

Del laboratorio a la clínica: desarrollo de perfiles genéticos en cáncer de mama

L. A. Henríquez-Hernández, L. Fernández-Pérez, N. Díaz-Chico¹

*Departamento de Ciencias Clínicas. ¹Departamento de Fisiología, Biología Molecular, Bioquímica y Genética.
Universidad de Las Palmas de Gran Canaria. Instituto Canario de Investigación del Cáncer (ICIC).
Las Palmas de Gran Canaria*

RESUMEN

El cáncer de mama es una enfermedad con un alto impacto en nuestra sociedad. La incidencia y prevalencia de esta patología es cada vez mayor. En su desarrollo intervienen factores genéticos, epigenéticos y ambientales. Aún hoy se desconocen la totalidad de procesos involucrados en el desarrollo, promoción y progresión del tumor mamario. Por todo ello, los esfuerzos en investigación dentro de este campo son cada día mayores. El objetivo no es sólo entender el comportamiento tumoral. Además se busca poder predecir el pronóstico, la respuesta a los tratamientos quimioterápicos y aumentar lo máximo los ratios de supervivencia. En la última década se ha desarrollado y perfeccionado la tecnología de *DNA microarray*, que permite realizar estudios genéticos y genómicos a gran escala. La aplicación de esta técnica al campo de la oncología ha generado muchos conocimientos y suscitado mucha expectativa. En esta revisión hemos hecho una aproximación técnica a la metodología de *DNA microarray*: sus principios básicos, en qué consiste, cómo interpretar los resultados, así como las diferentes estrategias que se pueden aplicar para visualizar, agrupar y comprender la inmensa cantidad de información generada con este tipo de experimentos. Este trabajo incluye un estudio detallado de las aplicaciones que tiene el *DNA microarray* en el cáncer de mama: estudios para mejorar el diagnóstico, para conocer mejor el pronóstico, para predecir la respuesta a los tratamientos, para identificar dianas terapéuticas y para generar test de diagnóstico clínico. Todas estas estrategias buscan llevar la metodología de *DNA microarray* del laboratorio a la práctica clínica rutinaria.

Palabras clave: Cáncer de mama. *Microarrays*. Perfil molecular. Factores pronóstico. Factores predictivos.

ABSTRACT

Breast cancer is a disease with a profound impact in our society. Incidence and prevalence of this pathology increase every year. In the breast cancer development genetic, epigenetic and environmental factors are involved. Many details of the entire process of initiation, promotion and tumoral progression are still unknown, and additional efforts are required. The main objective is not only to understand the tumour behaviour but also predict the prognosis, the response to chemotherapeutic treatments and increase the survival ratios are important goals. In the last ten years DNA microarray technology has been developed and improved. This is a high throughput put technology that permits genetic and genomic studies on a large scale. The implementation of DNA microarray in oncology has been generated a big amount of knowledge as well as a great expectancy. In this review we have carried out a technical approximation to DNA microarray methodology: basic principles, how to interpret the results as well as the different strategies available for visualize, cluster and understand the vast amount of information generated with this kind of experiments. We have described in detail the applications of DNA microarray to breast cancer: studies for improve the diagnosis, to predict the prognosis and the response to chemotherapy, to identify therapeutic targets and for to generate predictive test useful in the clinical practice. All these strategies are of translational meaning, trying to carry the DNA microarray from the laboratory to the daily clinical routine.

Key words: Breast cancer. Microarrays. Molecular profile. Prognostic factors. Predictive factors.

INTRODUCCIÓN

El cáncer de mama es una enfermedad compleja de origen multifactorial donde se ven involucrados factores genéticos, epigenéticos y ambientales. En las últimas décadas se ha producido un incremento significativo en el conocimiento de la genética del cáncer en general y del tumor mamario en particular. Se han descubierto numero-

genes relacionados con la patogénesis, el curso clínico y el pronóstico del cáncer de mama; como por ejemplo Her2/neu, p53, BRCA1/2 y un largo etcétera más. Durante años este tipo de estudios se han abordado de manera limitada, gen a gen, variable a variable. En la actualidad, y desde hace casi una década, los estudios genéticos y genómicos se realizan a gran escala, debido fundamentalmente al desarrollo de la tecnología de *DNA microarrays*. En esta revisión se pretende realizar un acercamiento global a esta metodología y encuadrarla en sus aplicaciones al cáncer de mama.

DNA MICROARRAY. PRINCIPIOS BÁSICOS

Esta tecnología surgió a mediados de la última década del siglo pasado y supuso una auténtica revolución en el campo de la biología molecular, puesto que permitía estudiar cientos de genes en un único ensayo, maximizando el tiempo y los resultados hasta niveles no conocidos hasta ese momento (1). Desde entonces la metodología se ha mejorado mucho y en la actualidad se puede estudiar el genoma completo de cualquier especie en un único ensayo, lo que implica medir decenas de miles de genes de una sola vez. Pero quizás donde mayor avance y perfeccionamiento ha habido sea en las herramientas informáticas necesarias para procesar toda esa información, así como en los programas de ordenador capaces de dar sentido a los datos generados y orientar los estudios posteriores; imprescindibles y necesarios.

La técnica de *DNA microarray* hace posible el estudio simultáneo de los niveles de expresión génica de miles de genes (2). Esto es posible mediante la adhesión a un soporte físico (de diversos materiales) de fragmentos específicos de ADN, a los que se le unirá por hibridación la muestra marcada con un fluorocromo (Fig. 1). Existen multitud de plataformas para realizar *microarrays* (3), que se diferencian entre sí por la manera de manufacturar las plataformas o por la distinta preparación de la muestra. En general, los *microarrays* se pueden dividir en dos grupos: de marcaje único (de la muestra) o de doble marcaje (Fig. 1).

En el primer caso cada muestra (control y problema) es marcada con el mismo fluorocromo e hibridada por separado en dos *microarrays*; en el segundo, cada muestra se marca con un fluorocromo diferente y se hibridan de manera conjunta sobre un único microchip. Este último sistema favorece las comparaciones directas entre un control y un problema, aunque hace más compleja las comparaciones entre más de dos muestras, necesitándose en ese caso una referencia común. Existen igualmente distintas estrategias para abordar los estudios de *microarray*, siendo determinante para el éxito de la investigación decidir el diseño correcto y que más convenga a nuestro diseño experimental (Fig. 2).

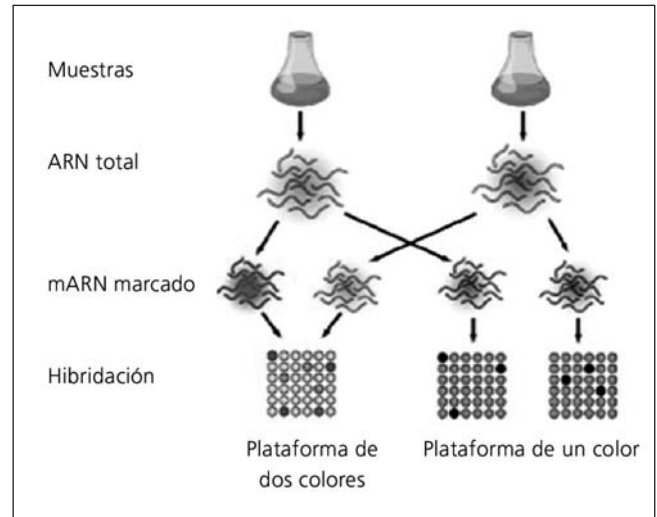


Fig. 1. Imagen que resume los pasos básicos de un experimento con microarray: la muestra, obtención del ARN de la misma y posterior marcaje con fluorocromos y por último la hibridación. Además se detallan las dos principales plataformas de microarrays, de dos colores y de un color.

