

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 384 107**

51 Int. Cl.:

**C12Q 1/68** (2006.01)

**G01N 33/574** (2006.01)

**C12N 15/10** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Número de solicitud europea: **05821698 .7**

96 Fecha de presentación: **04.11.2005**

97 Número de publicación de la solicitud: **1815014**

97 Fecha de publicación de la solicitud: **08.08.2007**

54 Título: **Indicadores moleculares de pronóstico de cáncer de mama y predicción de la respuesta al tratamiento**

30 Prioridad:  
**05.11.2004 US 625442 P**

45 Fecha de publicación de la mención BOPI:  
**29.06.2012**

45 Fecha de la publicación del folleto de la patente:  
**29.06.2012**

73 Titular/es:  
**GENOMIC HEALTH, INC.  
301 PENOBSCOT DRIVE  
REDWOOD CITY, CA 94063, US y  
NSABP FOUNDATION, INC.**

72 Inventor/es:  
**BAKER, Joffre B.;  
BRYANT, John L.;  
PAIK, Soonmyung y  
SHAK, Steven**

74 Agente/Representante:  
**Linage González, Rafael**

ES 2 384 107 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín europeo de patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre concesión de Patentes Europeas).

## DESCRIPCIÓN

Indicadores moleculares de pronóstico de cáncer de mama y predicción de la respuesta al tratamiento

5 **Campo de la invención**

La presente invención se refiere a indicadores moleculares cuantitativos que pueden guiar decisiones clínicas en el cáncer de mama, tal como cáncer de mama negativo para ganglios linfáticos, positivo para receptor de estrógenos (ESR1). En particular, la invención se refiere a determinados genes, cuya expresión variada indica la probabilidad de recurrencia del cáncer de mama resecado de forma quirúrgica en pacientes que no se tratan con un agente terapéutico de manera adyuvante. Además, la invención se refiere al uso de la medición cuantitativa de la expresión de determinados genes, incluyendo el gen ESR1, que se miden como una variable continua, para determinar (a) la probabilidad de una respuesta beneficiosa al agente terapéutico antiestrógeno, tal como tamoxifeno; y (b) la magnitud potencial de la respuesta beneficiosa a la quimioterapia.

15 **Descripción de la técnica relacionada**Estudios de expresión génica

20 Los oncólogos disponen de varias opciones de tratamiento, incluyendo diferentes combinaciones de fármacos quimioterápicos que están caracterizados como “de referencia”, y varios fármacos que no tienen especificaciones declaradas para el tratamiento de un cáncer particular, pero para los que existe evidencia de eficacia en este cáncer. La mayor probabilidad de buen desenlace del tratamiento requiere que se identifiquen los pacientes con mayor riesgo de enfermedad metastásica y se asignen al tratamiento contra el cáncer óptimo disponible. En particular, es importante determinar la probabilidad de la respuesta del paciente a los fármacos terapéuticos “de referencia”, tales como ciclofosfamida, metotrexato, 5-fluorouracilo, antraciclinas, taxanos y fármacos antiestrógenos, tal como tamoxifeno, porque tienen eficacia limitada y un espectro de efectos secundarios a menudo graves. La identificación de pacientes que tienen la mayor o menor probabilidad de necesitar y responder a los fármacos disponibles podría, por tanto, aumentar el beneficio neto que estos fármacos ofrecen, y disminuir la morbilidad y toxicidad netas, mediante una selección de pacientes más inteligente.

Actualmente, las pruebas de diagnóstico usadas en la práctica clínica son analitos individuales, y por tanto no capturan el valor potencial de las relaciones conocidas entre docenas de diferentes marcadores. Además, las pruebas de diagnóstico a menudo se basan en la inmunohistoquímica, que no es cuantitativa. La inmunohistoquímica a menudo produce diferentes resultados en diferentes laboratorios, en parte debido a que no están normalizados los reactivos, y en parte debido a que las interpretaciones son subjetivas. Las pruebas basadas en ARN, aunque cuantitativas de manera potencialmente alta, no se han usado debido a la percepción de que el ARN se destruye en muestras tumorales tal como se preparan de manera rutinaria, concretamente fijadas en formalina y embebidas en parafina (FPE), y debido a que es inconveniente obtener y almacenar muestras de tejido nuevo de pacientes para su análisis.

45 Durante las últimas dos décadas, la biología molecular y la bioquímica han revelado centenas de genes cuyas actividades influyen en el comportamiento de las células tumorales, su estado de diferenciación, y su sensibilidad o resistencia a determinados fármacos terapéuticos. Sin embargo, con pocas excepciones, no se ha aprovechado el estado de estos genes con el fin de tomar de manera rutinaria decisiones clínicas sobre los tratamientos con fármacos. En los últimos años, varios grupos han publicado estudios sobre la clasificación de diversos tipos de cánceres mediante análisis de la expresión génica en micromatrices de miles de genes (véase, por ejemplo Golub *et al.*, *Science* 286:531-537 (1999); Bhattacharjæ *et al.*, *Proc. Natl. Acad. Sci. USA* 98:13790-13795 (2001); Chen-Hsiang *et al.*, *Bioinformatics* 17 (Supl. 1):S316-S322 (2001); Ramaswamy *et al.*, *Proc. Natl. Acad. Sci. USA* 98:15149-15154 (2001); Martin *et al.*, *Cancer Res.* 60:2232-2238 (2000); West *et al.*, *Proc. Natl. Acad. Sci. USA* 98:11462-114 (2001); Sorlie *et al.*, *Proc. Natl. Acad. Sci. USA* 98:10869-10874 (2001); Yan *et al.*, *Cancer Res.* 61:8375-8380 (2001)). Sin embargo, estos estudios aún no han producido pruebas usadas de manera rutinaria en la práctica clínica, en gran parte debido a que las micromatrices requieren ARN de tejido nuevo o congelado y tales muestras no se presentan en cantidad suficiente para permitir la validación clínica de las firmas moleculares identificadas.

60 En los últimos tres años, se ha hecho posible obtener un perfil de la expresión génica de cientos de genes en tejido fijado en formalina, embebido en parafina (FPE) usando la tecnología de RT-PCR. Se han descrito métodos que son altamente sensibles, precisos y reproducibles (Cronin *et al.*, *Am. J. Pathol.* 164:35-42 (2004)). Debido a que miles de muestras de tejido clínico FPE archivados existen con registros clínicos asociados, tales como supervivencia, historia de tratamiento con fármacos, etc., la capacidad para someter a ensayo ahora de forma cuantitativa la expresión génica en este tipo de tejido permite estudios clínicos rápidos que relacionan la expresión de determinados genes con el pronóstico del paciente y la probabilidad de respuesta a los tratamientos. El uso de datos generados mediante estudios clínicos pasados permite resultados rápidos debido a que los acontecimientos clínicos son históricos. Por el contrario, por ejemplo, si se desea llevar a cabo un estudio de supervivencia en pacientes con cáncer recién incluidos, generalmente sería necesario esperar muchos años para determinar que se han producido

un número de muertes estadísticamente suficiente.

#### Pronóstico y predicción de cáncer de mama

- 5 El cáncer de mama es el tipo más común de cáncer entre mujeres en los Estados Unidos, y es la causa principal de muertes por cáncer entre mujeres de 40 – 59 años de edad.

En la actualidad sólo se usan clínicamente de manera rutinaria pocas pruebas moleculares en el cáncer de mama. Los ensayos inmunohistoquímicos para determinar proteínas de receptor de estrógenos (ESR1) y receptor de progesterona (PGR) se usan como base para la selección de pacientes para el tratamiento con fármacos antiestrógenos, tales como tamoxifeno (TAM). Además, se usan inmunohistoquímica o hibridación *in situ* fluorescente de ErbB2 (Her2) (que miden la proteína y el ADN, respectivamente) para seleccionar pacientes con los fármacos antagonistas de Her2, tales como trastuzumab (Herceptin®; Genentech, Lie, South San Francisco, CA).

- 15 Debido a que las pruebas actuales para el pronóstico y para la predicción de la respuesta a la quimioterapia son inadecuadas, las estrategias de tratamiento contra el cáncer de mama varían entre los oncólogos (Schott y Hayes, J. Clin. Oncol. PMID 15505274 (2004); Hayes, Breast 12:543-9 (2003)). Generalmente, los pacientes negativos para ganglios linfáticos cuyos tumores se halla que son positivos para ESR1 se tratan con un fármaco antiestrógeno, tal como TAM, y los pacientes cuyos tumores se halla que son negativos para ESR1 se tratan con quimioterapia. A menudo, a los pacientes positivos para ESR1 también se les prescribe quimioterapia además de terapia con antiestrógeno, aceptando los efectos secundarios tóxicos de la quimioterapia con el fin de disminuir modestamente el riesgo de recurrencia del cáncer. Las toxicidades incluyen, neuropatía, náuseas y otros síntomas gastrointestinales, pérdida del cabello y deterioro cognitivo. Se ha de temer la recurrencia debido a que el cáncer de mama recurrente es habitualmente metastásico y poco sensible al tratamiento. Claramente existe una necesidad de identificar a aquellos pacientes que corren un riesgo sustancial de recurrencia (es decir, proporcionar información de pronóstico) y es probable que respondan a la quimioterapia (es decir, proporcionar información predictiva). Igualmente, existe una necesidad de identificar a aquellos pacientes que no corren un riesgo significativo de recurrencia o que no es probable que respondan a la quimioterapia, puesto que a estos pacientes se les debe evitar la exposición innecesaria a estos fármacos tóxicos.

