

EXPLORACIÓN DE MICROMATRICES DE ADN MEDIANTE ANÁLISIS MULTIRRESOLUCIÓN

José Luis Sánchez López, Andrés Santos

Departamento de Ingeniería Electrónica. Escuela Técnica Superior de Ingenieros de Telecomunicación. Universidad Politécnica de Madrid. Ciudad Universitaria s/n. Madrid 28040.
Correo electrónico: andres@die.upm.es jlsanchez@die.upm.es

RESUMEN

El objeto de este artículo es presentar un nuevo método multi-resolución para la exploración de micromatrices. Combina varios modos de exploración, con diferentes grados de resolución, para la obtención y análisis de imágenes de dispositivos tales como las micromatrices de ADN (MMADN). Consta de una exploración a baja resolución que localiza los puntos sonda en la MMADN y una exploración de alta resolución que explora sólo las zonas de interés, restringidas alrededor del punto sonda. Gran parte de la imagen de una MMADN suele ser fondo que no necesita ser explorado a alta resolución. La novedad de este sistema radica en evitar explorar a alta resolución toda la superficie de la MMADN, concentrándose únicamente en las zonas que contienen información útil.

1. ESTADO DE LA TÉCNICA ANTERIOR

La Genómica (ciencia que estudia todo el ADN celular) ha pasado de ser estructural (secuenciadora) a funcional (usa las secuencias para entender las funciones génicas). Se trata no sólo de conocer cada gen, sino de estudiar cuándo y cómo se expresa, mediante la observación simultánea del comportamiento conjunto de genes y redes génicas, gracias a las micromatrices de ADN.

1.1. Tecnología de micromatrices de ADN

El procedimiento de análisis genético con micromatrices de ADN tiene cuatro etapas fundamentales [9]:

- **Diseño y construcción de la MMADN:** Se depositan sobre el portaobjetos de un microscopio, de vidrio o silicio, conjuntos de muestras de hebras de ADN, formando una matriz microscópica de puntos sonda o "spots". Cada sonda puede contener hasta 10^6 hebras idénticas, muestras de una secuencia o gen característicos. Una MMADN puede llegar a albergar 400.000 sondas en 1.6 cm^2 . Existen diversos métodos para fijar las hebras al porta: crecimiento fotolitográfico, microdepósito, impresión por chorro de tinta.

- **Preparación y ejecución de los experimentos:** Se preparan las muestras del ADN "objetivo" bajo estudio: se extrae el ARN mensajero de dos tipos de célula: unas sanas o de control, otras de prueba que están enfermas o han sido tratadas con algún fármaco. Después se transforma el ARNm

en ADNc y se etiqueta con un marcador fluorescente, Cy3 (verde, 543 nm) para el control y Cy5 (rojo, 633 nm) para la prueba. Se mezclan las muestras y se vierten sobre las sondas. Las muestras objetivo sólo hibridarán con aquellas sondas cuya secuencia sea exactamente complementaria a la suya.

- **Exploración y Medida:** La tercera etapa es en la que se enmarca el objeto de este estudio. Al hibridar, las muestras objetivo quedan fijas a las sondas. Si se excitan los marcadores fluorescentes (por ejemplo mediante microscopía confocal láser) se obtendrá un patrón de colores que revelará la expresión génica de cada experimento. El color de cada punto indicará la abundancia relativa de ADNc hibridado a la sonda: si es verde ese gen sólo se expresa en la célula de control; si es rojo sólo se activa en la célula enferma o tratada; si es amarillo se expresa por igual en condiciones de control o de prueba; si es negro no se expresa en ningún caso.

El proceso clásico de análisis de imagen consta a su vez de 3 etapas: *Ubicación* (localización geométrica del centro de los puntos), *Segmentación* (clasificación de los píxeles como señal o fondo), *Extracción de Información* (intensidades de los puntos y los fondos de imagen, medidas de calidad) [14].

- **Análisis de los resultados:** La última etapa pertenece a la bioinformática, siendo las principales técnicas aplicadas: reconocimiento de patrones, métodos de agrupación o "clustering", minería de datos y tecnologías de bases de datos.