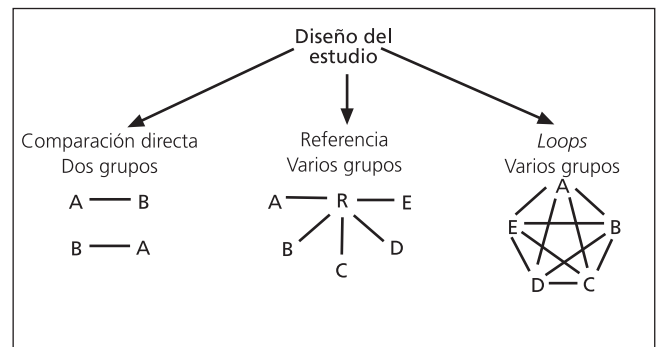


Fig. 2. Representación gráfica de los principales tipos de diseños experimentales para microarrays. La comparación directa entre dos grupos es la más sencilla. Además están las comparaciones de diversos grupos experimentales con un único grupo control (Referencia) y la comparación de todos los grupos con todos los grupos (Loop).

La fase de hibridación de la muestra con el microchip lleva varias horas (14-18 horas). Una vez transcurrido ese tiempo, los chips se lavan y se leen en un escáner de fluorescencia apropiado para *microarrays*. Se generan entonces una serie de imágenes que deben ser procesadas para obtener datos comprensibles de regulación génica. En cualquier diseño experimental son tan importantes los replicados técnicos como los biológicos. Estos últimos son especialmente relevantes en los experimentos con *microarrays* para obtener resultados significativos. Independientemente del tipo de plataforma usada, las diferencias entre las eficiencias de hibridación de las muestras hacen difícil extrapolar los niveles de fluorescencia de cada punto del chip perteneciente a cada gen, en niveles abso-

lutos de expresión de mRNA. Dichos niveles se expresan como un ratio que representa el cambio de expresión entre las condiciones estudiadas (control frente a problema) (4). La característica básica de este valor generado es que tiene una distribución asimétrica; entre 1 e infinito en caso de inducción y entre 0 y 1 en caso de inhibición. Estos ratios son generalmente transformados logarítmicamente antes de continuar con cualquier análisis posterior, resultando en una distribución de valores de expresión simétrica tanto en caso de inducción como de inhibición. No obstante, diferencias de eficiencia de marcaje, de cantidad de muestra o de ruido de fondo hacen que los datos transformados no puedan ser usados directamente para analizarlos; antes deben ser normalizados (5). En un principio los métodos de normalización se basaban en la mediana de la intensidad de la señal o bien en la expresión constante de algunos genes con independencia de las condiciones experimentales. Hoy día ambos métodos están obsoletos. En la actualidad se utilizan metodologías de normalización no lineal como el de Lowess (6), que equilibra la intensidad de la señal global de fluorescencia asumiendo que la mayoría de los genes no deben cambiar sus niveles de expresión (Fig. 3).

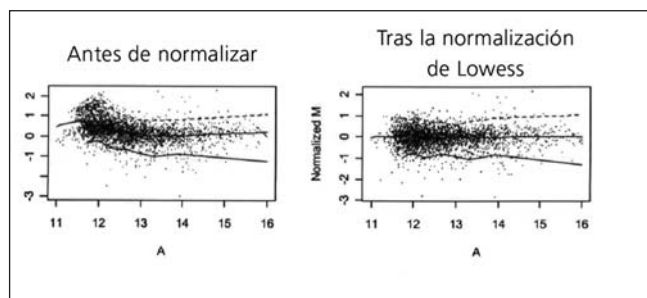


Fig. 3. Se representa una típica figura de microarray antes y después de aplicar la normalización de Lowess. Se muestra a M frente a A , donde M (\log_2 del ratio entre los fluorocromos) está en el eje Y , frente a A (la media del \log_2 de la intensidad) está en el eje X , y donde cada gen está representado por un único punto.

Este sistema de normalización debe ser considerado como óptimo, siempre y cuando no nos encontremos ante modelos experimentales que impliquen cambios severos e importantes en la maquinaria de transcripción y otros procesos celulares determinantes.

DIFERENCIAS SIGNIFICATIVAS DE EXPRESIÓN

Tras el análisis de imagen y las primeras normalizaciones de los datos se procede a analizar los mismos con el objetivo de conocer aquellos genes que se encuentren significativamente regulados. Este proceso requiere de potentes test estadísticos específicos para *microarrays*. El alto coste de este tipo de experimentos

hace que el número de replicados sea escaso, lo que supone un problema importante a la hora de realizar estadística. Los modernos programas informáticos de análisis de microchips solventan este tipo de problemas comparando la variación de expresión de cada gen en concreto con la variación global. Atrás quedan por tanto estrategias estadísticas aplicadas como el t-test, totalmente obsoletas y erróneas. El p-valor es otro parámetro que debe ser corregido en los test estadísticos. Si se diseña un experimento de *microarrays* con 20.000 genes y aplicamos un p-valor de 0,01, teóricamente obtendremos 200 genes significativamente regulados, cuando la realidad puede ser un número bastante menor. Por ello al p-valor se le aplica un test de corrección múltiple (MTC) que ajusta el p-valor de cada gen en función del número de replicados realizados. Este tipo de estrategias son bastante conservativas y pueden reducir el poder para reconocer resultados significativos. No obstante se asume que en este tipo de experimentos va a existir un número de falsos positivos. El parámetro estadístico que va a controlar esto se conoce como *false discovery rate* (FDR), y generalmente es del 5%. Es decir, el 5% de los genes significativamente regulados pueden ser falsos positivos. En función del objetivo final del estudio planteado se modificará el grado de restricción de los distintos parámetros estadísticos (7).

VISUALIZANDO DIFERENCIAS DE EXPRESIÓN

Los experimentos con *microarrays* generan una cantidad importante de datos numéricos que en muchas ocasiones son difíciles de interpretar de manera general. Por ello son muy útiles las representaciones gráficas derivados de los mismos. Los programas de análisis informático incluyen diferentes opciones para la visualización de los resultados. Uno de los procedimientos estándar de visionado de resultados es el *scatter plot*. Genes que se encuentren expresados de manera similar entre las condiciones ensayadas aparecerán representados cercanos a una diagonal establecida, mientras que aquellos que estén diferencialmente expresados aparecerán por arriba o por debajo de esta línea según estén inducidos o inhibidos. Los *scatter plots* se usan además para comparar resultados entre diferentes plataformas o para ensayos de reproducibilidad (8). Establecer un coeficiente de correlación entre los ratios derivados de experimentos diferentes es interesante ya que permite conocer aquellos genes con resultados más reproducibles, y por tanto más fiables una vez aplicados los test estadísticos de significación. Los diagramas de Venn son muy útiles para identificar relaciones biológicas significativas entre los genes regulados. Este tipo de diagramas solapan círculos con genes en común (Fig. 4), preferiblemente de manera escalada para tener así una mejor idea visual de la magnitud de la coincidencia.