- 30 Los factores pronóstico difieren de los factores predictivos de tratamiento en el cáncer de mama. Los factores pronóstico son aquellas variables relacionadas con la historia natural del cáncer de mama, que influyen en las tasas de recurrencia y el desenlace de los pacientes una vez que han desarrollado cáncer de mama. Los parámetros clínicos que se han asociado con un peor pronóstico incluyen, por ejemplo, afectación de los ganglios linfáticos, aumento del tamaño del tumor y tumores de grado alto. Los factores pronóstico se usan frecuentemente para clasificar a pacientes en subgrupos con diferentes riesgos de recidiva iniciales. Por el contrario, los factores predictivos de tratamiento son variables relacionadas con la probabilidad de una respuesta beneficiosa en un paciente individual a un tratamiento, tal como terapia con antiestrógeno o quimioterapia, independiente del pronóstico.

- 40 Existe una gran necesidad de factores pronóstico y predictivos cuantitativos precisos que puedan ayudar al médico asistencial a realizar elecciones de tratamiento inteligentes, adaptadas a las necesidades de un paciente particular, basándose en un análisis de riesgo-beneficio bien fundado.

- 45 El documento WO 2004/065583 da a conocer conjuntos de genes, cuya expresión es importante en el diagnóstico y/o el pronóstico del cáncer de mama.

- 50 El documento WO 03/078662 da a conocer métodos para medir niveles de ARNm en tejidos tumorales sometidos a biopsia, incluyendo material de biopsia embebido en parafina archivado. También se dan a conocer conjuntos de genes de cáncer de mama en el diagnóstico y el tratamiento del cáncer de mama, y métodos para asignar las opciones de tratamiento más óptimas a un paciente con cáncer de mama basándose en el conocimiento derivado de estudios de expresión génica.

#### **Sumario de la invención**

- 55 Según un aspecto de la presente invención, se proporciona un método para la determinación cuantitativa de la probabilidad de una respuesta beneficiosa en un paciente con cáncer de mama positivo para ESR1 al tratamiento con un fármaco antiestrógeno tal como se especifica en la reivindicación 1.

- 60 En una realización, el fármaco antiestrógeno es TAM.

El método predictivo de tratamiento de la presente invención puede comprender la etapa de preparar un informe para el paciente, incluyendo una recomendación de tratamiento.

- 65 En una realización adicional, se recomienda el tratamiento con fármaco antiestrógeno sin quimioterapia cuando la expresión de ESR1 para el paciente es superior al valor de expresión medido en la misma prueba para el porcentaje

de pacientes con cáncer de mama negativo para ganglios y positivo para ER, y se sabe que el paciente está por lo demás en un grupo de riesgo bajo.

En otra realización, la variable es el nivel de expresión del transcrito de ARNm de ESR1, o su producto de expresión.

En una realización adicional, el método incluye la etapa de determinar la puntuación de recurrencia para el paciente.

En una realización particular, tras determinar la puntuación de recurrencia, se recomienda un tratamiento que comprende quimioterapia cuando la expresión de ESR1 para el paciente es diferente de cero pero es inferior al valor de expresión medido en la misma prueba para un porcentaje particular de pacientes con cáncer de mama negativo para ganglios y positivo para ER, y se sabe que el paciente está por lo demás en un grupo que tiene un riesgo alto de recurrencia.

### Breve descripción de las figuras

La figura 1 muestra el aumento absoluto en proporción de pacientes tratados con TAM que están libres de enfermedad distante a los 10 años como una función de la medición cuantitativa de ESR1.

La figura 2 muestra el beneficio absoluto del tratamiento con tamoxifeno (TAM) como una función de la expresión de ESR1. Las líneas horizontales a través de las barras representan límites de confianza del 95%.

La figura 3 muestra el beneficio absoluto del tratamiento con TAM como una función del cuartil en el que la expresión de ESR1 disminuye. Las líneas horizontales a través de las barras representan límites de confianza del 95%. Los puntos de corte que definen los cuartiles de la expresión de ESR1, se basan en valores  $C_T$  normalizados con respecto a la referencia derivados del análisis de la población de pacientes B-14 del NSABP.

### Descripción detallada de la realización preferida

#### A. Definiciones

A menos que se defina lo contrario, los términos técnicos y científicos usados en el presente documento tienen el mismo significado que el que entendería comúnmente un experto habitual en la técnica al que pertenece esta invención. Singleton *et al.*, Dictionary of Microbiology and Molecular Biology 2<sup>a</sup> ed., J. Wiley & Sons (Nueva York, NY 1994); y Webster's New World™ Medical Dictionary, 2<sup>a</sup> edición, Wiley Publishing Inc., 2003, proporcionan al experto en la técnica una guía general para muchos de los términos usados en la presente solicitud. Para los fines de la presente invención, se definen los siguientes términos a continuación.

La expresión "respuesta beneficiosa" significa una mejora en cualquier medida del estado del paciente, incluyendo aquellas medidas usadas habitualmente en la técnica, tal como supervivencia global, supervivencia sin progresión, supervivencia sin recurrencia, y supervivencia sin recurrencia distante. La supervivencia sin recurrencia (RFS, *recurrence-free survival*) se refiere al tiempo (en años) desde la cirugía hasta la primera recurrencia local, regional o distante. La supervivencia sin recurrencia distante (DFRS, *distant recurrence-free survival*) se refiere al tiempo (en años) desde la cirugía hasta la primera recurrencia anatómicamente distante. El cálculo de estas medidas en la práctica puede variar de estudio a estudio dependiendo de la definición de los acontecimientos que van a ser o bien excluidos o bien no considerados.

El término "micromatriz" se refiere a una disposición ordenada de elementos de matriz que pueden hibridarse, preferiblemente sondas de polinucleótidos, en un sustrato.

El término "polinucleótido", cuando se usa en singular o plural, generalmente se refiere a cualquier polirribonucleótido o polidesoxirribonucleótido, que puede ser ARN o ADN no modificado o ARN o ADN modificado. Por tanto, por ejemplo, los polinucleótidos tal como se definen en el presente documento incluyen, sin limitación, ADN mono y bicatenario, ADN que incluye regiones mono y bicatenarias, ARN mono y bicatenario, y ARN que incluye regiones mono y bicatenarias, moléculas híbridas que comprenden ADN y ARN que pueden ser monocatenarios o, más normalmente, bicatenarios o incluyen regiones mono y bicatenarias. Además, el término "polinucleótido" tal como se usa en el presente documento se refiere a regiones tricatenarias que comprenden ARN o ADN o tanto ARN como ADN. Las hebras en tales regiones pueden proceder de la misma molécula o de moléculas diferentes. Las regiones pueden incluir todas de una o más de las moléculas, pero más normalmente implican sólo una región de algunas de las moléculas. Una de las moléculas de la región de triple hélice a menudo es un oligonucleótido. El término "polinucleótido" incluye específicamente ADNc. El término incluye ADN (incluyendo ADNc) y ARN que contienen una o más bases modificadas. Por tanto, los ADN o ARN con estructuras principales modificadas para su estabilidad o por otros motivos son "polinucleótidos" tal como se concibe este término en el presente documento. Además, los ADN o ARN que comprenden bases inusuales, tal como inosina, o bases modificadas, tales como bases tritadas, se incluyen dentro del término "polinucleótidos" tal como se define en el presente documento. En general, el término "polinucleótido" abarca todas las formas modificadas química, enzimática y/o metabólicamente de polinucleótidos no modificados, así como todas las formas químicas de ADN y ARN

características de virus y células, incluyendo células simples y complejas.

El término “oligonucleótido” se refiere a un polinucleótido relativamente corto, que incluye, sin limitación, desoxirribonucleótidos monocatenarios, ribonucleótidos mono o bicatenarios, híbridos de ARN:ADN y ADN bicatenarios. Los oligonucleótidos, tales como oligonucleótidos de sonda de ADN monocatenario, a menudo se sintetizan mediante métodos químicos, por ejemplo usando sintetizadores de oligonucleótidos automatizados que están disponibles comercialmente. Sin embargo, los oligonucleótidos pueden prepararse mediante una variedad de otros métodos, que incluyen técnicas mediadas por ADN recombinante *in vitro* y mediante la expresión de ADN en células y organismos.

La expresión “expresión génica” describe la conversión de la información de la secuencia génica de ADN en ARN transcrito (el transcrito de ARN no sometido a corte y empalme inicial o el ARNm maduro) o el producto proteico codificado. La expresión génica puede monitorizarse midiendo los niveles de o bien el ARN entero o bien los productos proteicos del gen o las subsecuencias.

La expresión “amplificación génica” se refiere a un proceso mediante el cual se forman múltiples copias de un gen o fragmento génico en una línea celular o célula particular. La región duplicada (un tramo de ADN amplificado) a menudo se denomina “amplicón”. A menudo, la cantidad de ARN mensajero (ARNm) producido, es decir, el nivel de expresión génica, también aumenta en proporción al número de copias hechas del gen particular expresado.