1.2. Tipos de escáner para micromatrices de ADN

En la exploración de las micromatrices existen dos tipos principales de escáneres: confocales y CCD.

En los escáneres de microscopía confocal el lector típico está inspirado en el diseño original de Stanford, [10, 11, 12], siendo los principales fabricantes Affymetrix, Agilent, Axon Instruments y Perkin Elmer [6]. Cada sistema tiene características peculiares adicionales, pero el lector sigue siendo básicamente un microscopio confocal de fluorescencia con iluminación láser. Son sistemas muy complejos y caros. Surgen problemas al desplazar durante la exploración la MMADN o el sistema de excitación / detección.

Por otro lado se han publicado trabajos [1, 2, 3, 4, 7, 8] que investigan alternativas a los escáneres confocales utilizando tecnologías CMOS o CCD. Todavía están en fase experimental, hay muy pocos sistemas comerciales de este tipo disponibles en el mercado ("Proximal CCD Imager" de la desaparecida Genometrix, Inc. [1]), suelen ser también caros y más lentos que los confocales. Al contrario que los confocales, los escáneres CCD muestran una gran diversidad. Hay dos corrientes: la convencional explora la MMADN con la CCD usada como cámara, y la otra usa la CCD como soporte sobre el que crecer la propia MMADN, explotando sus propiedades eléctricas. En el primer caso, más frecuente,

Este trabajo ha sido parcialmente subvencionado por el Ministerio de Educación, Cultura y Deportes a través de una beca FPU (Ref. AP2002-2716).

perdura el problema de desplazar la MMADN o la CCD. Las principales dificultades de esta tecnología son el enfoque y la resolución.

En ambos casos, confocal y CCD, la exploración de la MMADN es exhaustiva manteniendo la máxima resolución, sin tener en cuenta si la zona explorada tiene información útil (spots) o si sólo es fondo del porta (la fluorescencia del fondo es el principal factor que limita la discriminación de la señal [5]). Estas metodologías dan lugar a imágenes muy ruidosas que ocupan casi 100 MB por canal (Cy5-rojo y Cy3-verde).

2. SISTEMA ESCÁNER MULTI-RESOLUCIÓN

El objeto de este trabajo es una metodología de exploración de micromatrices de ADN: una primera etapa a baja resolución localiza el objetivo y después una segunda, a alta resolución, lo fija y explora. La idea consiste en emular las dos foveas de que disponen las aves rapaces, cooperantes a la hora de buscar y seguir una presa [13]. Una “fovea” “ve” el conjunto, los máximos de los puntos (*Ubicación*), y otra inspecciona cada punto concreto. El sistema combina un modo de exploración tosca, que “sobrevuela” la MMADN para localizar puntos, con un modo de exploración fina, que mide con precisión moviéndose en entornos restringidos alrededor de las direcciones localizadas.

2.1. Componentes del sistema

El método consiste en dos etapas (fig. 1): una exploración de baja resolución (EBR) y una de alta resolución (EAR). Para cada canal de fluorescencia, la MMADN es explorada primero en la etapa EBR, que determina una rejilla de “Regiones Margen” (RM), y después por la EAR que explora únicamente aquellas zonas de la MMADN que caen dentro de las RM, dando lugar a la imagen final. Cada subsistema desempeña una labor concreta cuyas funciones se detallan a continuación:

- *Funciones del Subsistema de Baja Resolución:*

Toma como entrada la MMADN impresa en el porta del microscopio.

Lleva a cabo una primera exploración a baja resolución: cada punto sonda es representado con muy pocos píxeles (~10) de “gran” tamaño (~100 µm).

Presenta al usuario una imagen preliminar, útil para detectar errores en la fase de preparación química de la MMADN (presencia / ausencia de spots de control), evaluar la eficacia del lavado o detectar precozmente espurios de gran tamaño (partículas de polvo).

Detecta saturaciones en la imagen, información muy útil para implementar un Control Automático de Ganancia (CAG) en la EAR.