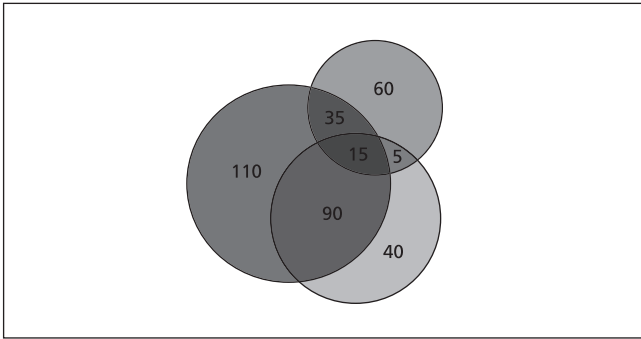


Fig. 4. Ejemplo de diagrama de Venn en el que se representan tres listas de genes. Las listas presentan genes en común que se representan como solapamientos entre ellas. Los valores indican número de genes.

El solapamiento entre las listas de genes debe ir acompañado de un p-valor derivado de test hipergeométricos (9). Este tipo de representación gráfica de los resultados debe realizarse una vez hayan finalizado los análisis estadísticos, para evitar falsos solapamientos y resultados erróneos. Los estudios que incorporan mayor número de replicados incrementan su potencial para identificar diferencias de expresión menores. Es decir, de manera ideal las comparaciones para originar un diagrama de Venn deben ser hechas a partir de listas de genes que posean el mismo número de replicados. Aunque existen otras estrategias para visualizar los resultados, estas son las más comunes y más usadas.

CLUSTERING O AGRUPAMIENTO DE LOS DATOS

A pesar de las maniobras estadísticas y de representación usadas, la cantidad de datos generados por este tipo de experimentos hace muy difícil identificar algún tipo de orden o estructura en los mismos, sobre todo si el número de microchips usados es grande. Para reducir la complejidad del conjunto de datos aparentemente sin sentido, es posible agrupar aquellos genes que muestren patrones de expresión similares en las diferentes condiciones experimentales de nuestro diseño; es lo que se conoce como *clustering* o agrupamiento. Existen numerosos métodos para realizar esto (10), pero de manera general se pueden agrupar en dos tipos: agrupamientos supervisados o no supervisados. Los primeros necesitan un grupo de muestras de clasificación conocida que se usará para clasificar un nuevo conjunto de datos generados. Este tipo de método es el usado en investigación del cáncer, por ejemplo para predecir la evolución clínica de pacientes nuevos en base a su perfil de expresión tumoral comparándolo con perfiles de expresión de pacientes de evolución conocida (11). Los agrupamientos no supervisados, por el contrario, no requieren ningún conoci-

to previo al análisis, y por eso son extensamente usados. Los agrupamientos jerárquicos pertenecen a este tipo de *clustering*, y permiten ir agrupando genes por niveles en función de la mayor similitud entre ellos. De esta manera se construye una estructura en forma de árbol donde grupos de genes con patrones de expresión similares quedan conectados (Fig. 5). Los agrupamientos jerárquicos permiten incluso agrupar los diferentes experimentos realizados además de agrupar a los propios genes de cada experimento. De esta manera se pueden establecer similitudes entre modelos experimentales diferentes, estrategia que es interesante en campos como la endocrinología.

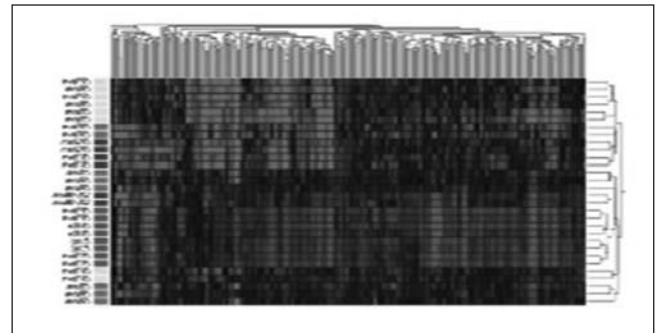


Fig. 5. Imagen que representa un cluster jerárquico a partir de experimentos con microarray. A la derecha se agrupan en forma de árbol los diferentes experimentos realizados y arriba se agrupan los genes en base a su similitud de expresión en los diferentes experimentos realizados.

Se recomienda no obstante tener precaución a la hora de interpretar los resultados obtenidos a partir de los agrupamientos jerárquicos, tomando siempre un punto de corte común a lo largo del árbol para que las relaciones observadas sean comparables y reales. Otro método de agrupamiento de genes bastante usado es el llamado *k-mean clustering*. Este método agrupa a todos los genes regulados de manera conjunta para después irlos separando en grupos más pequeños en función de la similitud de los patrones de expresión. El principal problema que tiene esta estrategia es que el número de grupos debe ser definido *a priori* por el investigador, y no siempre es el correcto. Se recomienda usar primero otro tipo de agrupamientos que permitan después hacer una buena elección en el número de grupos generados por el *k-mean clustering*. Los análisis de los componentes principales (PCA o *principal component analysis*) son un buen método previo al *k-mean clustering*. Este tipo de análisis busca reducir la complejidad de los datos generados centrándose en los datos con mayor variación, a los que considerará como los *componentes principales* del total. Como resultado, un conjunto de datos complejo que contenga información de múltiples experimentos podrá ser visualizado en una gráfica bi- o tridimensional con los genes que posean mayores diferencias de expresión. Existe un método de partición más sofisticado conocido

como mapas de organización propia (SOM o *self-organizing maps*), que tienen la ventaja de que además describen la relación entre los grupos generados. Por último cabe destacar el *signature algorithm*, otro método de agrupamiento que permite que determinados genes pertenezcan a más de un grupo, o incluso, que un gen no pertenezca a ninguno (12). Este tipo de método permite una representación más natural de los datos y es muy útil a la hora de valorar regulaciones génicas biológicas donde un mismo gen puede estar participando en diferentes rutas de regulación controladas por diferentes factores de transcripción, por ejemplo. Cada método de agrupamiento tiene sus ventajas y sus inconvenientes, y debe ser seleccionado uno u otro en base a lo que se desea conocer (13,14). Es habitual utilizar varias estrategias de agrupamientos para evitar errores en los grupos y tener una mayor fiabilidad a la hora de interpretar los resultados.