Los factores pronóstico son aquellas variables relacionadas con la historia natural de cáncer de mama, que influyen en las tasas de recurrencia y el desenlace de los pacientes una vez que han desarrollado cáncer de mama. Los parámetros clínicos que se han asociado con un peor pronóstico incluyen, por ejemplo, afectación de los ganglios linfáticos, aumento de tamaño del tumor y tumores de grado alto. Los factores pronóstico se usan frecuentemente para clasificar a los pacientes en subgrupos con diferentes riesgos de recidiva iniciales. Por el contrario, los factores predictivos de tratamiento son variables relacionadas con la probabilidad de una respuesta beneficiosa en un paciente individual a un tratamiento, tal como terapia con antiestrógeno o quimioterapia, independiente del pronóstico.

El término “pronóstico” se usa en el presente documento para referirse a la probabilidad de muerte que puede atribuirse a cáncer o progresión del cáncer, incluyendo recurrencia y propagación metastásica de una enfermedad neoplásica, tal como cáncer de mama, durante la historia natural de la enfermedad. Los factores pronóstico son aquellas variables relacionadas con la historia natural de enfermedades neoplásicas, tales como cáncer de mama, que influyen en las tasas de recurrencia y el desenlace de la enfermedad una vez que el paciente desarrolló la enfermedad neoplásica, tal como cáncer de mama. En este contexto, “desenlace natural” significa desenlace en ausencia de tratamiento adicional. Por ejemplo, en el caso de cáncer de mama, “desenlace natural” significa desenlace tras la resección quirúrgica del tumor, en ausencia de tratamiento adicional (tal como, quimioterapia o tratamiento por radiación). Frecuentemente se usan los factores pronóstico para clasificar a los pacientes en subgrupos con diferentes riesgos iniciales, tales como riesgos de recidiva iniciales.

El término “predicción” se usa en el presente documento para referirse a la probabilidad de que un paciente responda o bien favorablemente o bien desfavorablemente a un fármaco o un conjunto de fármacos, y también el grado de esas respuestas. Por tanto, los factores predictivos de tratamiento son aquellas variables relacionadas con la respuesta en un paciente individual con respecto a un tratamiento específico, independiente del pronóstico. Los métodos predictivos de la presente invención pueden usarse clínicamente para tomar decisiones de tratamiento eligiendo las modalidades de tratamiento más apropiadas para cualquier paciente particular. Los métodos predictivos de la presente invención son herramientas valiosas en la predicción de si es probable que un paciente responda favorablemente a un régimen de tratamiento, tal como terapia con antiestrógeno, tal como tratamiento con TAM solo o en combinación con quimioterapia y/o radioterapia.

La expresión supervivencia “a largo plazo” se usa en el presente documento para referirse a la supervivencia durante al menos 3 años, más preferiblemente durante al menos 8 años, lo más preferiblemente durante al menos 10 años tras cirugía u otro tratamiento.

El término “tumor”, tal como se usa en el presente documento, se refiere a todo crecimiento y proliferación de células neoplásicas, ya sean malignas o benignas, y todas las células y tejidos precancerosos y cancerosos.

Los términos “cáncer” y “canceroso” se refieren a o describen el estado fisiológico en mamíferos que se caracteriza normalmente por crecimiento celular desregulado. Los ejemplos de cáncer incluyen, pero no se limitan a, cáncer de mama, cáncer de ovario, cáncer de colon, cáncer de pulmón, cáncer de próstata, cáncer hepatocelular, cáncer gástrico, cáncer de páncreas, cáncer de cuello uterino, cáncer de hígado, cáncer de vejiga, cáncer de las vías urinarias, cáncer de tiroides, cáncer renal, carcinoma, melanoma y cáncer de cerebro.

La “patología” del cáncer incluye todos los fenómenos que comprometen el bienestar del paciente. Esto incluye, sin limitación, crecimiento celular anómalo o incontrolable, metástasis, interferencia con el funcionamiento normal de células vecinas, liberación de citocinas u otros productos de secreción a niveles anómalos, supresión o

agravamiento de la respuesta inflamatoria o inmunológica, neoplasia, tumor premaligno, tumor maligno, invasión de tejidos u órganos circundantes o distantes, tales como ganglios linfáticos, etc.

5 En el contexto de la presente invención, la referencia a “al menos uno”, “al menos dos”, “al menos tres”, “al menos cuatro”, “al menos cinco”, etc. de los genes enumerados en cualquier conjunto de genes particular significa uno cualquiera y cualquiera de las combinaciones de los genes enumerados.

10 La expresión cáncer “negativo para ganglios”, tal como cáncer de mama “negativo para ganglios”, se usa en el presente documento para hacer referencia a cáncer que no se ha propagado a los ganglios linfáticos.

Las expresiones “corte y empalme” y “corte y empalme de ARN” se usan de manera intercambiable y se refieren al procesamiento de ARN que elimina intrones y une exones para producir ARNm maduro con la secuencia codificante continua que se desplaza al interior del citoplasma de una célula eucariota.

15 En teoría, el término “exón” se refiere a cualquier segmento de un gen interrumpido que está representado en el producto de ARN maduro (B. Lewin. Genes IV Cell Press, Cambridge Mass. 1990). En la teoría, el término “intrón” se refiere a cualquier segmento de ADN que se transcribe pero se elimina del transcrito mediante corte y empalme junto con los exones a cada lado de éste. De forma operativa, las secuencias de exones se producen en la secuencia de ARNm de un gen tal como se define mediante los números Ref. SEQ. De forma operativa, las secuencias de intrones son secuencias intercaladas dentro del ADN genómico de un gen, separadas por secuencias de exones y que tienen secuencias consenso de corte y empalme GT y AG en sus límites 5' y 3'.

## B. Descripción detallada

25 La práctica de la presente invención empleará, a menos que se indique lo contrario, técnicas convencionales de biología molecular (incluyendo técnicas recombinantes), microbiología, biología celular y bioquímica, que están dentro de la habilidad de la técnica. Tales técnicas se explican completamente en la bibliografía, tal como, “Molecular Cloning: A Laboratory Manual”, 2ª edición (Sambrook *et al.*, 1989); “Oligonucleotide Synthesis” (MJ. Gait, ed., 1984); “Animal Cell Culture” (R.I. Freshney, ed., 1987); “Methods in Enzymology” (Academic Press, Inc.); 30 “Handbook of Experimental Immunology”, 4ª edición (D.M. Weir & CC. Blackwell, eds., Blackwell Science Inc., 1987); “Gene Transfer Vectors for Mammalian Cells” (J.M. Miller & M.P. Calos, eds., 1987); “Current Protocols in Molecular Biology” (F.M. Ausubel *et al.*, eds., 1987); y “PCR: The Polymerase Chain Reaction”, (Mullis *et al.*, eds., 1994).

### B.1. Descripción general de la invención

35 Durante los últimos dos años Genomic Health, Inc y colaboradores (Esteban *et al.*, Proc Am Soc Clin Oncol 22: página 850, 2003 (abstract 3416); Cobleigh *et al.* Soc Clin Oncol 22: página 850, 2003 (abstract 3415); Soule *et al.*, Proc Am Soc Clin Oncol 22: página 862, 2003 (abstract 3466)) notificaron varios estudios clínicos exploratorios de la expresión génica en cáncer de mama temprano, que tenían como objetivo hallar una firma molecular para el riesgo de recurrencia. Estos estudios usaron RT-PCR cuantitativa para someter a prueba 250 marcadores génicos 40 candidatos en muestras de tejido embebido en parafina, congelado (FPE, *frozen, paraffin-embedded*) que tenían registros clínicos asociados. Se realizó un análisis en los tres estudios con el fin de examinar si podían identificarse genes que se relacionaban de manera constante con el riesgo de recurrencia en un grupo diverso de pacientes. Basándose en estos resultados univariantes, se diseñaron modelos de múltiples genes y se analizaron en los tres 45 estudios. Se desarrolló un único ensayo de múltiples genes, que consistía en 16 genes relacionados con cáncer y 5 genes de referencia, para someterse a prueba de modo prospectivo en estudios de validación clínica. Se generó un algoritmo denominado puntuación de recurrencia (RS, *Recurrence Score*), que utiliza las mediciones de estas 21 especies de ARNm y realiza una lectura del riesgo de recurrencia en una escala de 100 puntos.

50 Para someter a prueba la validez clínica de esta prueba y algoritmo de puntuación de recurrencia, se llevó a cabo un ensayo clínico ciego con criterios de valoración identificados de manera prospectiva. Este ensayo de validación se centró en pacientes tratados con TAM solo en las ramas de inscripción y aleatorizada de la población de ensayo clínico del estudio B-14 del proyecto nacional de cirugía adyuvante de mama e intestinos (NSABP, *National Surgical Adjuvant Breast and Bowel Project*) (Fisher B, Costantino JP, Redmond CK, *et al.*: Endometrial cancer in TAM-treated breast cancer patients: Findings from the National Surgical Adjuvant Breast y Bowel Project (NSABP) B-14. J Natl Cancer Inst 86:527-537 (1994)). Genomic Health, Inc. y el NSABP llevaron a cabo el ensayo de RT-PCR de 21 genes en 668 muestras de tejido de cáncer de mama derivadas de estos pacientes y se calculó una puntuación 55 recurrencia para cada paciente.

