applications of the bio-arrays. The bio-arrays are a versatile technology with wide possibilities for the scientific advance in the XXI century. Its application had shown a significant growth in the last years and, together with this, the genetic analysis on the scientific research. In the same way that the microchip had impressed velocity to the development of the computerized science, the bio-arrays give velocity to the molecular biology and the science in general.

These represent thousands of samples contained on a support in an organized way and, due to the tendency to be use it as a miniaturized form, will be possible interrogate the genome in only one bio-arrays experiment. The hybridization of these supports of bio-array allows monitor the expression and the sequence of the genes in them organized, being processed and interpreted its results through software's designed for these ends. Most of the current applications of this technology are related with the biomedical science, but their future use is of unimaginable dimensions for the current science.

Keywords: DNA microarrays, bio-arrays, gene expression, protein microarray, tissue microarray, genomic, SNP

Introducción

El conocimiento del genoma es mucho más que el descubrimiento del orden de las bases nucleotídicas en su secuencia. Su lectura ha logrado marcar una nueva era en las investigaciones científicas y ha permitido un acelerado avance al desarrollo de la biomedicina, la agropecuaria, la industria y la bioinformática, entre otras múltiples disciplinas.

El desarrollo vertiginoso de la era post-genómica dependerá en gran medida de lo eficiente que sea la comunidad científica mundial en desarrollar y validar herramientas biológicas relacionadas con el estudio del genoma. Entre estas herramientas se encuentran los bio-arreglos, tecnología que permite el estudio de sistemas biológicos complejos y la generación masiva de datos experimentales. Su utilización permite hacer análisis abarcadores y de amplias dimensiones sobre los eventos que están involucrados en los procesos patológicos y fisiológicos del desarrollo de los animales y las plantas.

Los bio-arreglos pueden ser diseñados para contener en un mismo soporte el genoma completo de los microorganismos o miles de genes eucariontes.

En los laboratorios donde se aplique esta tecnología se hace imprescindible el desarrollo e implementación exitosa de algoritmos de trabajo, disponer de equipos específico y de las estrategias de normaliza-

ción de las señales quimioluminiscentes, fluorescentes o radioactivas. Estos aspectos garantizaran los buenos resultados a obtener con esta tecnología de alta complejidad en su diseño, que necesita un sistema eficiente de análisis de la expresión génica, así como reproducibilidad entre experimentos.

Todo el proceso de análisis de la expresión de genes se desarrolla a través de diferentes etapas entre las que se encuentran la fabricación y diseño del bio-arreglo, la preparación de la sonda a utilizar en la hibridación del soporte, la colección de las muestras, su procesamiento y la lectura e interpretación de las intensidades de las señales. El proceso en su conjunto se realiza con un equipamiento cada vez más sofisticado y costoso que no muchos laboratorios poseen, pero existen variantes comerciales que permiten a los investigadores acceder a distintas etapas de esta tecnología y realizar a menor costo estudios sobre la variabilidad de la expresión y la estructura génica.

Según el diseño y las posibilidades de cada laboratorio, es necesario tener sistemas robotizados para la aplicación de forma masiva de miles de muestras sobre los soportes de bio-arreglos en el menor tiempo posible, lo que siempre debe ocurrir con la misma reproducibilidad. También es importante contar con equipos que controlen con exactitud la temperatura y

el tiempo de hibridación y un escáner lector de fluorescencia, quimioluminiscencia y/o radiactividad de gran sensibilidad y altamente programable (Figura 1). Los resultados de los bio-arreglos son procesados con potentes e inteligentes programas bioinformáticos que permiten normalizar, reducir la dimensionalidad e interpretar el perfil de expresión de los genes.

La Tecnología de los bio-arreglos: definición

Se entiende por bio-arreglos, la disposición de forma organizada de muestras individuales sobre un soporte sólido de forma tal que sea posible su análisis en paralelo. La eficiencia y la novedad de los bio-arreglos son completamente dependientes de la densidad del análisis y la miniaturización que alcanzan.

Según el tipo de muestra aplicada en los soportes, los tipos de soportes, la densidad y la miniaturización que alcanzan, se encuentran en la literatura diferentes tipos de denominaciones para esta tecnología, las más citadas son: *cDNA microarrays*, *DNA chip*, *Gene array*, *Genome array*, *Genome chip*, *Protein chip*, *Gene chip* y *Bio chip* (todas en inglés) [1].

Varias etapas componen la preparación y análisis de los bio-arreglos [2]. La mayor parte de este proceso necesita de una eficiente y potente base bioinformática, la cual se inserta de forma predominante dentro de las diferentes etapas que conforman esta tecnología. El uso de la bioinformática permite almacenar y obtener los datos necesarios para el diseño experimental mediante la creación de potentes bases de datos de fácil uso, de alta seguridad y confiabilidad para almacenar en ellas información valiosa obtenida de la investigación científica [3].

La creación de programas inteligentes para el análisis e interpretación de los resultados obtenidos durante la etapa de hibridación y revelado de señales del bio-arreglo, son de vital importancia para el éxito y



Figura 1. La figura representa los cinco componentes fundamentales para el diseño, hibridación, cuantificación y análisis de los bio-arreglos.

desarrollo de esta tecnología. Estos programas deben realizar la interpretación de los miles de resultados numéricos que son generados en un solo experimento de bio-arreglos, deben realizar diseños experimentales, mostrar las mejores opciones para el desarrollo de la investigación (seleccionar el tipo de muestra y su localización en el soporte), el procesamiento de la imagen y la cuantificación del resultado, la normalización de los datos obtenidos en la cuantificación, el análisis estadístico y la reducción de la dimensionalidad del sistema [4].

En la Figura 2 se representan las etapas que componen el proceso de obtención y análisis de los arreglos de ADN.

En este trabajo que revisa la información esencial publicada entre los años 1999 al 2002, se exponen cada una de estas etapas con mayor profundidad.