CLASIFICACIÓN FUNCIONAL DE LOS GENES (ONTOLOGÍA)

Las listas de genes generadas, una vez estos han sido analizados estadísticamente y agrupados, no tienen información relativa a la actividad biológica de los mismos. Esto se puede realizar anotando los genes en bases de datos públicas como Unigene o ENSEMBL. Desde el año 2000 se ha puesto en marcha el proyecto Ontología Génica (GO o *gene ontology*) que tiene como objetivo proporcionar una estructura y un lenguaje común que pueda ser usado sistemáticamente para describir los diferentes genes. GO consiste en 3 ontologías que clasifican a los genes en función de los procesos biológicos en los que esté involucrado, la función molecular y el componente celular. De esta manera es posible asignarle a cada gen una función putativa (9,15,16). Además la asignación de un gen a una clase determinada viene acompañada de un p-valor, que es fiable siempre que esa clase en particular no posea un número pequeño de genes asignados (< 5). Como un mismo gen puede estar en múltiples categorías se puede aplicar un FDR a la clasificación ontológica para que el valor p sea más realista. GO es una buena herramienta para descubrir nuevos genes involucrados en determinados procesos biológicos. Si un gen desconocido aparece repetidamente en un grupo de genes relacionados con una determinada función, podemos inferir que ese gen desconocido va a estar participando en la función vista.

ESTRATEGIAS FUTURAS

La tecnología conocida como *DNA microarray* está disponible desde finales de los noventa del siglo pasado. La tecnología en sí, sus aplicaciones y sus métodos de análisis están en continuo desarrollo. El número de genes por chip ha crecido exponencialmente y se ha pasado de

tener unas pocas centenas a decenas de miles e incluso genomas completos. Ya no sólo se trabaja con expresión, sino que existen chips con *splicings* alternativos, con polimorfismos, con exones... por lo que los métodos de análisis son cada vez más complejos y sofisticados; y deben seguir mejorándose. A día de hoy es posible que los métodos actuales queden totalmente obsoletos dentro de unos años. No obstante, el uso de esta tecnología ha permitido aumentar el conocimiento científico genético enormemente, también en campos como la oncología. Aplicar este conocimiento generado en los laboratorios a la práctica clínica es uno de los mayores desafíos de la comunidad científica.

PERFILES DE EXPRESIÓN GÉNICO EN CÁNCER DE MAMA

Desde la introducción de la tecnología de *DNA microarray* a mediados de los noventa, el cáncer de mama ha sido el carcinoma más estudiado en relación a sus perfiles de expresión (2). La alta incidencia así como la importancia social de este tipo de cáncer ha propiciado el uso de esta tecnología con el objeto de identificar grupos de genes que ayuden a comprender la biología tumoral, ayuden a maximizar los tratamientos y permitan anticiparse al pronóstico; así como para descubrir nuevas dianas terapéuticas. El objetivo final es trasladar todo el conocimiento generado a la práctica clínica mediante el uso de microchips de pequeño tamaño y fácil procesado, que contengan los genes considerados relevantes y que aportarán una información genética individual útil para el paciente. Para ello hay que superar las controversias creadas en torno a esta metodología y que ponen en tela de juicio su validez y reproducibilidad. La validación de resultados en muestras de pacientes de cohortes independientes es un paso indispensable que permitirá el acceso de esta técnica a la práctica clínica rutinaria (17). Debido a que los datos generados en los experimentos con *microarrays* llevan un importante número de test estadísticos y análisis ulteriores, a veces es difícil comparar resultados de modelos similares generados en laboratorios diferentes. Por ello se ha llegado al consenso internacional de que los datos crudos, sin procesar, deben estar disponibles y ser publicados para que puedan ser accesibles y comparados en las mismas condiciones en diferentes laboratorios del mundo sin tener que reproducir la fase experimental.

ESTUDIOS PARA CONOCER LA BIOLOGÍA DEL CÁNCER DE MAMA

El conocimiento de la biología del cáncer de mama no ha crecido demasiado desde hace unos años a pesar de que se conocen datos de expresión de decenas de miles de genes procedentes de diversas líneas tumorales y espe-

cimenes de tumor. En el año 2000, Perou y cols. publicaron en *Nature* un artículo en el que demostraban que diversos patrones de expresión determinaban una clasificación biológica de los tumores mamarios (18). Para ello agruparon los genes usando métodos no supervisados de *cluster* jerárquicos. Usaron 65 muestras procedentes de 42 pacientes. Este estudio ha quedado validado posteriormente por Sorlie y cols. (19). El trabajo de Perou ha permitido conocer diversos subtipos tumorales en base a los perfiles de expresión de los genes, a saber: luminal A, luminal B, HER2 (EGFR2) sobreexpresado, basal y normal. Los 3 más agresivos desde el punto de vista clínico son el luminal B, HER2 y el basal. Estudios posteriores han corroborado estos datos usando plataformas de microchips diferentes y análisis distintos en diferentes laboratorios (20,21). El grupo de genes característico de cada subtipo tumoral es reproducible en todo caso. Es importante destacar el hecho de que los tumores del subtipo basal no disponen en la actualidad de tratamientos contra dianas terapéuticas (22). Los estudios de *microarrays* para este subtipo tumoral han sido ampliados con estudios inmunohistoquímicos. El patrón de expresión se correlaciona con una serie de copias genómicas aberrantes conocidas como CGH (23). Chang y cols. esbozaron una interesante teoría posteriormente comprobada que sugería que los mecanismos moleculares propios de los sistemas de reparación de heridas son los mismos que aparecen en los procesos de invasión y progresión tumoral (24). Este patrón de expresión se relaciona con un peor pronóstico de carcinomas mamarios. El mismo patrón predice lo mismo en el caso del cáncer gástrico y de pulmón. Algo similar se ha descubierto para el caso de los genes relacionados con respuesta a hipoxia. El pronóstico es peor si el mismo patrón de expresión aparece en el tejido tumoral. Un estudio publicado por Clarke y cols. en 2003 proponían un conjunto de 186 genes, entre los que estaban genes relacionados con la apoptosis, con la proliferación y con la quimiotaxis, y a los que llamó "patrón de invasividad", y que mostraban una fuerte asociación con el pronóstico de la enfermedad (25). Este patrón de invasividad es válido para otros tipos tumorales como el de pulmón o próstata. Los datos generados por Chang y Clarke en conjunto permiten realizar una estratificación nueva del riesgo de cáncer. Un 5-10% de los tumores de mama son de carácter hereditario. En este caso, las mutaciones germinales en los genes BRCA1/2 se consideran determinantes en la herencia del tumor. Conocer las diferencias en los patrones de expresión de los tumores hereditarios frente a los esporádicos es interesante y determinante. Se ha observado que existe un conjunto de genes relacionados con la adhesión celular y la migración (lamininas, colágenos y fibronectinas) que caracterizan a los tumores con mutaciones en BRCA1 y no a los tumores esporádicos (26). Es más, tumores de tipo familiar negativos para mutaciones en BRCA1/2 tienen perfiles de expresión diferentes a aquellos del mismo tipo que sí poseen alteraciones en estos dos genes (27). El conjunto de

resultados ha abierto nuevas vías de investigación para intentar comprender el comportamiento biológico de los tumores mamarios de tipo familiar hereditario.