60 Los puntos de corte especificados previamente de la puntuación de recurrencia clasificaron a los pacientes en una de tres categorías: riesgo bajo, riesgo intermedio y riesgo alto de recurrencia de enfermedad distante. La proporción de los 668 pacientes clasificados como de riesgo bajo, intermedio y alto mediante el ensayo de RT-PCR fueron del 51%, del 23% y del 27%, respectivamente. Las estimaciones de Kaplan-Meier y los intervalos de confianza del 95% para las tasas de recurrencia distante a los 10 años fueron del 6,8% (4,0%, 9,6%), del 14,3% (8,3%, 20,3%), del 65 30,5% (23,6%, 37,4%), respectivamente, para los grupos de riesgo bajo, intermedio y alto; la tasa para el grupo de riesgo bajo fue significativamente menor que la tasa para el grupo de riesgo alto ( $p < 0,001$ ). En un modelo de Cox

5 multivariante que relaciona la recurrencia distante con respecto a la puntuación de recurrencia, la edad y el tamaño del tumor, la puntuación de recurrencia proporciona un poder predictivo significativo ( $p < 0,001$ ) que va más allá de la edad y el tamaño del tumor. Este estudio validó la puntuación de recurrencia como un poderoso factor pronóstico de recurrencia distante en pacientes sin ganglios afectados que tienen tumores que son positivos para ESR1 y tratadas con TAM (Paik *et al.* Breast Cancer Research and Treatment 82, Suplemento 1: página S10, 2003 (abstract 16)).

10 La invención dada a conocer en el presente documento se deriva, en parte, del estudio de pacientes en la rama de placebo del estudio clínico B-14 del NSABP (B-14) y, en parte, de la comparación de pacientes en la rama de placebo de B-14 con respecto a pacientes tratados con TAM con pacientes en las ramas de inscripción y aleatorizada del estudio B-14 del NSABP. Se analizó de manera cuantitativa el tejido de cáncer de mama derivado de los pacientes tratados con placebo, usando un ensayo de RT-PCR para cuantificar la expresión de dieciséis genes relacionados con cáncer y cinco genes de referencia.

15 El análisis de expresión génica cuantitativo dio como resultado la identificación de indicadores moleculares de pronóstico. Basándose en el análisis de la relación entre la expresión génica y la supervivencia sin recurrencia distante en la rama de placebo del ensayo B-14 del NSABP B-14, se ha identificado un conjunto de genes, cuyos niveles de expresión son indicativos del desenlace si no se proporciona tratamiento adicional al paciente. El desenlace puede manifestarse en diversas mediciones que incluyen supervivencia, supervivencia sin recurrencia y 20 supervivencia sin recurrencia distante, todas las cuales están dentro del alcance de la invención.

25 Los genes y grupos de genes de pronóstico identificados pueden usarse individualmente o en combinaciones particulares para predecir la probabilidad de desenlace para pacientes particulares. Los indicadores de pronóstico incluyen, específicamente, el grupo de proliferación (BIRC5 + MK1 + MYBL2 + CCNB1 + STK6), el grupo de invasión (CTSL2 + MMP11), y uno o más de los genes individuales: CCNB1, BIRC5, MYBL2, PGR, STK6, MK1, GSTM1, GAPD, RPLP0 y MMP11.

30 En otro aspecto, el análisis de la expresión génica dado a conocer en el presente documento dio como resultado la identificación de indicadores moleculares de respuesta beneficiosa a fármacos antiestrógenos, tales como TAM, basándose en el análisis de la relación entre la expresión génica y la supervivencia sin recurrencia distante en pacientes no tratados de la rama de placebo de B-14 así como pacientes tratados con TAM de las ramas tanto de inscripción como de aleatorización de B-14.

35 Basándose en el análisis de interacción de la relación entre la expresión génica y la supervivencia sin recurrencia distante en las cohortes de placebo y tratamiento combinadas, se han identificado un conjunto de genes o grupos de genes, cuyos niveles de expresión son indicativos de respuesta beneficiosa al tratamiento con TAM. Estos genes y grupos de genes se usan en combinaciones particulares para predecir la probabilidad de respuesta beneficiosa al tratamiento con TAM, u otro fármaco antiestrógeno, para pacientes individuales. Específicamente, estos genes/grupos de genes son: el grupo de ESR1 (ESR1 + PGR + BCL2 + SCUBE2).

40 Un hallazgo significativo de la invención es que los niveles cuantitativos de ESR1 se relacionan con la probabilidad de beneficio con TAM como una variable continua a lo largo de una escala de expresión de 14 puntos. Por tanto, para un paciente individual es posible proporcionar una estimación cuantitativa de la probabilidad de beneficio para este agente terapéutico, con niveles de expresión génica de ESR1 superiores que se correlacionan con una mayor 45 posibilidad de respuesta. Esta información puede utilizarse de varias maneras. Proporciona una evaluación más refinada de la probabilidad de una respuesta beneficiosa al tratamiento con TAM y otras terapias con antiestrógeno, que las que han estado disponibles previamente. El TAM tiene efectos secundarios significativos, que incluyen el desarrollo de cáncer uterino, trombosis de venas profundas, embolia pulmonar y cataratas (Physicians Desk Reference 2002). De manera similar, otros fármacos antiestrógenos, tales como, toremifeno (Fareston®, Orion, Corp.), anastrozol (Arimidex®, AstraZeneca Pharmaceuticals) y acetato de megasterol, tienen efectos secundarios graves. Como resultado de la presente invención, los pacientes y sus oncólogos pueden usar ahora la puntuación de ESR1 para evaluar el riesgo frente al beneficio cuando decidan si es apropiado el tratamiento con TAM u otra terapia con antiestrógeno.

55 Aunque la expresión de ESR1 (principalmente tal como se determina mediante medición inmunohistoquímica a nivel proteico) se usa de manera rutinaria en la práctica clínica para determinar si un paciente debe tratarse con TAM, basándose en el estado "positivo para ESR1" o "negativo para ESR1", los hallazgos que subyacen a la presente invención se relacionan con pacientes que ya están definidos como "positivos para ESR1" mediante los criterios convencionales. Según la presente invención, es posible determinar la probabilidad de una respuesta beneficiosa al 60 tratamiento con TAM, o al tratamiento con otros fármacos antiestrógenos, entre este grupo de pacientes.

Los niveles de ESR1 pueden usarse junto con la puntuación de recurrencia (tratada a continuación) para determinar si a pacientes individuales se les debe prescribir TAM (u otro fármaco antiestrógeno) solo, o TAM (u otro fármaco antiestrógeno) más quimioterapia.

65 La invención permite adicionalmente el diseño de una prueba particular, un ejemplo de la cual se facilita en el

ejemplo a continuación, con puntos de corte de expresión de ESR1 precisos que predicen un nivel alto, intermedio o bajo del beneficio del tratamiento con TAM o el tratamiento con otros fármacos antiestrógenos.

En diversas realizaciones de la invención, están disponibles diversos enfoques tecnológicos para la determinación de los niveles de expresión de los genes dados a conocer incluyendo, sin limitación, RT-PCR, micromatrices, análisis en serie de la expresión génica (SAGE) y análisis de la expresión génica mediante secuenciación masiva en paralelo de firmas (MPSS), que se tratarán en detalle a continuación. En realizaciones particulares, el nivel de expresión de cada gen puede determinarse con respecto a diversas características de los productos de expresión del gen que incluyen exones, intrones, epítomos de proteína y actividad de proteína.

## B.2 Obtención del perfil de expresión génica

En general, los métodos de obtención del perfil de expresión génica pueden dividirse en dos grandes grupos: métodos basados en el análisis de hibridación de polinucleótidos, y métodos basados en la secuenciación de polinucleótidos. Los métodos usados más comúnmente conocidos en la técnica para la cuantificación de la expresión de ARNm en una muestra incluyen transferencia de tipo Northern e hibridación *in situ* (Parker y Barnes, *Methods in Molecular Biology* 106:247-283 (1999)); ensayos de protección de ARNasa (Hod, *Biotechniques* 13:852-854 (1992)); y reacción en cadena de la polimerasa con transcripción inversa (RT-PCR) (Weis *et al.*, *Trends in Genetics* 8:263-264 (1992)). Alternativamente, pueden emplearse anticuerpos que pueden reconocer dúplex específicos, incluyendo dúplex de ADN, dúplex de ARN y dúplex híbridos de ADN-ARN o dúplex de ADN-proteína. Los métodos representativos para el análisis de la expresión génica basado en secuenciación incluyen análisis en serie de la expresión génica (SAGE), y análisis de la expresión génica mediante secuenciación masiva en paralelo de firmas (MPSS).

Dos procesos biológicos implicados comúnmente en la tumorigénesis incluyen amplificación génica y metilación del ADN. Ambos procesos dan como resultado la expresión anómala de genes importantes en la formación o el progreso de tumores. Los métodos que monitorizan la amplificación génica y la metilación de ADN pueden, por tanto, considerarse métodos sustitutos para obtener del perfil de expresión génica.