Sistema de Manejo de la Información del Laboratorio (LIMS)

Uno de los problemas más importantes de la era post genómica es que el desarrollo de tecnologías de alto flujo de tamizaje han provocado un vertiginoso crecimiento de la información. Como característica general, el ritmo de crecimiento de los bancos de datos biológicos es superior al ritmo de procesamiento de datos. De aquí se comprende la necesidad de contar con sistemas automatizados para el manejo de datos que permitan la integración de la información generada por los grupos de investigación con la información anterior que existe en los bancos de datos. Todo este proceso permite que se mantenga la información organizada de forma tal que se facilite su análisis y se garantice su integridad y seguridad [5].

Los bio-arreglos son una de las tecnologías que permiten llevar a cabo experimentos masivos. En cada experimento de bio-arreglos pueden ser evaluadas miles de muestras individuales frente a un estímulo determinado.

La información que se genera en un experimento de bio-arreglos se puede dividir en tres clases: diseño experimental y procedimientos que permiten reproducir el experimento, información relativa a las muestras y controles utilizados en el ensayo y, por último, el análisis de los resultados y las conclusiones derivadas del experimento.

Esta información generada por los bio-arreglos es sumamente heterogénea. Para cada experimento se deben almacenar las características de diseño de los bio-arreglos usados, las condiciones de preparación de las muestras de hibridación y de su marcaje para ser utilizadas como sonda; el registro y cuantificación de las imágenes de los arreglos y el análisis de los resultados, entre las más importantes. (Figura 3). La forma más racional de almacenar este volumen de información ha sido heredada de los sistemas de manejo de datos del laboratorio (LIMS, siglas en inglés) que se emplean desde la década de los años 80 en el control de la calidad y en los procesos de producción de las industrias químicas, farmacéuticas y metalúrgicas, entre otras.

El LIMS es una metodología que ha sido adaptada a la plataforma investigativa con no pocos esfuerzos.

1. Cheung VG, Morley M, Aguilar F, Massimi A, Kucherlapati R and Childs G. Making and reading microarrays. *Nat Genet* 1999;21:15-9.

2. Holloway AJ, Van Laar RK, Tothill RW, Bowtell DD. Options available-from start to finish-for obtaining data from DNA microarrays II. *Nat Genet* 2002;32 Suppl 2:481-9.

3. Steinlechner M, Parson W. Automation and high through-put for a DNA database laboratory: development of a laboratory information management system. *Croat Med J* 2001;42:252-5.

4. Slonim DK. From patterns to pathways: gene expression data analysis comes of age. *Nat Genet* 2002;32 Suppl 2:502-8.

5. Altman RB, Klein TE. Challenges for biomedical informatics and pharmacogenomics. *Annu Rev Pharmacol Toxicol* 2002;42:113-33.

En el área que corresponde a la bioinformática, el LIMS está llamado a resolver el problema causado por la reciente explosión de datos derivados de la amplia utilización por los investigadores de las tecnologías de tamizaje de alto flujo.

Un LIMS diseñado para bioinformática debe incluir información específica referida a la documentación que fundamenta cada experimento, ya sea su diseño como sus conclusiones, información que permita relacionar los resultados de un experimento con los datos que existen en bancos de datos de secuencias de proteínas y ácidos nucleicos, de estructura de proteínas, en bancos de datos específicos como pueden ser datos de enzimas, de rutas metabólicas, de anticuerpos, de antígenos y muchos más que dependen fundamentalmente de los objetivos del experimento [6].

El depósito de los datos experimentales de los microarreglos de ADN se está convirtiendo en una condición para la publicación de los resultados en varias de las más importantes revistas científicas y es probable que otras muchas otras comiencen a exigirlo [7]. Algunas de las bases de datos públicas donde se pueden depositar los resultados experimentales son por ejemplo ArrayExpress del Instituto Europeo de Bioinformática (<http://www.ebi.ac.uk/arrayexpress>), CIBEX en Japón y GEO de los Institutos Nacionales de Salud de EE.UU.

Blancos e interacciones en los soportes de bio-arreglos

En los soportes de bio-arreglos pueden estar organizadas muestras de origen biológico o sintético. Estas muestras, las cuales van a funcionar como blancos de unión de sus moléculas complementarias en forma de ARN, ADN y proteínas, van a generar diferentes interacciones moleculares entre los blancos y las sondas utilizadas en cada experimento de bio-arreglos (Tabla 1)

Sonda experimental

Una de las etapas más importante para el éxito y confiabilidad del resultado es la selección adecuada del tipo de muestra a estudiar y la preparación de la sonda a utilizar en los experimentos de hibridación. En lo posible la sonda debe proceder de poblaciones celulares homogéneas [10], de las cuales se aisle un ARN integro y abundante que permita generar un ARN mensajero verdaderamente representativo del ambiente celular que se está estudiando. La homogeneidad de la muestra, garantiza que se esté analizando una población celular completamente transformada por la acción de algún agente químico, físico o biológico.

Las fuentes de muestras para la obtención de ADN genómico y ARN total, pueden ser a partir de cultivo de tejidos, de microorganismos bajo diferentes condiciones experimentales, biopsias obtenidas a partir de órganos y tejidos las cuales pueden ser procesadas mediante la técnica de microdissección de tejidos a láser (LCM, siglas en Inglés).

De todas las fuentes anteriores, la más confiable, pero también la más costosa y laboriosa es la de LCM [11]. Esta técnica permite el aislamiento de células de interés a partir de tejidos heterogéneos