Crea la rejilla de RM dentro de las cuales se encuentran los spots (fig. 2): durante la impresión de la MMADN, el dispositivo impresor, o “pin”, está sometido a imprecisiones, a las que se une el hecho de que el volumen de líquido depositado es variable, y éste puede esparcirse más o menos al entrar en contacto con la superficie de la MMADN. Estos efectos dan lugar a una desviación (Δ) de la posición canónica que le correspondería idealmente a cada sonda, de tal forma que el radio de la Región Margen es:

$$R_{RM} \approx (R_{SPOT} + H) \times Q \quad (1)$$

Donde H representa la holgura del pin que depositó la muestra a analizar en la MMADN (debido a vibraciones o imprecisiones del robot impresor), y Q es un coeficiente de

difusión que recoge el efecto combinado de la variabilidad del volumen depositado y del mayor o menor “esparcimiento” de la gota que contiene la muestra sobre la superficie de la MMADN. A partir de la imagen de baja resolución, se crean perfiles de intensidad de Filas y Columnas, y de la localización de los picos de estos perfiles se determina la posición de los centros de las RM, es decir las coordenadas de la mencionada rejilla.

La EBR está dotada de cierta inteligencia para que descarte, como elementos de la rejilla, espurios de tamaño similar al de los puntos sonda y que impliquen una distorsión de la configuración esperable de la rejilla.

- *Funciones del subsistema de Alta Resolución:*

Toma como entradas la MMADN impresa en el porta del microscopio, la rejilla de RM creada por la EBR y la información necesaria para el CAG sobre puntos de la imagen con problemas de saturación.

Lleva a cabo una segunda exploración a alta resolución: cada punto sonda es representado con muchos píxeles (~100) de “pequeño” tamaño (~5 µm).

Su campo de actuación se restringe a las RM.

En cada RM forzosamente unos píxeles son del spot y otros del fondo. Lo interesante es que es fondo “local”, con lo que, al adquirir la imagen, se pueden determinar los umbrales de segmentación y calcular los cocientes Señal / Fondo. Así, el procedimiento global del escáner lleva implícito este proceso combinado de *Ubicación* y *Presegmentación*: las RM son las semillas de la fase de *Segmentación* del análisis de imagen.

Tiene un CAG que controla de forma individual en cada spot la intensidad de iluminación y / o la ganancia de sensores.

No todas las zonas del porta saturan la imagen, con lo que parece sensato que las ganancias de los sensores y las intensidades de iluminación se ajusten a las necesidades locales de la zona que estén explorando en cada momento. Así se evita la “desaparición” de spots poco expresados (en términos genéticos) en algunas zonas de la MMADN, debido a la disminución forzada de las intensidades o de las ganancias para evitar la saturación de spots más expresados. Actualmente se utilizan los 16 bits de señal para codificar los brillos absolutos y después obtener la relación de brillos de un canal a otro, cuando lo que realmente interesa codificar con precisión de 16 bits es esa relación de brillos.

Presenta al usuario la imagen final y la almacena.

2.2. Ventajas del sistema escáner multi-resolución

Al llevar a cabo una **exploración “personalizada”** de cada spot, permite implementar la ganancia variable de los sensores, el **CAG**, imposible de otro modo. El sistema es **más rápido** que los convencionales ya que más de la mitad de la MMADN suele ser fondo que no necesita ser explorado a alta resolución. Ésta es la principal ventaja del escáner: aporta inteligencia al proceso de exploración, ya que se concentra en la parte de la MMADN que aporta información. Además facilita el posterior análisis de imagen, al eliminar el ruido de la parte de fondo inútil de la imagen, de una forma imposible para el resto de escáneres actuales.

El **tamaño del fichero en disco se reduce** drásticamente. En la superficie de una MMADN se pueden distinguir claramente 2 zonas que no interesa explorar a alta resolución (no sólo porque no contengan información sino porque además contienen ruido en forma de polvo y otros espurios). Son los fondos **intrabloque** y **extrabloque** (fig. 3).

- El fondo **intrabloque** es la superficie acumulada de huecos intersticiales entre cada 4 puntos sonda más los medios huecos intersticiales presentes entre cada par de puntos sonda pertenecientes a la frontera del bloque.