La amplificación génica es una alteración común en muchos cánceres que puede conducir a la expresión elevada de oncogenes celulares (Meltzer, P. *et al.*, *Cancer Genet Cytogenet.* 19:93 (1986). En el cáncer de mama, existe una buena correlación entre la amplificación génica de ERBB2 y la sobreexpresión de ERBB2 (Nagai, M. A. *et al.*, *Cancer Biother* 8:29 (1993), Savinainen, K.J. *et al.*, *Am. J. Pathol.* 160:339 (2002)). La amplificación del gen ERBB2 que conduce a su sobreexpresión, se correlaciona con un pronóstico pobre (Press, M.F. *et al.*, *J. Clin. Oncol.* 15:2894 (1997), Slamon, D.J. *et al.*, *Science* 244:707 (1989)) y es predictivo de la respuesta a la terapia anti-HER2 en combinación con quimioterapia convencional (Seidman, A.D. *et al.*, *J. Clin. Oncol.* 19:1866 (2001)).

También se ha demostrado que la metilación del ADN es una alteración común en el cáncer que conduce a la expresión elevada o disminuida de una amplia gama de genes (Jones, P.A. *Cancer Res.* 65:2463 (1996)). En general, la hipometilación de islas de CpG en las regiones promotoras y elementos reguladores da como resultado un aumento de la expresión génica, incluyendo muchos oncogenes (Hanada, M., *et al.*, *Blood* 82:1820 (1993), Feinberg, A.P. y Vogelstein, B. *Nature* 301:89 (1983)). Debido a que la metilación del ADN se correlaciona con el nivel de expresión génica específica en muchos cánceres, sirve como sustituto útil para la obtención del perfil de expresión de tumores (Toyota, M. *et al.*, *Blood* 97: 2823 (2001), Adorjan, P. *et al.* *Nucl. Acids. Res.* 10:e21 (2002)).

### a. PCR con transcriptasa inversa (RT-PCR)

De las técnicas enumeradas anteriormente, el método cuantitativo más sensible y más flexible es RT-PCR, que puede usarse para comparar los niveles de ARNm en diferentes poblaciones de muestra, en tejidos normales y tumorales, con o sin tratamiento con fármaco, para caracterizar patrones de expresión génica, para diferenciar entre ARN estrechamente relacionados y para analizar la estructura del ARN.

La primera etapa es el aislamiento del ARNm de una muestra diana. El material de partida es normalmente ARN total aislado de líneas de células tumorales o tumores humanos, y líneas celulares o tejidos normales correspondientes, respectivamente. Por tanto, el ARN pueden aislarse de una variedad de tumores primarios, incluyendo líneas de células tumorales o tumor de mama, pulmón, colon, próstata, cerebro, hígado, riñón, páncreas, bazo, timo, testículos, ovario, útero, etc., con ADN reunido de donantes sanos. Si la fuente de ARNm es un tumor primario, puede extraerse ARNm, por ejemplo de muestras archivadas de tejido congelado o fijado (por ejemplo fijado en formalina) y embebido en parafina.

En la técnica se conocen bien métodos generales para la extracción de ARNm y se dan a conocer en libros de texto convencionales de biología molecular, incluyendo Ausubel *et al.*, *Current Protocols of Molecular Biology*, John Wiley and Sons (1997). Los métodos para la extracción de ARN de tejidos embebidos en parafina se dan a conocer, por ejemplo, en Rupp y Locker, *Lab Invest.* 56:A (1987), y De Andrés *et al.*, *BioTechniques* 18:42044 (1995). En particular, puede realizarse el aislamiento del ARN usando un kit de purificación, un conjunto de tampones y proteasa procedentes de fabricantes comerciales, tales como Qiagen, según las instrucciones del fabricante. Por

ejemplo, puede aislarse el ARN total de células en cultivo usando minicolumnas RNeasy de Qiagen. Otros kits de aislamiento de ARN disponibles comercialmente incluyen el kit de purificación de ARN y ADN completo MasterPure™ (EPICENTRE®, Madison, WI) y el kit de aislamiento de ARN de bloques de parafina (Ambion, Inc.). El ARN total de muestras de tejido puede aislarse usando ARN Stat-60 (Tel-Test). El ARN preparado a partir de tumores puede aislarse, por ejemplo, mediante centrifugación en gradiente de densidad de cloruro de cesio.

Puesto que el ARN no puede servir como molde para la PCR, la primera etapa en la obtención del perfil de expresión génica mediante RT-PCR es la transcripción inversa del molde de ARN para dar ADNc, seguido por su amplificación exponencial en una reacción de PCR. Las dos transcriptasas inversas más comúnmente usadas son la transcriptasa inversa de virus de mieloblastosis aviar (RT-VMA) y la transcriptasa inversa de virus de leucemia murina de Moloney (RT-VLMM). Normalmente, la etapa de transcripción inversa se realiza usando cebadores específicos, hexámeros al azar o cebadores de oligo-dT, dependiendo de las circunstancias y el objetivo de la obtención del perfil de expresión. Por ejemplo, el ARN extraído puede transcribirse de modo inverso usando un kit de PCR de ARN GeneAmp (Perkin Elmer, CA, EE.UU.), siguiendo las instrucciones del fabricante. Entonces, el ADNc derivado puede usarse como molde en la reacción de PCR posterior.

Aunque la etapa de PCR puede usar una variedad de ADN polimerasas dependientes de ADN termoestables, normalmente emplea la Taq ADN polimerasa, que tiene una actividad nucleasa en 5'-3' pero carece de actividad endonucleasa de corrección de lectura en 3'-5'. Por tanto, la PCR TaqMan® normalmente utiliza la actividad nucleasa en 5' de la Taq o Tth polimerasa para hidrolizar una sonda de hibridación unida a su amplicón diana, pero puede usarse cualquier enzima con una actividad nucleasa en 5' equivalente. Se usan dos cebadores de oligonucleótidos para generar un amplicón típico de una reacción de PCR. Se designa un tercer oligonucleótido, o sonda, para detectar la secuencia de nucleótidos ubicada entre los dos cebadores de PCR. La sonda no puede extenderse mediante la enzima Taq ADN polimerasa, y está marcada con un colorante fluorescente indicador y un colorante fluorescente extintor. Cualquier emisión inducida por láser del colorante indicador se extingue mediante el colorante de extinción cuando los dos colorantes se ubican bastante juntos cuando están en la sonda. Durante la reacción de amplificación, la enzima Taq ADN polimerasa escinde la sonda de una manera dependiente de la sonda. Los fragmentos de sonda resultantes se disocian en disolución, y la señal del colorante indicador liberado está libre del efecto de extinción del segundo fluoróforo. Se libera una molécula del colorante indicador por cada nueva molécula sintetizada, y la detección del colorante indicador no extinguido proporciona la base para la interpretación cuantitativa de los datos.

Puede realizarse RT-PCR TaqMan® usando equipos disponibles comercialmente, tales como, por ejemplo, ABI PRISM 7700™ Sequence Detection System™ (Perkin-Elmer-Applied Biosystems, Foster City, CA, EE.UU.), o Lightcycler (Roche Molecular Biochemicals, Mannheim, Alemania). En una realización preferida, se ejecuta el procedimiento con 5'-nucleasa en un dispositivo de PCR cuantitativa en tiempo real tal como el ABI PRISM 7700™ Sequence Detection System™. El sistema consiste en un termociclador, un láser, un dispositivo acoplado a carga (CCD), una cámara y un ordenador. El sistema amplifica muestras en un formato de 96 pocillos en un termociclador. Durante la amplificación, se detecta la señal fluorescente inducida por láser en el CCD. El sistema incluye software para ejecutar el instrumento y para analizar los datos.

Los datos del ensayo de 5'-nucleasa se expresan inicialmente como  $C_T$ , o el ciclo umbral. Tal como se trató anteriormente, se registraron los valores de fluorescencia durante cada ciclo y representan la cantidad de producto amplificado hasta este punto en la reacción de amplificación. El punto cuando la señal fluorescente se registró en primer lugar como estadísticamente significativa es el ciclo umbral ( $C_T$ ).

Para minimizar errores y el efecto de variación muestra a muestra, la RT-PCR se realiza habitualmente usando uno o más genes de referencia como patrones internos. El patrón interno ideal se expresa a un nivel constante entre los diferentes tejidos, y no se ve afectado por el tratamiento experimental. Los ARN más frecuentemente usados para normalizar patrones de expresión génica son ARNm para los genes de mantenimiento gliceraldehído-3-fosfato-deshidrogenasa (GAPD) y  $\beta$ -actina (ACTB).

Una variación más reciente de la técnica de RT-PCR es la PCR cuantitativa en tiempo real, que mide la acumulación del producto de PCR a través de una sonda fluorogénica bimarcada (es decir, sonda TaqMan®). La PCR en tiempo real es compatible tanto con PCR competitiva cuantitativa, en la que se usa el competidor interno para cada secuencia diana para normalización, como con la PCR comparativa cuantitativa que usa un gen de normalización contenido dentro de la muestra, o un gen de mantenimiento para la RT-PCR. Para detalles adicionales véase, por ejemplo Held *et al.*, Genome Research 6:986-994 (1996).