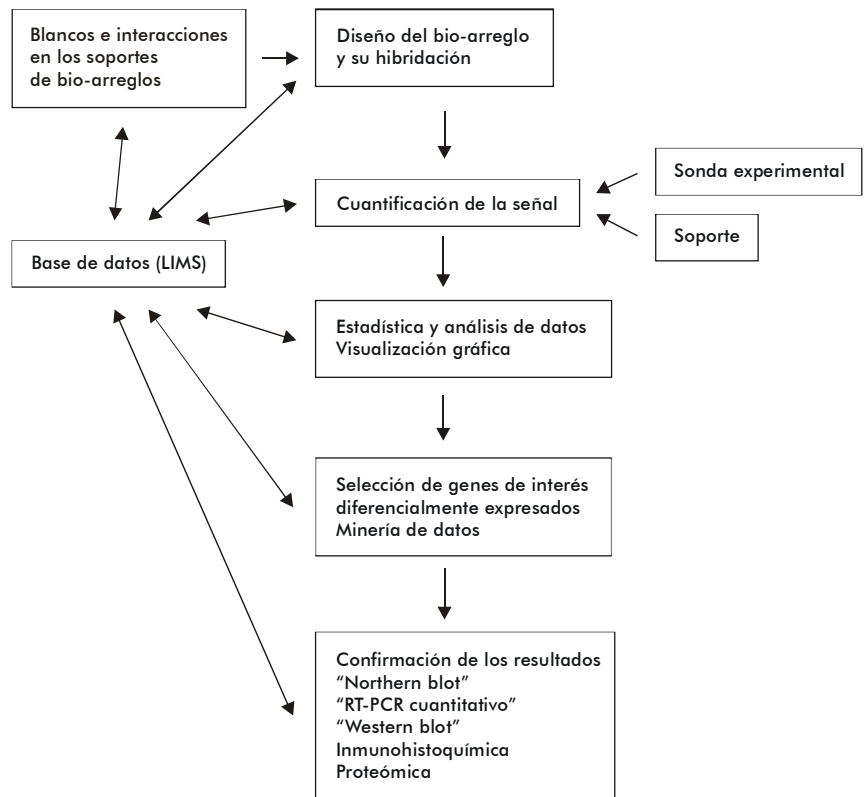


Figura 2. Diferentes etapas conforman un experimento en bio-arreglos. Desde el diseño y la ubicación de los blancos sobre el soporte, hasta la aplicación de herramientas bioinformáticas para el análisis de los resultados.

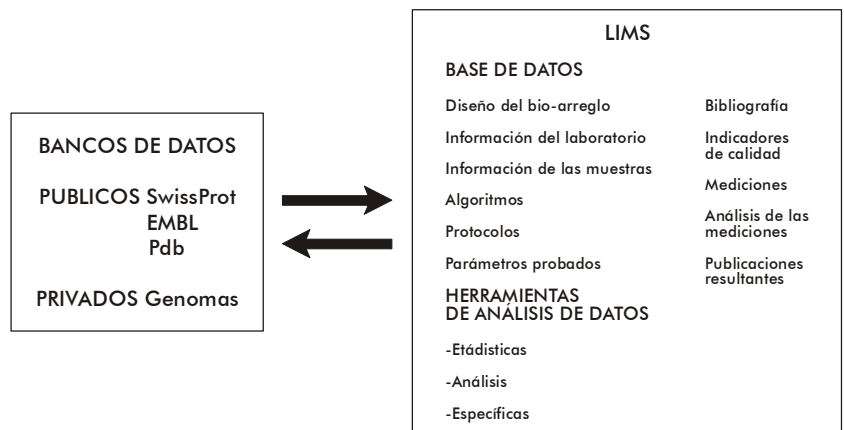


Figura 3. El LIMS proporciona estructuras que aseguran la integridad de los datos, que garantizan la integración de estos datos a los bancos de datos públicos y privados e implementa herramientas de análisis de datos.

prefijados Tiene diferentes etapas de trabajo como son: la ubicación individual de los blancos de interés, el disparo a láser sobre estos blancos, la adhesión de las células al film de transferencia y la separación de las células de interés del tejido necrótico y el estroma. Es así como de forma pura se pueden aislar placas de Alzheimer del cerebro, glomérulos del riñón, células neoplásicas de cáncer de mama, y otras. Se evita, en todas las ocasiones, seleccionar células no relaciona-

6. Searls DB. Using bioinformatics in gene and drug discovery. Drug Discov Today 2000;5:135-43.

7. Stoekert CJ, Causton HC, Ball CA. Microarray databases: standards and ontologies. Nat Genet 2002;32 Suppl 2:469-73.

das con el daño o transformación celular que se desea estudiar [12].

Material de soporte, diseño e hibridación

El material de soporte para el diseño del bio-arreglo puede ser rígido o flexible, lo que le imprime características propias a algunas de las etapas de su preparación (Tabla 2).

Según el soporte: nylon, vidrio o plástico, se definen las características del bio-arreglo, como son la naturaleza y la densidad de muestras, el diámetro del punto de aplicación, la dimensionalidad y la miniaturización, así como el tipo de procesamiento de la señal, ya sea fluorescencia, radioactividad o quimioluminiscencia (Figura 4).

Cuando se utilizan los filtros de nylon, que es la forma menos costosa de diseño de los bio-arreglos, sólo se puede aspirar a una baja densidad de clones en cada filtro. Los aplicadores comerciales para este tipo de bio-arreglos, permiten obtener un diámetro de la aplicación dentro de los límites de la clasificación como bio-macroarreglos. Sin embargo, a pesar de que este diseño no favorece la dimensionalidad, ni la miniaturización, se puede considerar fácil de realizar y de reproducir en laboratorios con pocos recursos. Quizás su diseño sea limitado y no permita descubrir nuevos genes, ni tener una imagen global de cómo responde el genoma ante los cambios, pero es posible su aplicación en la generación de métodos de diagnóstico con la utilización de genes relacionados con funciones específicas.

En la Figura 5 se muestran los resultados obtenidos en la hibridación de un arreglo de ADN, cuyo diseño experimental permitió la estandarización de esta tecnología en la división de Farmacéuticos del Centro de Ingeniería Genética y Biotecnología (CIGB) de La Habana, Cuba. Los clones de ADN se seleccionaron al azar de una biblioteca comercial de ADN complementario (ADNc) de ratón de 11 días de desarrollo (Clontech, USA). En las columnas C1-C4 del diseño del arreglo se colocaron un total de 88 plásmidos de secuencias conocidas, los cuales portaban genes humanos relacionados con factores de crecimiento, interleucinas e interferones, genes que codifican para la proteína β amiloide del Alzheimer, quinasas e inhibidores del ciclo celular, genes que codifican para las proteínas del citoesqueleto, vectores plasmídicos, factores transcripcionales y otros. Además se tuvo en cuenta incluir genes expresados de forma constitutiva, como son el gliceraldehído 3-fosfato deshidrogenasa (GAPDH) y la β actina humana. Se utilizaron como controles negativos genes de plantas (metil transferasa de álamo) y vectores plasmídicos sin insertos génicos. Este filtro con un total de 2304 aplicaciones se hibridó con una sonda de ADN perteneciente al vector pGAD10 (Clontech, USA), el cual se utilizó para construir la librería ADNc. Como se observa en la figura, la mayoría de los ADN aplicados al filtro, hibridaron con la sonda de ADN utilizada en la experiencia, revelándose además el reconocimiento de la sonda marcada con algunos de los 88 plásmidos controles de secuencias conocidas.