La superficie de un hueco **intersticial** viene dada por:

$$f_{intra} = D^2 - \pi R_{RM}^2 \quad (2)$$

y el número de huecos intersticiales es:

$$NH = B \times [(N - 1) \times (N + 1)] \quad (3)$$

donde D es la distancia entre spots (centro a centro), R_{RM} es el radio de la Región Margen (si $\langle R_{SPOT} \rangle$ es el radio medio del spot, R_{RM} varía entre $R_{RM\ MIN} = \langle R_{SPOT} \rangle = 0.19\ mm$ y $R_{RM\ MAX} = D/2 = 0.34\ mm$), B es el número de bloques, N es el número de columnas o filas de spots por bloque (suponiendo bloques cuadrados). Con lo que la superficie total del fondo intrabloque es:

$$F_{intra} = NH \times f_{intra} \quad (4)$$

En el caso de una MMADN típica ($N = 24$; $D = 0.69\ mm$; $\langle R_{RM} \rangle = 0.27\ mm$) $F_{intra} = 1725\ mm^2$. Teniendo en cuenta que cada bloque mide (de acuerdo con los valores típicos propuestos) $16.04 \times 15.97\ mm = 256.16\ mm^2$, el $(143.75 / 256.16) \times 100 = 56.12\ %$ de la superficie de cada bloque es fondo intrabloque.

- El fondo **extrabloque** es la superficie exterior a los bloques de puntos sonda. Depende del tamaño del porta y de la configuración de la MMADN (número y tamaño de los bloques de puntos sonda). Por ejemplo, en el caso de una MMADN típica (de área total: $118.12 \times 40\ mm$) de 12 bloques dispuestos en 2 columnas y 6 filas $F_{extra} = 1650.88\ mm^2$. Es decir, el $(1650.88 / 4724.8) \times 100 = 34.94\ %$ de la superficie del porta es fondo extrabloque.

- La superficie de **fondo total** ahorrada (en el caso de una MMADN típica con los parámetros supuestos) es $F_{Total} = F_{intra} + F_{extra} = 1725 + 1650.88 = 3375.88\ mm^2$. El sistema escáner multi-resolución ahorra por término medio explorar superficie inútil de la MMADN en un porcentaje igual al 71.45%. **En el mejor caso**, las RM coinciden con el punto sonda ($R_{RM\ MIN}$) $F_{Total\ MAX} = 2484 + 1650.88 = 4134.88\ mm^2$, **Ahorro_{MAXIMO} = 87.51%**. **En el peor caso**, RM tangentes ($R_{RM\ MAX}$) $F_{Total\ MIN} = 759 + 1650.88 = 2409.88\ mm^2$, **Ahorro_{MINIMO} = 51%**.

3. CONCLUSIONES

Este trabajo presenta un nuevo método de exploración para escanear micromatrices de ADN, aportando inteligencia al proceso, empleando alta resolución de forma selectiva, sólo allí donde sea menester, allí donde resida la información útil. Un enfoque multi-resolución nos ofrece la clave para alcanzar tal propósito.

Esta metodología ofrece la posibilidad de incluir un CAG durante la EAR, lo que evita la saturación de los spots muy expresados y la pérdida de la información de aquellos que lo están poco.

También permite eliminar ruido de la señal de fondo al explorar sólo las zonas que contienen spots, descartando las zonas de fondo inútil que sólo albergan polvo y otros espurios. Cada RM contiene píxeles que corresponden al fondo local de cada spot.

El método permite ahorrar la exploración a alta resolución de entre el 51% y el 87% de la superficie de la MMADN.

4. AGRADECIMIENTOS

La MMADN aquí mencionada, y sobre la cual están basados los cálculos, es la “CMT™ Yeast-S228c Gene Array” de Corning Life Sciences. La información y experiencia sobre su configuración, función y uso práctico fue proporcionada por el Centro de Secuenciación Automática de ADN de la Facultad de Farmacia de la Universidad Complutense de Madrid (Dr. J. Arroyo Nombela y Dr. J.M. Rodríguez).