ESTUDIOS PARA MEJORAR EL DIAGNÓSTICO DEL CÁNCER DE MAMA

Mientras que los perfiles de expresión proporcionan información relevante que permite incrementar el conocimiento acerca del desarrollo y la promoción tumoral, su valor no está tan comprobado como herramienta clínica. La aplicación de la tecnología de *microarrays* al diagnóstico constituye la mejor manera de ir introduciendo esta metodología dentro de la práctica clínica. Las alteraciones en p53 juegan un papel relevante en la progresión tumoral y en la resistencia a los tratamientos. Las técnicas inmunohistoquímicas no diferencian deleciones y mutaciones, por lo que son insuficientes en investigación. Affymetrix, una empresa especializada en *microarrays* que tiene desarrolladas sus propias plataformas y programas de análisis, ha descubierto un patrón de expresión compuesto por 32 genes relacionados con proliferación y crecimiento celular, transporte de iones, transcripción y biología del cáncer de mama, que diferencia los tumores con p53 sin mutar y mutado. Se distinguen grupos de datos independientes con capacidad para predecir el progreso clínico con mejores resultados que los obtenidos con solo el estado de mutación de p53 (28). Los estudios de expresión se han relacionado en diversos trabajos con la histología del cáncer de mama. El carcinoma lobular infiltrante prolifera generalmente más despacio que el ductal, y se caracteriza por tener pérdida de la expresión de E-caderina (29) así como ser más positivos a los receptores de estrógenos y progesterona con mayor frecuencia (30). Ambos tipos tumorales a nivel histológico pueden verse solapados y no ser siempre distinguibles. Esto en la práctica clínica no es importante porque el tratamiento va a ir en dependencia de otros factores como el estatus de HER2 o de los receptores hormonales. Un estudio con *microarrays* comparó 17 carcinomas lobulares frente a 106 ductales y descubrió un grupo de 11 genes que diferenciaban claramente un tipo tumoral del otro (E-cadherin, survivin, cathepsin B, TP11, SPRY1, SCYA14, TFAP2B, thrombospondin 4, osteopontin, HLA-G y CHC1) (31). Estos resultados son prometedores aunque aún necesitan validación puesto que otros estudios no han encontrado diferencias tan claras entre los dos tipos tumorales a nivel de expresión. Hannemann y cols. hicieron una comparación entre 40 casos de carcinoma ductal *in situ* frente a 40 tumores invasivos, y obtuvieron una lista de 35 genes que distingue ambos grupos (32). Además consiguieron distinguir los carcinomas ductales bien diferenciados de los peor diferenciados a nivel de expresión. Estos datos están validados aunque precisan de validaciones externas y traslado a la clínica. La determinación de BRCA1/2 es compleja y laboriosa. Se sabe que

fibroblastos con estos genes mutados presentan diferentes perfiles de expresión después de ser irradiados (33). La teoría es que si estos genes están mutados, deben producirse cambios en el ARN que pueden ser medidos por técnicas como el *microarray*. Esta hipótesis está en desarrollo y es posible que en un futuro inmediato se descubra algún grupo de genes que diferencie a los pacientes con BRCA1/2 mutado de los no mutados y que además tenga relevancia en la clínica.

ESTUDIOS PARA CONOCER MEJOR EL PRONÓSTICO DE LA ENFERMEDAD

El grado histológico es uno de los factores pronóstico más importantes en el cáncer de mama (34). Entre el 30-60% de los tumores mamarios están moderadamente diferenciados (grado 2), y presentan un pronóstico medio; por lo que son un grupo numeroso de tumores sobre los que por ejemplo, pocas decisiones se puede tomar respecto al tratamiento. Sotiriou y cols., usando tecnología Affimetrix, identificaron 97 genes que diferenciaban perfectamente tumores de grado 1 de los de grado 3. Estos genes, relacionados con progresión y proliferación celular, se encontraban más expresados en los tumores más pobremente diferenciados. Usando estos datos clasificaron los tumores grado 2 en dos grupos de riesgo, relacionados con mayor o menos supervivencia libre de enfermedad con independencia de otros parámetros clínicos o el estatus de receptores hormonales. El cáncer de mama de tipo luminal A posee unos niveles de expresión génica menores a otros de tipo luminal: lo que sostiene la idea de que los niveles de expresión de genes relacionados con proliferación pueden definir el peor pronóstico del subtipo luminal B. Dai y cols. demostraron que el nivel de expresión de genes relacionados con proliferación poseía mayor relevancia a la hora del pronóstico que otros parámetros como la edad o el estado de los receptores hormonales (35). No obstante, conocer el grado histológico por parte de los patólogos es algo tan barato y rutinario que hoy en día es impensable que pueda ser sustituido por otro tipo de técnicas como la de *microarrays*. El Instituto del Cáncer de Holanda ha conseguido identificar un conjunto de 70 genes relacionados con el pronóstico tumoral de pacientes metastásicas (n = 78) menores de 55 años, con 5 años de evolución de la enfermedad, con ganglios linfáticos negativos y que no habían recibido quimioterapia adyuvante (36,37). Los genes están relacionados con el ciclo celular, la metástasis, la angiogénesis y la invasión. Estos resultados se han validado en una cohorte de pacientes con ganglios linfáticos tanto positivos como negativos, así como con tratamiento adyuvante y sin él (n = 295). Este grupo de 70 genes ha sido explorado por otros grupos de investigación siendo validados sus resultados y ampliados a otros tipos de situaciones clínicas. Hay que destacar que este grupo de genes se conserva tanto en el caso de tumores primarios como en tumores

metastáticos, lo que sostiene la idea de que la capacidad de metástasis es inherente al tumor, y que el fenómeno de diseminación a distancia únicamente refleja la biología específica del tumor primario (38). Se encuentran publicados los perfiles genéticos que predicen la metástasis del cáncer de mama a hueso y a pulmón (39,40), lo que es de una importancia vital a la hora de aumentar los controles y revisiones de estas pacientes así como para ajustar los tratamientos (por ejemplo uso de bifosfonatos en caso de riesgo de metástasis ósea). Wang y cols., usando chips de Affimetrix en un grupo de 115 pacientes con cáncer de mama, identificaron un grupo de 76 genes con capacidad para predecir la posible recaída de pacientes con ganglios linfáticos negativos no tratadas (41). Los hallazgos se han optimizado diferenciando perfectamente aquellos tumores receptores de estrógeno positivo de los negativos. Estos genes están relacionados con procesos de replicación y reparación del ADN, ciclo celular, apoptosis y respuesta inmune. Este conjunto de 76 genes ha sido validado por otras instituciones (42). Los modelos pronóstico tienen una serie de limitaciones inherentes a ellos mismos. En principio, diferentes poblaciones y distintos análisis estadísticos modifican un determinado pronóstico; lo que hace imprescindible el uso optimizado de análisis supervisados. Pero más importante es tal vez el componente no controlable de esta serie de modelos. La dieta, los hábitos alimenticios, el ejercicio, y una alta variabilidad personal junto con el "destino" de la enfermedad, hacen que los modelos pronóstico estén lejos de la perfección.