La utilización del vidrio como soporte para el diseño de los bio-arreglos tiene múltiples ventajas en relación con el nylon. Por ejemplo el vidrio es un

Tabla 1. Posibles interacciones que se generan de la hibridación de los bio-arreglos, en dependencia de las muestras ancladas al soporte.

Muestras sobre el soporte	Interacciones
1-ADN (oligonucleótidos, productos de PCR, bacterias)	ADN-ADN, ADN-ARN, ADN-proteína
2-Proteínas (extractos celulares, enzimas, anticuerpos, péptidos)	Enzima-sustrato, enzima-efector, enzima-inhibidor, Antígeno-anticuerpo, anticuerpo-ARN, anticuerpo-ADN, anticuerpo-superficies celulares, ADN-proteína, proteína-moléculas pequeñas, proteína-proteína, proteína-receptor
3-Células	Hormona-receptor, anticuerpo-receptor y otros ligandos específicos
4-ARN (ARN de transferencia, ARN ribosomal, ARN mensajero, oligonucleótidos)	ARN-ADN, ARN-ARN, ARN-proteína
5-Tejidos [8, 9]	Proteínas-anticuerpos específicos Inmunohistoquímica, inmunofluorescencia e hibridación in situ

Tabla 2. Posibles formatos para el diseño y preparación de los soportes de bio-arreglos.

Características	Nylon	Vidrio	Plástico
Densidad (número de genes/cm ²)	< 20 genes/cm ²	10 000-30 000 genes/cm ²	≈ 100 genes/cm ²
Propiedades físicas del soporte	Poroso y blando	No poroso y rígido	No poroso y rígido
Definición según el diámetro del punto de aplicación	Macroarreglos > 300 micras	Microarreglos <250 micras (miniaturización)	Microarreglos <250 micras
Tamaño del soporte	8 x 12 cm	8 x 8 mm- 1.28 x 1.28 cm	8 x 12 cm
Marcaje de la sonda	Radioactivo y quimioluminiscencia	Fluorescencia	Radioactivo
Fabricación	Mecánica o manual	Fotolitografía, "Ink-jet", mecánica	Mecánica
Sensibilidad para la obtención de la señal	Alta	Baja	Alta
Blancos (muestras ancladas al soporte)	ADN complementario (PCR, bibliotecas)	Oligonucleótidos	Oligonucleótidos
Obtención de la imagen	Fosfoimager y autoradiografía	Escáner fluorescente	Fosfoimager y autoradiografía

material que soporta altas temperaturas, sufre menos deterioro por su manipulación, facilita el enlace e inmovilización efectiva de las muestras en su superficie y no es poroso lo que permite utilizar cantidades muy pequeñas de solución de hibridación que favorece que la cinética de interacción entre las moléculas sea mucho mayor. La hibridación de este tipo de diseño se realiza con moléculas marcadas con fluoróforos que emiten señales a diferentes longitudes de ondas, lo que permite hibridar un solo bio-arreglo con diferentes fluoróforos simultáneamente, opción que no es posible realizar en los bio-arreglos de soportes de nylon [1].

Los soportes de plástico a su vez tienen ventajas sobre el vidrio y el nylon, ya que además de ser de un material rígido y no poroso, permiten realizar la hibridación con muestras marcadas radioactivamente que es una forma de detección de señales mucho más sensible que la fluorescencia y quimioluminiscencia [13].

Procesamiento y Análisis de datos

El objetivo de la mayoría de los experimentos de microarreglos de ADN es interrogar el patrón de expresión diferencial de miles de transcripts en muestras biológicas. Por su complejidad estos experimentos tienen muchas fuentes de variabilidad lo que implica diseños experimentales adecuados que permita a los

8. Bolli M, Kocher T, Adamina M, Guller U, Dalquen P, Haas P, et al. Tissue microarray evaluation of melanoma antigen E (MAGE) tumor-associated antigen expression: potential indications for specific immunotherapy and prognostic relevance in squamous cell lung carcinoma. *Ann Surg* 2002;236:785-93.

9. Bubendorf L, Nocito A, Moch H, Sauter G. Tissue microarray (TMA) technology: miniaturized pathology archives for high-throughput in situ studies. *J Pathol* 2001; 195:72-9.

10. Miura K, Bowman ED, Simon R, Peng AC, Robles AJ, Jones RT, et al. Laser capture microdissection and microarray expression analysis of lung adenocarcinoma reveals tobacco smoking- and prognosis-related molecular profiles. *Cancer Res* 2002;62:3244-50.

11. Mikulowska-Mennis A, Taylor TB, Vishnu P, Michie SA, Raja R, Horner N, et al. High-quality RNA from cells isolated by laser capture microdissection. *Biotechniques* 2002;33:176-9.

12. Bonaventure P, Guo H, Tian B, Liu X, Bittner A, Roland B, et al. Nuclei and subnuclei gene expression profiling in mammalian brain. *Brain Res* 2002;943: 38-47.

especialistas lograr un óptimo uso de la información obtenida [14]. Esta variabilidad puede dividirse en tres niveles: la variabilidad biológica, intrínseca de cada organismo y que está influenciada por factores ambientales, genéticos y por la complejidad de la muestra; la variabilidad técnica, introducida durante el procedimiento de extracción, marcaje e hibridación de las muestras y los errores de la medición asociados a las lecturas de las señales y que puede estar influenciado por múltiples factores.