Las etapas de un protocolo representativo para obtener el perfil de expresión génica usando tejidos fijados, embebidos en parafina como la fuentes de ARN, que incluyen el aislamiento, la purificación, la extensión con cebadores y la amplificación del ARNm se facilitan en diversos artículos de revistas publicadas {por ejemplo: T.E. Godfrey *et al.* J. Molec. Diagnostics 2: 84-91 (2000); K. Specht *et al.*, Am. J. Pathol. 158: 419-29 (2001); Cronin *et al.*, Am J Pathol 164:35-42 (2004)}. En resumen, un procedimiento representativo se inicia con el corte de secciones de aproximadamente 10  $\mu$ m de espesor de muestras de tejido tumoral embebidas en parafina. Entonces, se extrae el ARN y se eliminan las proteínas y el ADN. Tras el análisis de la concentración del ARN, pueden incluirse etapas

de reparación y/o amplificación de ARN, si es necesario, y el ARN se somete a transcripción inversa usando promotores específicos de gen seguido por RT-PCR.

b. Micromatrices

5 La expresión génica diferencial también puede identificarse o confirmarse usando la técnica de micromatrices. Por tanto, el perfil de expresión de genes asociados con cáncer de mama puede medirse en tejido tumoral o bien reciente o bien embebido en parafina, usando la tecnología de micromatrices. En este método, las secuencias de polinucleótido de interés (incluyendo ADNc y oligonucleótidos) se disponen en placas o en matrices, en un sustrato  
10 de microchip. Entonces se hibridan las secuencias dispuestas en matrices con sondas de ADN específicas de células o tejidos de interés. Al igual que en el método de RT-PCR, la fuente de ARNm normalmente es ARN total aislado de líneas de células tumorales o tumores humanos, y líneas celulares o tejidos normales correspondientes. Por tanto, el ARN puede aislarse de una variedad de líneas de células tumorales o tumores primarios. Si la fuente de ARNm es un tumor primario, el ARNm puede extraerse, por ejemplo, de muestras de tejido congelado o embebido  
15 en parafina y fijado (por ejemplo fijado en formalina) archivadas, que se preparan de manera rutinaria y se conservan en la práctica clínica diaria.

En una realización específica de la técnica de micromatrices, se aplican insertos amplificados por PCR de clones de ADNc a un sustrato en una matriz densa. Preferiblemente, se aplican al menos 10.000 secuencias de nucleótidos al  
20 sustrato. Los genes dispuestos en micromatrices, inmovilizados sobre el microchip a 10.000 elementos cada uno, son adecuados para la hibridación en condiciones rigurosas. Pueden generarse sondas de ADNc marcadas de modo fluorescente a través de la incorporación de nucleótidos fluorescentes mediante transcripción inversa del ARN extraído de tejidos de interés. Las sondas de ADNc marcadas aplicadas al chip hibridan con especificidad a cada mancha de ADN sobre la matriz. Tras el lavado riguroso para eliminar sondas unidas de manera no específica, se  
25 explora el chip mediante microscopía láser confocal o mediante otro método de detección, tal como una cámara de CCD. La cuantificación de la hibridación de cada elemento dispuesto en matriz permite la evaluación de la abundancia del ARNm correspondiente. Con la fluorescencia bicolor, se hibridan las sondas de ADNc marcadas separadamente generadas a partir de dos fuentes de ARN por emparejamiento a la matriz. Por tanto se determina simultáneamente la abundancia relativa de los transcritos procedentes de las dos fuentes correspondientes para  
30 cada gen especificado. La escala miniaturizada de la hibridación proporciona una evaluación conveniente y rápida del patrón de expresión para un gran número de genes. Se ha demostrado que tales métodos tienen la sensibilidad requerida para detectar transcritos raros, que se expresan en pocas copias por célula, y para detectar de manera reproducible diferencias de al menos aproximadamente dos veces en los niveles de expresión (Schena *et al*, Proc. Natl. Acad. Sci. USA 93(2):106-149 (1996)). Los análisis por micromatrices pueden realizarse mediante equipos  
35 disponibles comercialmente, siguiendo los protocolos del fabricante, tal como usando la tecnología Affymetrix GenChip, o la tecnología de micromatrices de Incyte.

El desarrollo de métodos de micromatriz para el análisis a gran escala de la expresión génica hace posible buscar sistemáticamente marcadores moleculares de clasificación de cáncer y predicción de desenlace en una variedad de  
40 tipos de tumor.

c. Análisis en serie de la expresión génica (SAGE)

El análisis en serie de la expresión génica (SAGE) es un método que permite el análisis simultáneo y cuantitativo de  
45 un gran número de transcritos génicos, sin necesidad de proporcionar una sonda de hibridación individual para cada transcrito. En primer lugar, se genera un marcador de secuencia corta (aproximadamente 10 -14 pb) que contiene suficiente información para identificar de manera excepcional un transcrito, siempre que se obtenga el marcador a partir de una posición única dentro de cada transcrito. Entonces, muchos transcritos se unen entre sí para formar moléculas en serie grandes, que pueden secuenciarse, revelando la identidad de múltiples marcadores  
50 simultáneamente. El patrón de expresión de cualquier población de transcritos puede evaluarse cuantitativamente determinando la abundancia de marcadores individuales, e identificando el gen correspondiente para cada marcador. Para más detalles véase, por ejemplo Velculescu *et al*, Science 270:484-487 (1995); y Velculescu *et al*, Cell 88:243-51 (1997).

d. Análisis de la expresión génica mediante secuenciación masiva en paralelo de firmas (MPSS)

Este método, descrito por Brenner *et al*, Nature Biotechnology 18:630-634 (2000), es un enfoque de secuenciación que combina la secuenciación de firmas no basada en gel con clonación *in vitro* de millones de moldes en microperlas separadas de 5 µm de diámetro. En primer lugar, se construye una biblioteca de microperlas de moldes  
60 de ADN mediante clonación *in vitro*. Esto va seguido por el ensamblaje de una matriz plana de las microperlas que contienen el molde en una célula de flujo a una densidad alta (normalmente mayor de 3 x 10<sup>6</sup> microperlas/cm<sup>2</sup>). Los extremos libres de los moldes clonados en cada microperla se analizan simultáneamente, usando un método de secuenciación de firmas basado en fluorescencia que no requiere separación de fragmentos de ADN. Se ha demostrado que este método proporciona simultáneamente y de manera precisa, en una sola operación, cientos de  
65 secuencias de firmas de genes de una biblioteca de ADNc de levadura.

e. Descripción general del aislamiento, purificación y amplificación del ARNm

Las etapas de un protocolo representativo para obtener el perfil de expresión génica usando tejidos fijados, embebidos en parafina, como fuente de ARN, incluyendo aislamiento, purificación, extensión con cebadores y amplificación de ARNm se proporcionan en diversos artículos de revistas publicados (por ejemplo: T.E. Godfrey *et al.*, J. Molec. Diagnostics 2: 84-91 [2000]; K. Specht *et al.*, Am. J. Pathol. 158: 419-29 [2001]). En resumen, un procedimiento representativo se inicia con el corte de secciones de aproximadamente 10  $\mu\text{m}$  de espesor de muestras de tejido tumoral embebidas en parafina. Entonces, se extrae el ARN, y se elimina el ADN y las proteínas. Tras el análisis de la concentración de ARN, pueden incluirse etapas de reparación y/o amplificación de ARN, si es necesario, y se somete a transcripción inversa el ARN usando promotores específicos del gen seguido por RT-PCR. Finalmente, se analizan los datos para identificar la(s) mejor(es) opción(es) de tratamiento disponible(s) para el paciente basándose en el patrón de expresión génica característico identificado en la muestra de tumor examinada, dependiente de la probabilidad predicha de recurrencia de cáncer.

f. Conjunto de genes de cáncer de mama, subsecuencias génicas sometidas a ensayo y aplicación clínica de los datos de expresión génica

Un aspecto importante de la presente invención es usar la expresión medida de determinados genes por tejido de cáncer de mama para proporcionar información de pronóstico o predictiva. Para este fin, es necesario corregir (normalizar) tanto diferencias en la cantidad de ARN sometido a ensayo como la variabilidad de la calidad del ARN usado. Por tanto, el ensayo normalmente mide e incorpora la expresión de determinados genes de normalización, incluyendo genes de mantenimiento bien conocidos, tales como  $\beta$ -actina, GAPD, GUS, RPLO y TFRC, tal como se muestran en el ejemplo a continuación. Alternativamente, la normalización puede basarse en la señal media o mediana ( $C_T$ ) de todos los genes sometidos a ensayo o en un gran subconjunto de los mismos (enfoque de normalización global). En lo sucesivo, a menos que se observe lo contrario, expresión génica significa expresión normalizada.

g. Diseño de cebadores y sondas de PCR basados en intrones

Según un aspecto de la presente invención, se diseñan los cebadores y sondas de PCR basadas en secuencias de intrones presentes en el gen que va a amplificarse. Por consiguiente, la primera etapa en el diseño del cebador/la sonda es la descripción de las secuencias de intrones dentro de los genes. Esto puede realizarse mediante software disponible públicamente, tal como el ADN BLAT desarrollado por Kent, W.J., Genome Res. 12(4):656-64 (2002), o mediante el software BLAST incluyendo sus variaciones. Las etapas posteriores siguen métodos bien establecidos de diseño de cebadores y sondas de PCR.

Con el fin de evitar señales no específicas, es importante enmascarar secuencias repetitivas dentro de los intrones cuando se diseñan los cebadores y las sondas. Esto puede conseguirse fácilmente usando el programa Repeat Masker disponible en línea a través del Baylor College of Medicine, que examina secuencias de ADN frente a una biblioteca de elementos repetitivos y devuelve una secuencia de consulta en la que los elementos repetitivos están enmascarados. Entonces, pueden usarse las secuencias de intrones enmascaradas para diseñar secuencias de cebadores y sondas usando cualquier paquete de diseño de cebadores/sondas disponible comercialmente o en cualquier caso públicamente, tales como Primer Express (Applied Biosystems); MGB assay-by-design (Applied Biosystems); Primer3 (Steve Rozen y Helen J. Skaletsky (2000) Primer3 en el WWW para usuarios generales y para programadores biólogos. En: Krawetz S, Misener S (eds) Bioinformatics Methods and Protocols: Methods in Molecular Biology. Humana Press, Totowa, NJ, págs. 365-386).