ESTUDIOS PARA PREDECIR LA RESPUESTA TUMORAL A LA QUIMIOTERAPIA

El objetivo clínico más importante es tal vez usar la tecnología de *DNA microarrays* para intentar predecir la respuesta individual a los tratamientos sistémicos con quimioterápicos. Se trata por tanto de utilizar los perfiles de expresión genéticos para elaborar una predicción a la respuesta a los tratamientos. La doxorubicina y el 5-fluorouracilo inducen cambios de expresión tanto en líneas celulares como en pacientes (43). Así se ha observado que el subtipo tumoral luminal, en respuesta a estos tratamientos, modifica la expresión de genes relacionados con el control del ciclo celular que no son modificados en el caso del subtipo basal; lo que indica que la biología intrínseca del tumor va a condicionar su respuesta a las drogas. Usando distintas líneas celulares sensibles a la doxorubicina y utilizando microchips se ha conseguido descubrir un grupo de 79 genes que provocan resistencia a este tratamiento (44). Dentro del conjunto de genes destaca P-glicoproteína (codificada por el gen MDR1), y se ha correlacionado este patrón de expresión con una menor tasa de supervivencia. Estos resultados deben ser clínicamente validados, pero desde luego vuelven a demostrar que los niveles de expresión de determinados genes influirán en la respuesta a los diferentes tratamientos. Se han tratado igualmente a células de cáncer de mama con

estrógenos con el objeto de dilucidar un patrón de expresión de respuesta a esta droga (45). Se trata de 822 genes entre los que están el receptor de progesterona y otros genes relacionados con la proliferación celular, y que tienen capacidad para predecir la respuesta al tratamiento con tamoxifeno en pacientes con receptor de estrógenos positivo. Lo ideal en estudios futuros prospectivos, es usar esos modelos para intentar validar los diferentes patrones de expresión descubiertos, más que para intentar descubrir nuevos perfiles. Por ejemplo, un estudio con 82 pacientes tratadas de manera neoadyuvante con paclitaxel-fluorouracilo-doxorubicina-ciclofosfamida y que tuvieron una respuesta patológica completa, tenían unos niveles de expresión de los genes propios de los subtipos HER2 y basal; en comparación con los subtipos normal y luminal (46). Pero al explorar las listas de genes de los tipos basal y HER2 y realizar análisis supervisados, se observa que no existe correlación entre ambas listas de genes, lo que vuelve a indicar que tumores diferentes emplean mecanismos distintos de resistencia a drogas. Los tumores positivos al receptor de estrógenos (ER) se asocian generalmente a un mejor pronóstico y una mejor respuesta al tratamiento, sobre todo de tipo hormonal (47). No obstante, un porcentaje de pacientes recaen y en torno a un 40% desarrollan resistencia al tamoxifeno (25). La respuesta a estos fenómenos puede estar en los perfiles de expresión génica de los tumores ER-positivos. Jansen y cols. en 2005 utilizaron *DNA microarray* en muestras de tumor primario de mama de pacientes que habían recibido tamoxifeno. Las muestras tenían una antigüedad de entre 10 y 20 años. Identificaron 81 genes relacionados con la acción de los estrógenos, la apoptosis, la formación de la matriz extracelular y la respuesta inmune; que diferenciaban a pacientes sin recidiva de las recidivadas. Esta lista de genes se redujo a 66 tras una serie de estudios supervisados y de validación. Estos resultados son interesantes, ya que con una fiabilidad del 80% se puede predecir la resistencia al tamoxifeno y la posibilidad de recaída a largo plazo. No obstante, aunque existen estudios similares que dicen aproximadamente lo mismo, hay otros que discrepan en los datos, por lo que de nuevo se necesita validar todo este tipo de investigaciones para llevarlas a la clínica en un futuro próximo (48). La introducción de técnicas de PCR cuantitativa en tiempo real, que posibilitan la introducción de controles internos y que son más sensibles y específicas que el *DNA microarray*, permitirá ir validando con seguridad todos estos resultados interesantes y determinantes para el paciente oncológico.

ESTUDIOS PARA IDENTIFICAR DIANAS TERAPÉUTICAS

Las herramientas utilizadas hasta ahora como la histomorfología o el cariotipo, no proporcionaban información alguna relativa a la sensibilidad a los tratamientos. La tecnología de *microarrays* no sólo aporta información importante acerca de la biología y el comportamiento tu-

moral sino que además relacionan los genes que podrías ser dianas terapéuticas. Es el caso de ESR1 (codificado por ER) y ERBB2 (codificado por HER2), como genes importantes en la identificación de los subtipos tumorales y en la respuesta a los tratamientos (18). Así, el subtipo basal no expresa ER o HER2, pero sobreexpresa el receptor para el factor de crecimiento epidérmico (EGFR), que es diana de nuevas drogas desarrolladas para tratar el cáncer de colon, de cabeza y cuello o de pulmón (19,49). Algunos estudios han propuesto a las metaloproteinasas (50) o a NFκb (51) como dianas terapéuticas para determinados subtipos tumorales. Estos datos derivan de estudios con microchips en determinados modelos que permiten sacar este tipo de conclusiones. En algunos casos se ha sugerido que la resistencia al tamoxifeno pueda ser producto de la inducción del receptor de andrógenos (52), lo que implicaría resistencia a inhibidores de la aromataasa pero sensibilidad a terapias con anti-andrógenos en estas pacientes. Este tipo de hallazgos solo son posibles mediante el uso de metodologías de alto rendimiento como los *microarrays*, y pueden ser útiles para identificar nuevas dianas terapéuticas de drogas ya existentes (como los antiandrógenos) en grupos minoritarios de pacientes con esas características tumorales. En cualquier caso es un paso hacia la medicina de tratamiento personalizado, y como tal debe considerarse relevante. En este aspecto, los experimentos de validación son un paso imprescindible incluso cuando la diana terapéutica está definida. Los perfiles de expresión generan una importante cantidad de información que en ocasiones permite descubrir posibles nuevas vías de tratamiento contra el cáncer de mama.

TEST DIAGNÓSTICOS DE USO CLÍNICO

Existen algunos test diagnósticos de uso clínico comerciales. Es el caso de *Oncotype Dx*, que contiene 16 genes. Este test requiere el envío de las muestras embebidas en parafina a un laboratorio específico de EE. UU.; donde serán evaluadas y entrarán en un ensayo clínico prospectivo conocido como TAILORx (*Trial Assigning Individualized Options for Treatment*). Este estudio se inició en mayo de 2006 y ha reunido a más de 10.000 muestras de cáncer de mama de diverso tipo (ER-positivo y negativo, HER-2 negativo) de más de 900 lugares de EE. UU. El objetivo de este estudio es evaluar si esos 16 genes pueden ayudar a tomar decisiones relativas al tratamiento en pacientes que han recibido terapia hormonal adyuvante con o sin quimioterapia previa. MammaPrint® (53) es otro test diagnóstico compuesto por 70 genes en un pequeño microchip que además ha recibido la aprobación por parte de la FDA para su uso en clínica. En este caso la muestra de tumor debe enviarse a la compañía responsable del estudio situada en Holanda. Los 70 genes han sido validados para predecir recaídas tempranas así como para predecir el pronóstico de la enfermedad en pa-