El proceso de análisis de datos comienza con el procesamiento de las imágenes que comprende tres etapas: asignación de las coordenadas de cada señal, la clasificación de las señales como verdaderas o fondo y la extracción de los valores de intensidad por diferentes métodos (media, mediana e integración de la intensidad). En algunos programas se incluyen además procedimientos estadísticos que permiten evaluar la calidad de las mediciones. La precisión en la identificación de las señales y los procedimientos de cálculo de las intensidades varía entre los diferentes programas que se encuentran disponibles de forma pública o comercial y aún no existe un consenso de cuál es la mejor metodología para la extracción de los valores de intensidad [15].

Una vez calculados los valores de intensidad es necesario la transformación de los datos que permita la eliminación de los valores cuestionables o de baja calidad. Típicamente la primera transformación que se realiza se conoce como normalización. Este proceso es imprescindible para poder comparar intensidades entre diferentes hibridaciones.

Existen diferentes métodos para normalizar los resultados como son: normalización de la intensidad total y el uso de regresiones lineales, entre las más usadas. La normalización de los datos puede ser también global o local. La normalización local por ejemplo se aplica para las señales que han sido aplicadas con la misma aguja. La segunda transformación consiste en filtrar o eliminar los valores donde las señales no son estadísticamente diferentes a la del fondo del soporte o la de los controles negativos. Además se filtran las señales saturadas o cuya intensidad es muy próxima a los valores máximos de detección del equipo [16].

Los experimentos de microarreglos de ADN comprenden la medición de miles de señales por lo que se requiere de métodos que permitan reducir los datos a aquellos genes que son más variables. En los primeros experimentos de microarreglos de ADN sólo se analizaban los genes cuya expresión variaba más de dos veces respecto al control. En la actualidad se están desarrollando otros métodos de identificación de la expresión diferencial fundamentados en variantes de métodos estadísticos conocidos. Generalmente estos métodos se basan en realizar un cálculo estadístico y determinar la significación de lo observado [17]. Esta selección de los genes más informativos es el primer paso en la reducción de la complejidad de los datos.

Los datos de expresión génica se representan como una matriz de valores en los cuales los genes se colocan en las filas y las condiciones experimentales en las columnas. Según el experimento, los valores de expresión génica pueden ser utilizados para clasificar con-

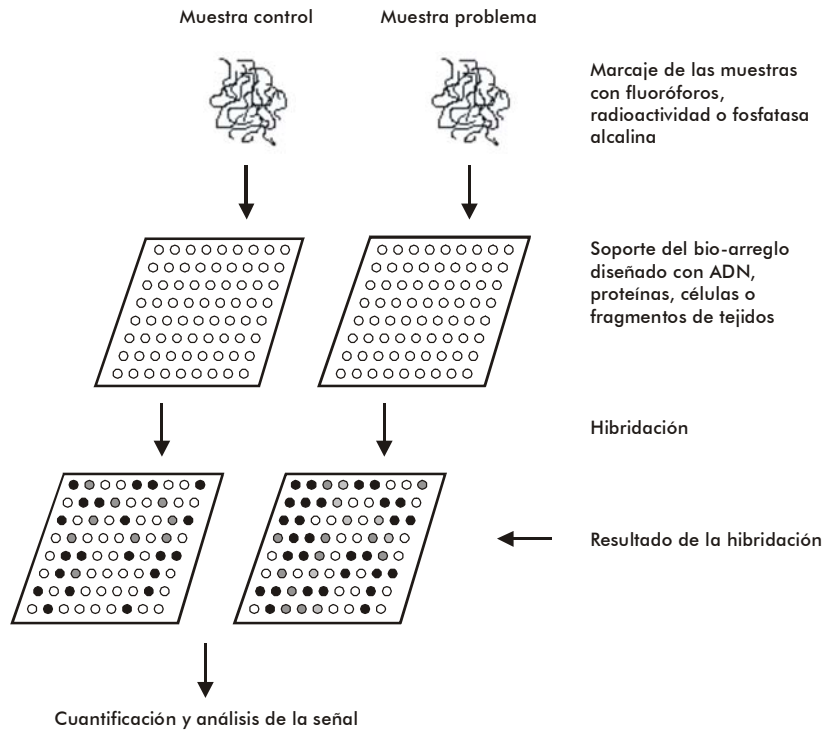


Figura 4. Etapas del proceso de obtención y marcaje de la sonda, diseño del soporte del bio-arreglos, hibridación y cuantificación de la señal.

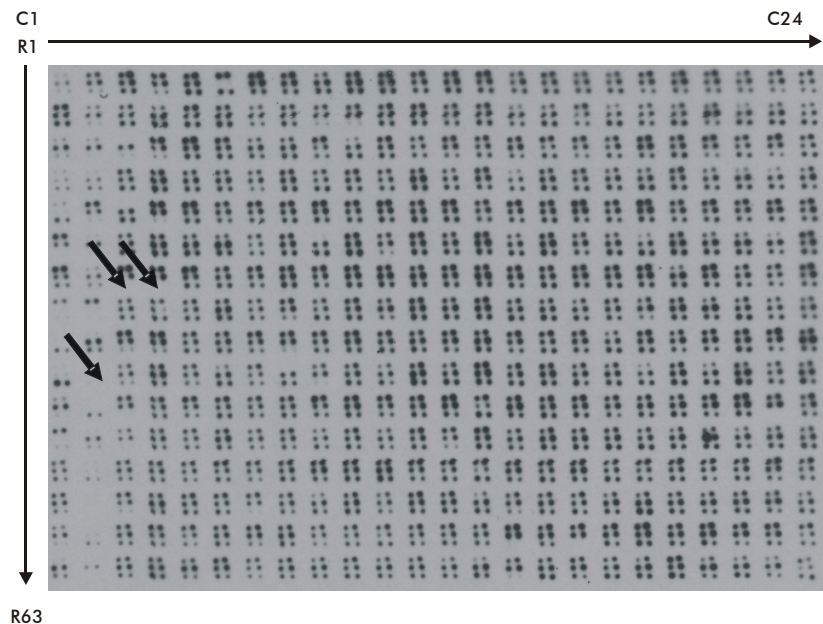


Figura 5. En el filtro de nylon (Nytran SuperCharge de Schleicher & Schuell, Germany) se aplicaron un total de 1152 clones de ADN por duplicados, y posteriormente se hibridaron con una región de ADN común a todos. La figura muestra el resultado de la hibridación, la cual se reveló mediante métodos quimioluminiscentes (Alkphos Direct Labelling Reagents, from Amersham Pharmacia Biotech UK). Con una flecha aparecen localizados alguno de los controles negativos utilizados en el filtro.

diciones (columnas) o patrones de expresión génica (filas). En ambos casos un primer paso consiste en agrupar los genes o las condiciones con patrones de expresión semejante. Estos procedimientos pueden

ser no supervisados como los métodos jerárquicos o supervisados donde los datos son entrenados para que puedan reconocer patrones [4].