Los factores más importantes considerados en el diseño de cebadores de PCR incluyen la longitud del cebador, la temperatura de fusión ( $T_m$ ), y el contenido en G/C, la especificidad, las secuencias de cebador complementario y la secuencia de extremo en 3'. En general, los cebadores de PCR óptimos tienen generalmente 17-30 bases de longitud y contienen aproximadamente el 20-80%, tal como, por ejemplo, aproximadamente el 50-60% de bases G+C. Normalmente se prefieren  $T_m$  de entre 50 y 80°C, por ejemplo de aproximadamente 50 a 70°C.

Para consultar directrices adicionales para el diseño de cebadores y sondas de PCR véase, por ejemplo Dieffenbach, CW. *et al.*, "General Concepts for PCR Primer Design" en: PCR Primer, A Laboratory Manual, Cold Spring Harbor Laboratory Press, Nueva York, 1995, págs. 133-155; Innis y Gelfand, "Optimization of PCRs" en: PCR Protocols, A Guide to Methods and Applications, CRC Press, Londres, 1994, págs. 5-11; y Plasterer, T.N. Primer select: Primer and probe design. Methods Mol. Biol. 70:520-527 (1997).

60 B.3 Algoritmos y métodos estadísticos

La presente invención aprovecha determinados algoritmos y métodos estadísticos, que se describen en la solicitud en tramitación junto con la presente n.º de serie 10/883.303.

65 Cuando se usa RT-PCR cuantitativa (qRT-PCR) para medir los niveles de ARNm, las cantidades de ARNm se expresan en unidades de  $C_T$  (ciclo umbral) (Heid *et al.*, Genome Research 6:986-994 (1996)). La suma promediada

de  $C_T$  de ARNm de referencia se establece como un número fijo tal como cero, y cada  $C_T$  de ARNm de prueba medido se proporciona en relación a este punto fijo. Por ejemplo, si para una muestra de tumor de cierto paciente se halla que el promedio de  $C_T$  de los 5 genes de referencia es de 31 y se halla que el  $C_T$  del gen de prueba X es de 35, el valor notificado para el gen X es de -4 (es decir 31-35).

5 Como primera etapa tras la determinación cuantitativa de los niveles de ARNm, los genes identificados en la muestra de tumor y que se sabe que están asociados con la patología molecular del cáncer se agrupan en subconjuntos. Por tanto, los genes que se sabe que están asociados con la proliferación constituirán el "grupo de proliferación" (eje o subconjunto). Los genes que se sabe que están asociados con la invasión del cáncer constituirán el "grupo de invasión" (eje o subconjunto). Los genes asociados con la(s) ruta(s) del receptor del factor de crecimiento clave(s) constituirán el "grupo de factor de crecimiento" (eje o subconjunto), también denominado grupo de GRB7. Los genes que se sabe que están implicados con la activación o señalización a través del receptor de estrógenos (ESR1) constituirán el "grupo de receptor de estrógenos (ER)" (eje o subconjunto), etc. Esta lista de subconjuntos es, por supuesto, no limitativa. Los subconjuntos (ejes) creados dependerán del cáncer particular, es decir cáncer de mama, próstata, de páncreas, de pulmón, etc. En general, los genes cuyas expresiones se sabe que están correlacionadas entre sí, o que se sabe que están implicados en la misma ruta se agrupan juntos.

20 En la siguiente etapa, el nivel de tumor medido de cada ARNm en un subconjunto se multiplica por un coeficiente que refleja su contribución dentro de un conjunto relativa con respecto al riesgo de recurrencia de cáncer y se suma este producto a los otros productos entre los niveles de ARNm en el subconjunto y sus coeficientes, dando un término, por ejemplo un término de proliferación, un término de invasión, un término de factor de crecimiento, etc. Por ejemplo, en el caso de cáncer de mama invasivo negativo para ganglios linfáticos el término de factor de crecimiento es (de 0,45 a 1,35) x GRB7 + (de 0,05 a 0,15) x ErbB2, tal como, por ejemplo 0,9 x GRB7 + 0,1 x ERBB2 (véase el ejemplo más adelante).

25 La contribución de cada término a la puntuación de recurrencia global se pondera mediante el uso de un coeficiente. Por ejemplo, en el caso de cáncer de mama invasivo negativo para ganglios linfáticos, el coeficiente del término de factor de crecimiento puede estar entre 0,23 y 0,70.

30 Adicionalmente, para algunos términos, tales como los términos de factor de crecimiento y de proliferación, se realiza una etapa adicional. Si la relación entre el término y el riesgo de recurrencia no es lineal, se usa una transformada funcional no lineal del término, tal como un umbral. Por tanto, en el cáncer de mama invasivo negativo para ganglios linfáticos, cuando el término de factor de crecimiento se halla a <8 el valor se fija a 8.

35 La suma de los términos obtenidos proporciona la puntuación de recurrencia (RS), que predice la probabilidad de recurrencia de cáncer en el transcurso normal de la enfermedad.

40 La escala de RS generada por el algoritmo de la presente invención puede ajustarse de diversas maneras. Por tanto, el intervalo podría seleccionarse de manera que la escala vaya desde 0 hasta 10, desde 0 hasta 50 o desde 0 hasta 100, por ejemplo.

45 Por ejemplo, en el enfoque de graduación particular descrito en el ejemplo a continuación, se calcula la puntuación graduada en una escala de 0 a 100. Por conveniencia, se añaden 10 unidades  $C_T$  a cada valor de  $C_T$  medido, y se calcula la RS no graduada tal como se describió anteriormente. Las ecuaciones para calcular la RS no graduada y la RS graduada se proporcionan en el siguiente ejemplo.

50 En el cálculo de la puntuación de recurrencia, o cualquier variable usada para calcular la puntuación de recurrencia, puede sustituirse cualquier gen por otro gen que se coexpresa en un conjunto de al menos 30 muestras de pacientes diferentes de ese tipo de tumor (tal como cáncer de mama) con el primer gen en el cáncer particular sometido a prueba con un coeficiente de Pearson  $\geq 0,5$ . De modo similar, cualquier gen individual, o gen dentro de un grupo de genes (subconjunto) incluido en los métodos de pronóstico y predictivos de la presente invención puede sustituirse por otro gen que se coexpresa con el primer gen en el cáncer particular sometido a prueba con un coeficiente de Pearson  $\geq 0,5$ .

#### 55 B.4 Tratamiento con fármacos antiestrógenos

60 Se sabe que los estrógenos promueven el crecimiento de algunos cánceres, tales como cánceres de mama, especialmente los que expresan el receptor de estrógenos (ESR1). Se han desarrollado varias terapias para bloquear el efecto de los estrógenos, o para disminuir los niveles de estrógenos en tal paciente, especialmente en pacientes con cáncer de mama positivo para ESR1.

65 Los fármacos antiestrógenos pueden clasificarse generalmente o bien como antagonistas de la unión de estrógenos al receptor de estrógenos o bien como inhibidores de la biosíntesis de estrógeno, tales como inhibidores de aromatasas.

El fármaco antiestrógenos usado más comúnmente TAM, que pertenece a la clase de antagonistas de la unión de

estrógenos, se toma normalmente por vía oral, una vez al día durante cinco años, tras la resección quirúrgica del cáncer de mama y/o quimio o radioterapia. Los estudios clínicos han demostrado que el uso de TAM como una terapia adyuvante tras la cirugía reduce el riesgo de recurrencia de cáncer, sin embargo, la respuesta de pacientes positivos para ESR1 a este tratamiento varía, y no hay factores pronóstico claros disponibles del grado de respuesta.

5 Otros fármacos antiestrógenos incluyen raloxifeno que, como TAM, bloquea el efecto de los estrógenos en tejido de mama y en cáncer de mama; y citrato de toremifeno, que está estrechamente relacionado con TAM, y puede ser una opción para mujeres posmenopáusicas con cáncer de mama metastásico.

10 El anastrozol, un inhibidor de aromatasas, actúa evitando que los estrógenos activen su receptor, bloqueando una enzima necesaria para la producción de estrógenos. El anastrozol es actualmente una opción para mujeres cuyo cáncer de mama avanzado continúa creciendo durante o tras el tratamiento con TAM.

15 El acetato de megestrol se usa normalmente para el tratamiento hormonal del cáncer de mama avanzado, habitualmente para mujeres cuyos cánceres no responden a TAM.

Todos los tratamientos están limitados por efectos secundarios graves, y por la falta de factores pronóstico fiables de respuesta del paciente, que podrían permitir a un médico hacer un análisis de riesgo-beneficio inteligente antes de recomendar un tratamiento particular.