cientes de tamaño tumoral menor de 2 cm e independientemente de su estatus de ER (54). Los resultados obtenidos muestran que este test tiene una capacidad pronóstica superior a *adjuvant online* y otros criterios para este tipo de pacientes (55). Existen otros proyectos de este tipo que intentan validar test predictivos a gran escala. Es el caso del MINDACT (*Microarray In Node-negative Disease may Avoid Chemotherapy Trial*) puesto en marcha en julio de 2006 y que trabaja con más de 6000 muestras de cáncer de mama nódulo linfático negativo. MINDACT tiene como objetivo saber si el 10-15% de tumores de mama diagnosticados en estadios tempranos pueden evitar la quimioterapia adyuvante utilizando solamente los datos generados por MammaPrint® y si además el test tiene también capacidad predictiva en cuanto a la respuesta a tratamiento (56). Los resultados derivados de todos estos proyectos son esperados por la comunidad científica con ansia, ya que pueden mejorar sustancialmente los tratamientos y pronósticos de las pacientes con cáncer de mama.

CONCLUSIONES

Existen un gran número de plataformas y métodos estadísticos para generar y analizar perfiles de expresión génica; y se espera que toda esta metodología se incremente en los años venideros. De momento, y como consenso en lo científico, se exige que los datos no procesados generados por los experimentos con *microarrays* sean de acceso público, lo cual facilita el manejo de los mismos por otros grupos de investigación y permite mayor facilidad para la validación externa de resultados. Respecto al cáncer de mama, se han publicado diversos perfiles de expresión relativos a la biología del tumor, el diagnóstico, el pronóstico y la respuesta al tratamiento. Algunos de estos resultados se han validado y están más cerca de tener algún tipo de aplicación clínica. En la actualidad hay puestos en marcha importantes estudios prospectivos que tienen como objetivo validar determinados perfiles de expresión influyentes en la práctica clínica. Los resultados de estos macroproyectos son esperados con ansia. Además, lo aprendido a lo largo de los años sobre el cáncer de mama tiene algunas aplicaciones sobre otros cánceres, lo que permite avanzar más rápido en otros frentes relativos a esta enfermedad.

BIBLIOGRAFÍA

- Hoheisel JD. Microarray technology: beyond transcript profiling and genotype analysis. *Nat Rev Genet* 2006; 7(3): 200-10.
- Schena M, Shalon D, Davis RW, Brown PO. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 1995; 270(5235): 467-70.
- Irizarry RA, Warren D, Spencer F, Kim IF, Biswal S, Frank BC, et al. Multiple-laboratory comparison of microarray platforms. *Nat Methods* 2005; 2(5): 345-50.
- Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci USA* 1998; 95(25): 14863-8.
- Quackenbush J. Microarray data normalization and transformation. *Nat Genet* 2002; 32 Suppl: 496-501.
- Yang YH, Dudoit S, Luu P, Lin DM, Peng V, Ngai J, Speed TP. Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Res* 2002; 30(4): e15.
- Hughes TR, Marton MJ, Jones AR, Roberts CJ, Stoughton R, Armour CD, et al. Functional discovery via a compendium of expression profiles. *Cell* 2000; 102(1): 109-26.
- Mahadevappa M, Warrington JA. A high-density probe array sample preparation method using 10- to 100-fold fewer cells. *Nat Biotechnol* 1999; 17(11): 1134-6.
- Khatiri P, Draghici S. Ontological analysis of gene expression data: current tools, limitations, and open problems. *Bioinformatics* 2005; 21(18): 3587-95.
- Slonim DK. From patterns to pathways: gene expression data analysis comes of age. *Nat Genet* 2002; 32 (suppl): 502-8.
- Ransohoff DF. Rules of evidence for cancer molecular-marker discovery and validation. *Nat Rev Cancer* 2004; 4(4): 309-14.
- Ihmels J, Friedlander G, Bergmann S, Sarig O, Ziv Y, Barkai N. Revealing modular organization in the yeast transcriptional network. *Nat Genet* 2002; 31(4): 370-7.
- McShane LM, Radmacher MD, Freidlin B, Yu R, Li MC, Simon R. Methods for assessing reproducibility of clustering patterns observed in analyses of microarray data. *Bioinformatics* 2002; 18(11): 1462-9.
- Churchill GA. Fundamentals of experimental design for cDNA microarrays. *Nat Genet* 2002; 32 (suppl): 490-5.
- Hosack DA, Dennis G Jr, Sherman BT, Lane HC, Lempicki RA. Identifying biological themes within lists of genes with EASE. *Genome Biol* 2003; 4(10): R70.
- Dennis G Jr, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, et al. DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol* 2003; 4(5): P3.
- Dupuy A, Simon RM. Critical review of published microarray studies for cancer outcome and guidelines on statistical analysis and reporting. *J Natl Cancer Inst* 2007; 99(2): 147-57.
- Perou CM, Sørlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, et al. Molecular portraits of human breast tumours. *Nature* 2000; 406(6797): 747-52.
- Sørlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A, et al. Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc Natl Acad Sci USA* 2003; 100(14): 8418-23.
- Calza S, Hall P, Auer G, Bjöhle J, Klaar S, Kronenwett U, et al. Intrinsic molecular signature of breast cancer in a population-based cohort of 412 patients. *Breast Cancer Res* 2006; 8(4): R34.
- Sotiriou C, Neo SY, McShane LM, Korn EL, Long PM, Jazaeri A, et al. Breast cancer classification and prognosis based on gene expression profiles from a population-based study. *Proc Natl Acad Sci U S A* 2003; 100(18): 10393-8.
- Yehiely F, Moyano JV, Evans JR, Nielsen TO, Cryns VL, et al. Deconstructing the molecular portrait of basal-like breast cancer. *Trends Mol Med* 2006; 12(11): 537-44.
- Chin K, DeVries S, Fridlyand J, Spellman PT, Roydasgupta R, Kuo WL, et al. Genomic and transcriptional aberrations linked to breast cancer pathophysiology. *Cancer Cell* 2006; 10(6): 529-41.
- Chang HY, Nuyten DS, Sneddon JB, Hastie T, Tibshirani R, Sørlie T, et al. Robustness, scalability, and integration of a wound-response gene expression signature in predicting breast cancer survival. *Proc Natl Acad Sci USA* 2005; 102(10): 3738-43.
- Clarke R, Liu MC, Bouker KB, Gu Z, Lee RY, Zhu Y, et al. Antiestrogen resistance in breast cancer and the role of estrogen receptor signaling. *Oncogene* 2003; 22(47): 7316-39.
- Berns EM, van Staveren IL, Verhoog L, van de Ouweland AM, Meijer-van Gelder M, Meijers-Heijboer H, et al. Molecular profiles of BRCA1-mutated and matched sporadic breast tumours: relation with clinico-pathological features. *Br J Cancer* 2001; 85(4): 538-45.
- Hedenfalk I, Ringner M, Ben-Dor A, Yakhini Z, Chen Y, Chebil G, et al. Molecular classification of familial non-BRCA1/BRCA2 breast cancer. *Proc Natl Acad Sci USA* 2003; 100(5): 2532-7.