Existen muchas soluciones en dominios públicos y privados que pueden ser utilizados para agrupar genes. Una lista muy actualizada de programas de análisis de arreglos de ADN puede ser encontrada en el sitio <http://ihome.cuhk.edu.hk/~b400559/>.

La mayor contribución al conocimiento de las funciones de los genes se obtendrá probablemente de la combinación de los datos de los arreglos de ADN y de otras fuentes de información genómica y biomédica. Los algoritmos capaces de integrar este tipo de información se están desarrollando aceleradamente. Varios grupos están combinando la información de microarreglos de ADN con el análisis de los promotores y la regulación génica [18]. Mucha de la información relevante acerca de la función de los genes esta contenida en la literatura médica. Los sistemas de minería de textos para la extracción automatizada de información serán de un gran valor para poder arribar a conclusiones exactas [19].

Validación de los experimentos de microarreglos de ADN

Existen dos formas de comprobar los resultados de los experimentos de los microarreglos de ADN: análisis *"in silico"* y análisis experimental.

El análisis *"in silico"* compara los resultados de los microarreglos de ADN con los resultados publicados en la literatura y existentes en bases de datos. La concordancia entre los resultados obtenidos por microarreglos de ADN entre grupos diferentes así como la información conocida, valida el comportamiento general del sistema y garantiza la seguridad de los datos obtenidos entre los que se incluyen los resultados únicos de ese experimento. Un ejemplo de validación *"in silico"* lo constituye el análisis de cuatro series de experimentos diferentes realizados en muestras de cáncer de próstata que pudieron confirmar la expresión diferencial de varios de los genes en todos los estudios realizados [20].

Los métodos de validación experimentales de los niveles de expresión génica comienzan típicamente por la misma muestra utilizada en el estudio de microarreglos de ADN inicial. Las metodologías que se utilizan varían y dependen en muchas ocasiones de la disponibilidad de reactivos apropiados como cebadores y anticuerpos. Para el estudio de los niveles de transcritos los métodos más frecuentes son: la reverso transcripción semi-cuantitativa (RT-PCR), RT-PCR en tiempo real, northern blot, ensayos de protección de ribonucleasas y las hibridaciones *"in situ"*. El RT-PCR en tiempo real ha sido el método de elección ya que requiere cantidades mínimas de ARNm, sin embargo requiere significativos esfuerzos para la optimización de las condiciones de amplificación. Adicionalmente a la validación de los resultados en los transcritos es igualmente importante su correlación con las concentraciones correspondientes de proteínas aunque hasta el momento no existe una correlación clara. Para la medición de la abundancia de las proteínas se utilizan las técnicas de inmunodetección en extractos celulares o en muestras de tejidos [21].

Aplicaciones de los Bio-arreglos

Esta tecnología tiene dos formas fundamentales de interrogar al genoma, tanto por su secuencia nucleotídica,

como por su expresión. Estas dos formas son: los estudios de polimorfismo génico individual (SNP, siglas en Inglés), que es la variación génica más frecuentemente encontrada en el genoma humano y el perfil de expresión de genes (GEP, siglas en inglés) (Figura 6).

El SNP puede ser de utilidad para el genotipaje y el diagnóstico de enfermedades. Este posibilitaría el uso de la terapia génica como método de cura transitoria o total de estas y la predicción de la respuesta farmacogenómica a los medicamentos, así mismo, permitiría realizar una terapia farmacológica personalizada y efectiva contra diferentes enfermedades [22].

Ampliamente utilizados en los estudios del cáncer, los bio-arreglos de SNP permiten entre otras múltiples aplicaciones la clasificación de tumores, la determinación de una terapia personalizada para combatirlos y los métodos de diagnósticos para conocer la predisposición a padecer algún tipo de cáncer. Un ejemplo de lo comentado anteriormente lo constituyen los microarreglos de ADN destinados a conocer la existencia de mutaciones en el gen p53. Este gen funciona como un supresor tumoral que induce a las células a autodestruirse cuando el daño de su ADN es irreparable. El gen p53 se encuentra frecuentemente mutado en los cánceres humanos. Las estadísticas apuntan a que aproximadamente entre un 30-70% de las enfermedades malignas contienen el gen p53 alterado en algún punto de su secuencia génica [23]. El progreso de la enfermedad no es igual para los enfermos de cáncer que tienen esta mutación en comparación con los que tienen el gen normal, los primeros tienen un pronóstico reservado, con una gran tendencia a la metástasis y a la resistencia a terapias con fármacos y radiaciones ionizantes. La importancia de aplicarles a los pacientes con cáncer una terapia personalizada y funcional, es vital para su supervivencia y el mejoramiento de la calidad de vida.

El otro grupo de aplicaciones de los bio-arreglos se centra en el estudio de la variación del perfil de expresión de genes (genómica funcional) [24]. Mediante esta aplicación es posible conocer cómo las células responden ante los cambios fisiológicos, ambientales y al estrés, mediante la cuantificación de los patrones de expresión de sus ARN mensajeros. Lo anterior tiene importantes implicaciones en el descubrimiento de nuevos medicamentos y blancos terapéuticos. Un ejemplo muy interesante de esta aplicación, es su utilización en el conocimiento de los genes que varían su expresión durante el proceso de envejecimiento, y el efecto que tiene sobre esta variabilidad la utilización de una dieta con restricción calórica [25]. De esta forma se han podido identificar seis genes los cuales varían su perfil de expresión en ratones viejos alimentados con y sin restricción calórica. Tres de estos genes han sido relacionados con los procesos de apoptosis (muerte celular programada) y los otros tres con la regulación del ciclo celular, la reparación del ADN y el recambio proteico. Con estos estudios se pudo conocer que los genes involucrados en los procesos de apoptosis y de recambio proteico disminuyen su expresión cuando existe restricción calórica [26].