5. BIBLIOGRAFÍA

- [1] S.J.Jr. Brignac, R. Gangadharan, M. McMahon, et al., “A Proximal CCD imaging system for high-throughput detection of microarray-based assays”, *IEEE Engineering in Medicine and Biology Magazine*, 18(2), pp. 120–122, Mar-Abr 1999.
- [2] R. Cubeddu, C. D’Andrea, A. Pifferi, et al., “Time resolved reading of DNA microarray by an intensified CCD camera”, *2000 Conference on Lasers and Electro-Optics Europe, Conference Digest*, p. 1, 2000.
- [3] M. Eggers, M. Hogan, A. Mallik, et al., “A versatile biochip for gene-based diagnostics”, *ELECTRO '96. Professional Program. Proceedings*, pp. 87–92, 1996.
- [4] J.P. Golden, F.S. Ligler, “A comparison of imaging methods for use in an array biosensor”, *Biosensors and Bioelectronics*, 17, pp. 719-725, Mar 2002.
- [5] D.J. Graves, H.J. Su, S. Addya, et al., “Four-laser scanning confocal system for microarray analysis”, *Biotechniques* 32(2), pp.346-354. Feb 2002.
- [6] A.J. Holloway, R.K. van Laar, R.W. Tothill et al., “Options available —from start to finish— for obtaining data from DNA microarrays II”, *The Chipping Forecast, Supl. Nature Genetics*, 32, pp. 481-489. Dic 2002.
- [7] J.B. Lamture, K.L. Beattie, B.E. Burke, et al., “Direct detection of nucleic acid hybridization on the surface of a charge coupled device”, *Nucleic Acids Res* 11;22(11), pp. 2121-2125, Jun 1994.
- [8] V.N. Morozova, A.V. Gavryushkinb y A.A. Deev, “Direct detection of isotopically labeled metabolites bound to a protein microarray using a charge-coupled device”, *Journal of Biochemical and Biophysical Methods* 51(1), pp. 57-67, Mar 2002.
- [9] “The Chipping Forecast”, *Supl. Nature Genetics*, 21(1), pp. 1-60. Enero 1999.
- [10] M. Schena, D. Shalon, R.W. Davis, et al., “Quantitative monitoring of gene expression patterns with a complementary DNA microarray”, *Science* 270, pp. 467-470, 1995.
- [11] D. Shalon, S.J. Smith y P.O. Brown, “A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization”, *Genome Res.* 6, pp. 639-645, 1996.
- [12] <http://cmgm.stanford.edu/pbrown/scanner.html>, Stanford, 2002.

[13] H.D. Wolpert, "Life lessons", *oemagazine, The monthly publication of SPIE*, 2(2), pp. 26–29, Feb 2002.

[14] Y.H. Yang, M.J. Buckley, S. Dudoit et al., "Comparison of methods for image analysis on cDNA microarray data", Dept. of Statistics, UC Berkley, pp. 1–40, Nov 2000.