28. Miller LD, Smeds J, George J, Vega VB, Vergara L, Ploner A, et al. An expression signature for p53 status in human breast cancer predicts mutation status, transcriptional effects, and patient survival. *Proc Natl Acad Sci USA* 2005; 102(38): 13550-5.
29. Cleton-Jansen AM. E-cadherin and loss of heterozygosity at chromosome 16 in breast carcinogenesis: different genetic pathways in ductal and lobular breast cancer? *Breast Cancer Res* 2002; 4(1): 5-8.
30. Coradini D, Pellizzaro C, Veneroni S, Ventura L, Daidone MG. Infiltrating ductal and lobular breast carcinomas are characterised by different interrelationships among markers related to angiogenesis and hormone dependence. *Br J Cancer* 2002; 87(10): 1105-11.
31. Korkola JE, DeVries S, Fridlyand J, Hwang ES, Estep AL, Chen YY, et al. Differentiation of lobular versus ductal breast carcinomas by expression microarray analysis. *Cancer Res* 2003; 63(21): 7167-75.
32. Hannemann J, Velds A, Halfwerk JB, Kreike B, Peterse JL, van de Vijver MJ. Classification of ductal carcinoma in situ by gene expression profiling. *Breast Cancer Res* 2006; 8(5): R61.
33. Kote-Jarai Z, Matthews L, Osorio A, Shanley S, Giddings I, Moreews F, et al. Accurate prediction of BRCA1 and BRCA2 heterozygous genotype using expression profiling after induced DNA damage. *Clin Cancer Res* 2006; 12(13): 3896-901.
34. Robbins P, Pinder S, de Klerk N, Dawkins H, Harvey J, Sterrett G, et al. Histological grading of breast carcinomas: a study of interobserver agreement. *Hum Pathol* 1995; 26(8): 873-9.
35. Dai H, van't Veer L, Lamb J, He YD, Mao M, Fine BM, et al. A cell proliferation signature is a marker of extremely poor outcome in a subpopulation of breast cancer patients. *Cancer Res* 2005; 65(10): 4059-66.
36. Van't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, et al. Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 2002; 415(6871): 530-6.
37. Van de Vijver MJ, He YD, van't Veer LJ, Dai H, Hart AA, Voskuil DW, et al. A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med* 2002; 347(25): 1999-2009.
38. Weigelt B, Glas AM, Wessels LF, Witteveen AT, Peterse JL, van't Veer LJ, et al. Gene expression profiles of primary breast tumors maintained in distant metastases. *Proc Natl Acad Sci USA* 2003; 100(26): 15901-5.
39. Smid M, Wang Y, Klijn JG, Sieuwerts AM, Zhang Y, Atkins D, et al. Genes associated with breast cancer metastatic to bone. *J Clin Oncol* 2006; 24(15): 2261-7.
40. Minn AJ, Gupta GP, Siegel PM, Bos PD, Shu W, Giri DD, et al. Genes that mediate breast cancer metastasis to lung. *Nature* 2005; 436(7050): 518-24.
41. Wang Y, Klijn JG, Zhang Y, Sieuwerts AM, Look MP, Yang F, et al. Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *Lancet* 2005; 365(9460): 671-9.
42. Foekens JA, Atkins D, Zhang Y, Sweep FC, Harbeck N, Paradiso A, et al. Multicenter validation of a gene expression-based prognostic signature in lymph node-negative primary breast cancer. *J Clin Oncol* 2006; 24(11): 1665-71.
43. Troester MA, Hoadley KA, Sørlie T, Herbert BS, Børresen-Dale AL, Lønning PE, et al. Cell-type-specific responses to chemotherapeutics in breast cancer. *Cancer Res* 2004; 64(12): 4218-26.
44. Györfy B, Serra V, Jürchott K, Abdul-Ghani R, Garber M, Stein U, et al. Prediction of doxorubicin sensitivity in breast tumors based on gene expression profiles of drug-resistant cell lines correlates with patient survival. *Oncogene* 2005; 24(51): 7542-51.
45. Oh DS, Troester MA, Usary J, Hu Z, He X, Fan C, et al. Estrogen-regulated genes predict survival in hormone receptor-positive breast cancers. *J Clin Oncol* 2006; 24(11): 1656-64.
46. Rouzier R, Perou CM, Symmans WF, Ibrahim N, Cristofanilli M, Anderson K, et al. Breast cancer molecular subtypes respond differently to preoperative chemotherapy. *Clin Cancer Res* 2005; 11(16): 5678-85.
47. Osborne CK. Tamoxifen in the treatment of breast cancer. *N Engl J Med* 1998; 339(22): 1609-18.
48. Fan C, Oh DS, Wessels L, Weigelt B, Nuyten DS, Nobel AB, et al. Concordance among gene-expression-based predictors for breast cancer. *N Engl J Med* 2006; 355(6): 560-9.
49. Nielsen TO, Hsu FD, Jensen K, Cheang M, Karaca G, Hu Z, et al. Immunohistochemical and clinical characterization of the basal-like subtype of invasive breast carcinoma. *Clin Cancer Res* 2004; 10(16): 5367-74.
50. Pusztai L, Sotiriou C, Buchholz TA, Meric F, Symmans WF, Esteva FJ, et al. Molecular profiles of invasive mucinous and ductal carcinomas of the breast: a molecular case study. *Cancer Genet Cytogenet* 2003; 141(2): 148-53.
51. Van Laere S, Van der Auwera I, Van den Eynden GG, Fox SB, Bianchi F, Harris AL, et al. Distinct molecular signature of inflammatory breast cancer by cDNA microarray analysis. *Breast Cancer Res Treat* 2005; 93(3): 237-46.
52. Doane AS, Danso M, Lal P, Donaton M, Zhang L, Hudis C, et al. An estrogen receptor-negative breast cancer subset characterized by a hormonally regulated transcriptional program and response to androgen. *Oncogene* 2006; 25(28): 3994-4008.
53. Glas AM, Floore A, Delahaye LJ, Witteveen AT, Pover RC, Bakx N, et al. Converting a breast cancer microarray signature into a high-throughput diagnostic test. *BMC Genomics* 2006; 7: 278.
54. Buyse M, Loi S, van't Veer L, Viale G, Delorenzi M, Glas AM, et al. Validation and clinical utility of a 70-gene prognostic signature for women with node-negative breast cancer. *J Natl Cancer Inst* 2006; 98(17): 1183-92.
55. Ravdin PM, Siminoff LA, Davis GJ, Mercer MB, Hewlett J, Gerson N, et al. Computer program to assist in making decisions about adjuvant therapy for women with early breast cancer. *J Clin Oncol* 2001; 19(4): 980-91.
56. Bogaerts J, Cardoso F, Buyse M, Braga S, Loi S, Harrison JA, et al. Gene signature evaluation as a prognostic tool: challenges in the design of the MINDACT trial. *Nat Clin Pract Oncol* 2006; 3(10): 540-51.