La genómica funcional tiene también un amplio campo de aplicación en la búsqueda de terapias más efectivas contra el cáncer. Durante el proceso de división de las células malignas, se han podido detectar genes

13. Lipshutz RJ, Fodor SP, Gingeras TR, Lockhart DJ. High density synthetic oligonucleotide arrays. *Nat Genet* 1999; 21:20-4.

14. Churchill GA. Fundamentals of experimental design for cDNA microarrays. *Nat Genet* 2002;32 Suppl 2:490-5.

15. Dopazo J, Zanders E, Dragoni I, Amphlett G, Falciani F. Methods and approaches in the analysis of gene expression data. *J Immunol Methods* 2001; 250:93-112.

16. Quackenbush J. Microarray data normalization and transformation. *Nat Genet* 2002;32 Suppl 2:496-501.

17. Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci U S A* 2001;98:5116-21.

18. Pilpel Y, Sudarsanam P, Church GM. Identifying regulatory networks by combinatorial analysis of promoter elements. *Nat Genet* 2001;29:153-9.

19. Jenson TK, Laegreid A, Komorowski J, Hovig E. A literature network of human genes for high-throughput analysis of gene expression. *Nat Genet* 2001;28:21-8.

20. Rhodes DR, Barrette TR, Rubin MA, Ghosh D, Chinnaiyan AM. Meta-analysis of microarrays: interstudy validation of gene expression profiles reveals pathway dysregulation in prostate cancer. *Cancer Res* 2002;62:4427-33.

21. Chuaiqui RF, Bonner RF, Best CJ, Gillespie JW, Flaig MJ, Hewitt SM, et al. Post-analysis follow-up and validation of microarray experiments. *Nat Genet* 2002;32 Suppl 2:509-14.

22. Ginsburg GS, McCarthy JJ. Personalized medicine: revolutionizing drug discovery and patient care. *Trends Biotechnol* 2001;19:491-6.

23. Ahrendt SA, Halachmi S, Chow JT, Wu L, Halachmi N, Yang SC, et al. Rapid p53 sequence analysis in primary lung cancer using an oligonucleotide probe array. *Proc Natl Acad Sci U S A* 1999;96:7382-7.

24. Petricoin EF, Hackett JL, Lesko LJ, Puri RK, Gutman SI, Chumakov K, et al. Medical applications of microarray technologies: a regulatory science perspective. *Nat Genet* 2002;32 Suppl 2:474-9.

25. Han E, Hilsenbeck SG. Array-based gene expression profiling to study aging. *Mech Ageing Dev* 2001;122:999-1018.

26. Han E, Hilsenbeck SG, Richardson A, Nelson JF. cDNA expression arrays reveal incomplete reversal of age-related changes in gene expression by calorie restriction. *Mech Ageing Dev* 2000;115:157-74.

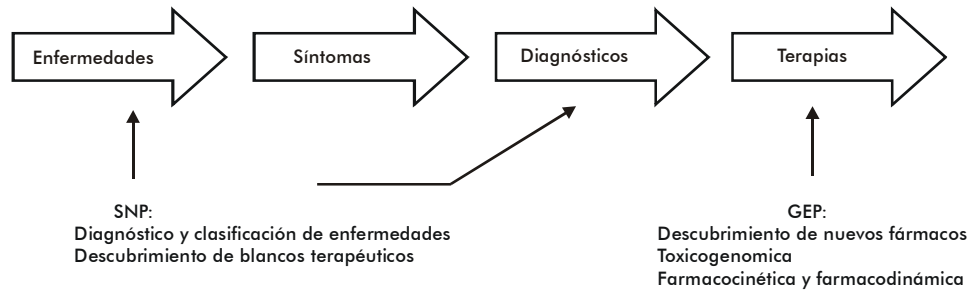


Figura 6. Mediante el uso de los bio-arreglos, la tendencia al tratamiento de las enfermedades será por las funciones alteradas y no por los síntomas del paciente. Con el uso de esta tecnología se puede llegar a predecir la posibilidad de padecer determinadas enfermedades, lo cual posibilitaría un tratamiento preventivo y efectivo para evitar su desarrollo.

que aumentan o disminuyen su expresión. En este sentido, los bio-arreglos pueden ser altamente informativos sobre cómo las células malignas responden ante la aplicación de un fármaco determinado. Esto permite aplicar una terapia personalizada y efectiva para cada tipo de malignidad celular.

El análisis temporal y espacial de la expresión de genes ha sido aplicado con éxito al estudio del proceso de cicatrización de tejidos, tanto en animales como en humanos. En animales se han valorado un total de 142 genes relacionados con los procesos de cicatrización y regeneración de tejidos, los cuales varían su expresión durante las diferentes etapas de la cicatrización de heridas (coagulación, inflamación, proliferación y remodelación) en las diferentes cepas de ratones estudiadas [27]. Los resultados obtenidos en estos estudios han permitido identificar nuevos genes implicados en el proceso de cicatrización de tejidos, los cuales aumentan su expresión durante la etapa de inflamación.

Sin lugar a dudas, el estudio del perfil de expresión de genes en diferentes procesos como el estado de hiperglicemia en el músculo esquelético de diabéticos tipo 2 [28], el daño tisular producido por quemaduras [29], el proceso de cicatrización en heridas corneales [30], el papel funcional de los genes durante el desarrollo de enfermedades inflamatorias como es la psoriasis [31], la identificación de genes que varían su expresión en el cerebro de las personas esquizofrénicas, muchos de ellos relacionados con las señales presinápticas [32], la interacción virus-hospedero [33] y otras innumerables aplicaciones biomédicas de la tecnología de los bio-arreglos, son una fuente inagotable de resultados científicos y de uso potencial en el diseño de métodos de diagnósticos, terapias génicas y fármacos para la prevención y cura de las enfermedades humanas.