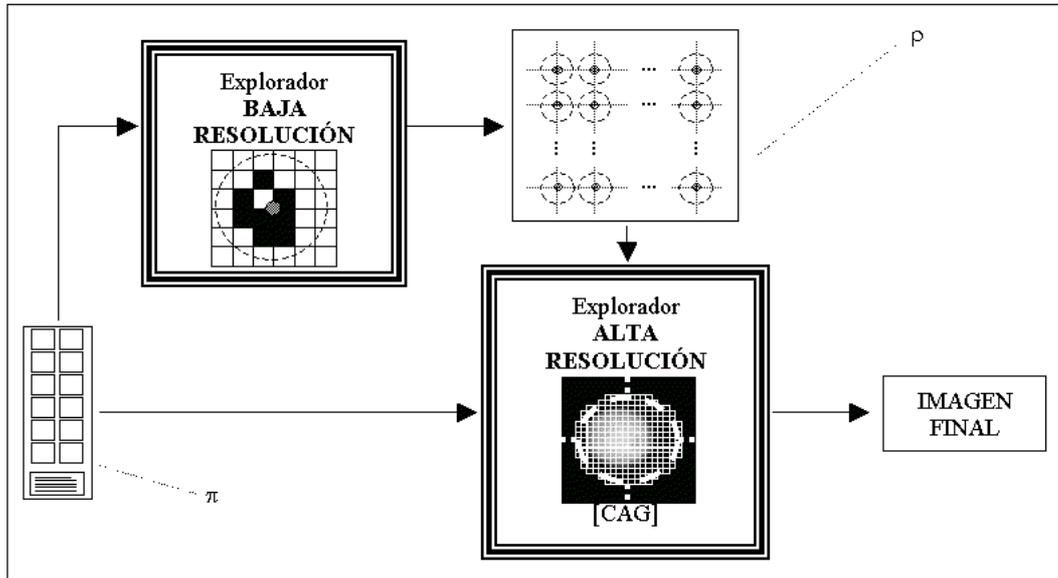


Figura 1: Diagrama de bloques del Sistema Escáner Multi-resolución. La MMADN (π) es explorada primero a Baja Resolución, y se determina la rejilla de Regiones Margen (ρ), después el subsistema de Alta Resolución, dotado de un CAG, explora sólo aquellas zonas de la MMADN que caen dentro de las Regiones Margen, dando lugar a la imagen final.

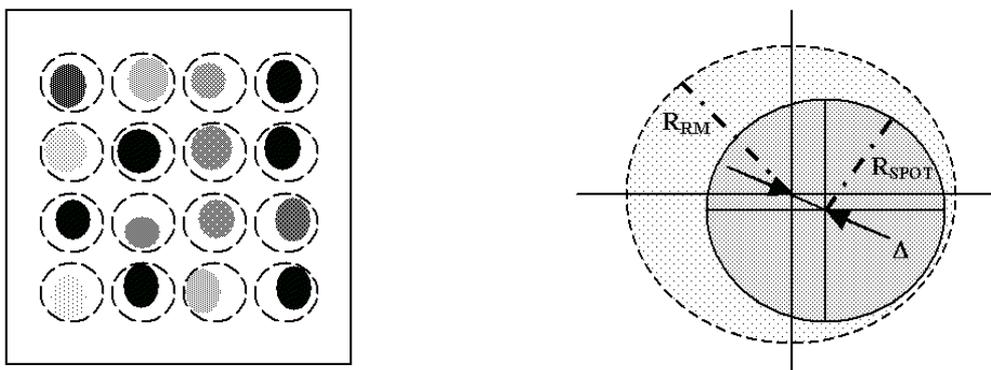


Figura 2: Aspecto real de la disposición, idealmente matricial, de los spots y detalle de la Región Margen. Los spots varían en tamaño y posición con respecto al caso ideal (en trazos discontinuos aparecen las RM correspondientes). A la derecha, un spot y su RM asociada (círculo circunscrito). Se muestran los radios (R_{SPOT} y R_{RM}) y la distancia entre centros (Δ).

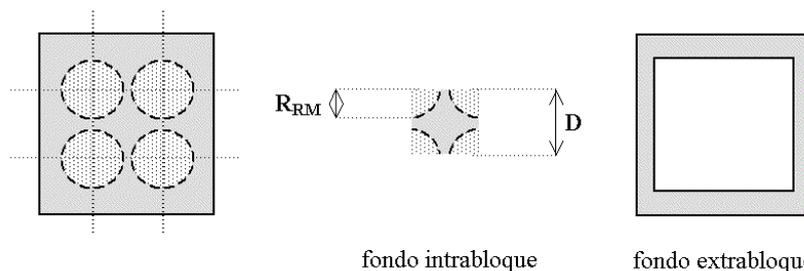


Figura 3: Zonas de fondo exteriores a las RM. Caso sencillo hipotético con tan sólo 4 spots. Existen dos tipos de fondo: el intrabloque, situado entre los spots, y el extrabloque, exterior al bloque de 2×2 spots de ejemplo. R_{RM} es el radio de la Región Margen, D es la distancia entre centros de Regiones Margen (es el paso de la rejilla de Regiones Margen).