Los bio-arreglos de proteínas y péptidos son una forma más novedosa del estudio de la genómica funcional, su desarrollo ha permitido revolucionar la proteómica y crear enormes oportunidades en el campo de la investigación biomédica [34]. La generación de un *chip* de proteínas requiere de mucho más esfuerzo que de un *chip* de ADN, esto es debido a que la molécula de ADN es muy estable durante los procesos de hibridación, sin embargo las proteínas son mucho más difíciles de producir y de manipular, siendo susceptibles a desnaturalizarse, a perder su estructura tridimensional, y su actividad específica durante el

proceso de extracción e hibridación [35]. Mediante la aplicación de este tipo de bio-arreglos se pueden conocer funciones proteicas, como la actividad enzimática, el reconocimiento proteína-proteína, la especificidad de anticuerpos ante diferentes antígenos, sus epitopes de reconocimiento, y la interacción proteínas-ácidos nucleicos, entre otras [36, 37].

Los microarreglos de tejidos son una herramienta poderosa para de forma rápida caracterizar el perfil fenotípico de un gran número de muestras. Su mayor aplicación hasta el momento está dirigida a conocer mediante métodos inmunohistoquímicos e hibridación *in situ*, cómo varía la expresión de algunos genes marcadores de malignidad en linfomas de Hodgkin [38], en células cancerosas del pulmón [8] y el carcinoma de colon [39].

Conclusiones

Dentro de la genómica funcional y estructural, los bio-arreglos son una tecnología de avanzada, la cual permite develar los misterios de la biología (genes y genomas). La revolución actual en la tecnología de los microarreglos de ADN, la bioinformática (biología computarizada) y el inminente desarrollo y aplicación de sus análogos en proteínas y tejidos, van a permitir que de forma acelerada se identifiquen nuevos fármacos y blancos terapéuticos. De forma similar, los avances en la identificación de los polimorfismos de nucleótidos (SNP) propiciarán el descubrimiento de un gran número de genes candidatos para la producción de nuevos métodos de diagnóstico y la caracterización molecular de muchas enfermedades humanas, de plantas y de animales.

Los experimentos realizados hasta el presente con la utilización de los bio-arreglos, han confirmado el patrón de expresión de genes de funciones conocidas, y brindado además información valiosa sobre genes con funciones desconocidas. El verdadero valor de su aplicación es la identificación de genes que sean significativamente sobre expresados o silenciados (en un 10 a un 20% en relación con el control negativo), los cuales jueguen un papel primordial en el desarrollo y progresión de las enfermedades. Por esto es de gran valor la aplicación de los avances obtenidos en las ciencias computarizadas y la clínica, a la organización e interpretación de los miles de datos generados en los bio-arreglos.

El uso futuro de esta novedosa tecnología está dirigido al estudio de la célula como una unidad integral,

27. Li X, Gu W, Masinde G, Covarrubias M, Mohan S, Baylink DJ. Temporal analysis of gene expression by microarray during wound healing. *Wounds* 2002;14:67-78.

28. Sreekumar R, Halvatsiotis P, Schimke JC, Nair KS. Gene expression profile in skeletal muscle of type 2 diabetes and the effect of insulin treatment. *Diabetes* 2002; 51:1913-20.

29. Spies M, Dasu MR, Svrakic N, Nestic O, Barrow RE, Perez-Polo JR, et al. Gene expression analysis in burn wounds of rats. *Am J Physiol Regul Integr Comp Physiol* 2002;283:R918-30.

30. Cao Z, Wu HK, Bruce A, Wollenberg K, Panjwani N. Detection of differentially expressed genes in healing mouse corneas, using cDNA microarrays. *Invest Ophthalmol Vis Sci* 2002;43:2897-904.

31. Oestreicher JL, Walters IB, Kikuchi T, Gilleaudeau P, Surette J, Schwertschlag U, et al. Molecular classification of psoriasis disease-associated genes through pharmacogenomic expression profiling. *Pharmacogenomics J* 2001;1:272-87.

32. Sawa A and Snyder SH. Schizophrenia: diverse approaches to a complex disease. *Science* 2002;296:692-5.

33. Fruh K, Simmen K, Luukkonen BG, Bell YC and Ghazal P. Virogenomics: a novel approach to antiviral drug discovery. *Drug Discov Today* 2001;6:621-7.

34. Houseman BT, Mrksich M. Towards quantitative assays with peptide chips: a surface engineering approach. *Trends Biotechnol* 2002;20:279-81.

35. Eickhoff H, Konthur Z, Lueking A, Lehrach H, Walter G, Nordhoff E, et al. Protein array technology: the tool to bridge genomics and proteomics. *Adv Biochem Eng Biotechnol* 2002;77:103-12.

36. Zhu H and Snyder M. Protein arrays and microarrays. *Curr Opin Chem Biol* 2001;5:40-5.

37. Bussow K, Konthur Z, Lueking A, Lehrach H, Walter G. Protein array technology. Potential use in medical diagnostics. *Am J Pharmacogenomics* 2001;1:37-43.

38. Hedvat CV, Hegde A, Chaganti RS, Chen B, Qin J, Filippa DA, et al. Application of tissue microarray technology to the study of non-Hodgkin's and Hodgkin's lymphoma. *Hum Pathol* 2002;33:968-74.

una entidad química, para lo cual no basta sólo con mirar las señales del control celular dirigidas a la síntesis del ARN mensajero y la composición del ADN genómico, sino también la cantidad real de proteínas que existe en la célula en condiciones fisiológicas y/o patológicas.

Recibido en Abril de 2002. Aprobado en Enero de 2003.

Agradecimientos:

Los autores queremos agradecer al Dr. Eduardo Pentón por habernos motivado y apoyado en la confección de este trabajo, y a los Drs. Gabriel Padrón y Rolando Rodríguez por contribuir al desarrollo de la bioinformática en el CIGB.

39. Hoos A, Nissan A, Stojadinovic A, Shia J, Hedvat CV, Leung DH, *et al.* Tissue microarray molecular profiling of early, node-negative adenocarcinoma of the rectum: a comprehensive analysis. *Clin Cancer Res* 2002;8:3841-9.