## 20 B.5 Quimioterapia contra el cáncer

Los agentes quimioterápicos usados en el tratamiento del cáncer pueden dividirse en varios grupos, dependiendo de su mecanismo de acción. Algunos agentes quimioterápicos dañan directamente el ADN y el ARN. Mediante la alteración de la replicación del ADN, tales agentes quimioterápicos o bien detienen completamente la replicación, o bien dan como resultado la producción de ADN o ARN sin sentido. Esta categoría incluye, por ejemplo, cisplatino (Platinol®), daunorubicina (Cerubidine®), doxorubicina (Adriamycin®) y etopósido (VePesid®). Otro grupo de agentes quimioterápicos contra el cáncer interfieren con la formación de nucleótidos o desoxirribonucleótidos, de modo que se bloquea la síntesis de ARN y la replicación celular. Ejemplos de fármacos de esta clase incluyen metotrexato (Abitrexate®), mercaptopurina (Purinethol®), fluorouracilo (Adrucil®) e hidroxiurea (Hydrea®). Una tercera clase de agentes quimioterápicos efectúa la síntesis o la ruptura de los husos mitóticos, y como resultado, interrumpen la división celular. Ejemplos de fármacos de esta clase incluyen vinblastina (Velban®), vincristina (Oncovin®) y taxanos, tales como, paclitaxel (Taxol®), y docetaxel (Taxotere®). El docetaxel está aprobado en la actualidad en los Estados Unidos para tratar pacientes con cáncer de mama metastásico o localmente avanzado tras el fracaso de la quimioterapia anterior, y pacientes con cáncer de pulmón de células no pequeñas metastásico o localmente avanzado tras el fracaso de la quimioterapia a base de platino anterior. La predicción de la respuesta del paciente a todos estos agentes quimioterápicos y a otros está específicamente dentro del alcance de la presente invención.

40 En una realización específica, la quimioterapia incluye el tratamiento con un derivado de taxano. Los taxanos incluyen, sin limitación, paclitaxel (Taxol®) y docetaxel (Taxotere®), que se usan ampliamente en el tratamiento del cáncer. Tal como se trató anteriormente, los taxanos afectan a estructuras celulares denominadas microtúbulos, que desempeñan un papel importante en las funciones celulares. En el crecimiento celular normal, los microtúbulos se forman cuando una célula comienza a dividirse. Una vez que la célula detiene la división, los microtúbulos se rompen o destruyen. Los taxanos detienen la ruptura de los microtúbulos; lo que bloquea la división de la célula cancerosa.

En otra realización específica, la quimioterapia incluye el tratamiento con un derivado de antraciclina, tal como, por ejemplo, doxorubicina, daunorubicina y aclacinomicina.

50 En una realización específica adicional, la quimioterapia incluye el tratamiento con un inhibidor de topoisomerasa, tal como, por ejemplo, camptotecina, topotecán, irinotecán, 20-S-camptotecina, 9-nitro-camptotecina, 9-amino-camptotecina o GI147211.

55 Se contempla específicamente el tratamiento con cualquier combinación de estos y otros fármacos quimioterápicos.

La mayoría de pacientes reciben quimioterapia inmediatamente tras la extirpación quirúrgica del tumor. Este enfoque se denomina comúnmente terapia adyuvante. Sin embargo, la quimioterapia puede administrarse también antes de la cirugía, denominándose tratamiento neoadyuvante. Aunque el uso de quimioterapia neoadyuvante se origina del tratamiento del cáncer de mama avanzado e inoperable, también ha ganado aceptación en el tratamiento de otros tipos de cánceres. La eficacia de la quimioterapia neoadyuvante se ha sometido a prueba en varios ensayos clínicos. En el ensayo B-18 del Nacional Surgical Adjuvant Breast and Bowel Project (NSAB B-18) (Fisher *et al.*, J Clin. Oncology 15:2002-2004 (1997); Fisher *et al.*, J Clin. Oncology 16:2672-2685 (1998)) la terapia neoadyuvante se llevó a cabo con una combinación de adriamicina y ciclofosfamida ("régimen AC"). En otro ensayo clínico, la terapia neoadyuvante se administró usando una combinación de 5-fluorouracilo, epirubicina y ciclofosfamida ("régimen FEC") (van Der Hage *et al.*, J. Clin. Oncol. 19:4224-4237 (2001)). Ensayos clínicos más nuevos también

han usado regímenes de tratamiento neoadyuvante que contienen taxanos. Véase, por ejemplo Holmes *et al.*, J. Natl. Cancer Inst. 83:1797-1805 (1991) y Moliterni *et al.*, Seminars in Oncology, 24:S17-10-S-17-14 (1999). Para información adicional sobre la quimioterapia neoadyuvante para cáncer de mama véase, Cleator *et al.*, Endocrine-Related Cancer 9:183-195 (2002).

5

#### B.6 Kits

Los materiales para su uso en los métodos de la presente invención son adecuados para la preparación de kits producidos según procedimientos bien conocidos. Por tanto, los kits comprenden agentes, que pueden incluir sondas y/o cebadores específicos de gen o selectivos de gen, para cuantificar la expresión de los genes dados a conocer para predecir el desenlace pronóstico o la respuesta al tratamiento. Tales kits pueden contener opcionalmente reactivos para la extracción de ARN de las muestras de tumor, en particular muestras de tejido fijado, embebido en parafina y/o reactivos para la amplificación de ARN. Además, los kits pueden comprender opcionalmente el/los reactivo(s) con una descripción o etiqueta de identificación o instrucciones relacionadas con su uso en los métodos de la presente invención. Los kits pueden comprender recipientes (incluyendo placas de microtitulación adecuadas para su uso en una implementación automatizada del método), cada una con uno o más de los diversos reactivos (normalmente en forma concentrada) utilizados en los métodos, que incluyen, por ejemplo, micromatrices fabricadas previamente, tampones, los nucleótidos trifosfato apropiados (por ejemplo, dATP, dCTP, dGTP y dTTP; o rATP, rCTP, rGTP y UTP), transcriptasa inversa, ADN polimerasa, ARN polimerasa y una o más sondas y cebadores de la presente invención (por ejemplo, poli(T) de longitud apropiada o cebadores aleatorios unidos a un promotor reactivo con la ARN polimerasa). Los algoritmos matemáticos usados para estimar o cuantificar la información de pronóstico o predictiva también son componentes apropiados de los kits.

10

15

20

Los métodos proporcionados por la presente invención también pueden automatizarse en su totalidad o en parte.

25

Todos los aspectos de la presente invención también pueden ponerse en práctica de manera que se incluya un número limitado de genes adicionales que se coexpresan con los genes dados a conocer, por ejemplo tal como se pone de manifiesto mediante los altos coeficientes de correlación de Pearson, en una prueba de pronóstico o predictiva además de y/o en lugar de los genes dados a conocer.

30

Habiendo descrito la invención, la misma se entenderá más fácilmente haciendo la referencia al siguiente ejemplo, que se proporciona a modo de ilustración, y no pretende limitar la invención en modo alguno.

#### Ejemplo

35

Estudio de la relación entre la expresión génica y el pronóstico y la probabilidad de respuesta beneficiosa con respecto a tamoxifeno en pacientes con cáncer de mama temprano

#### Métodos

40

Este estudio emplea tejido y datos del estudio B-14 de NSABP: "A Clinical Trial to Assess Tamoxifen in Patients with Primary Breast Cancer and Negative Axillary Nodes Whose Tumors are Positive for Estrogen Receptors" (Ensayo clínico para evaluar el tamoxifeno en pacientes con cáncer de mama primario y ganglios axilares negativos cuyos tumores son positivos para receptores de estrógenos). Los resultados de este ensayo los notificaron Fisher B, Costantino JP, Redmond CK, *et al.*: Endometrial cancer in TAM-treated breast cancer: Findings from the National Surgical Adjuvant Breast and Bowel Project (NSABP) B-14. J Natl Cancer Inst 86:527-537 (1994).

45

1.1 Se analizaron muestras fijadas, embebidas en parafina, de tejido de tumor de mama de hasta 450 pacientes, que se trataron a la entrada del estudio con placebo solo en el estudio B-14. Para cada paciente evaluable se evaluó cuantitativamente la expresión de 16 genes relacionados con cáncer y 5 genes de referencia mediante RT-PCR. Se evaluó la relación entre la recurrencia de la enfermedad y (a) la puntuación de recurrencia, (b) la expresión de genes en particular grupos de genes particulares, o (c) la expresión de genes individuales.

50

1.2 Criterios de inclusión.

55

1.2.1 Incluidas en el estudio B-14 de NSABP: "Ensayo clínico para evaluar el tamoxifeno en pacientes con cáncer de mama primario y ganglios axilares negativos cuyos tumores son positivos para receptores de estrógenos".

1.2.2 Aleatorización a placebo o a TAM en la parte controlada por placebo del estudio.

60

1.2.3 Clínicamente elegible con seguimiento.

1.3 Criterios de exclusión.

65

1.3.1 Sin bloqueo del tumor disponible desde el diagnóstico inicial en el archivo de NSABP.

1.3.2 Sin tumor o tumor muy pequeño (grupo 1) en bloqueo tal como se evalúa mediante examen del portaobjetos con H&E.

1.3.3 ARN insuficiente (<275 ng) para el análisis por RT-PCR.

1.3.4 C<sub>T</sub> no normalizado promedio para los 5 genes de referencia >35.

1.4 Panel génico.

1.4.1 Se llevó a cabo el análisis de 16 genes relacionados con cáncer y 5 genes de referencia enumerados en la tabla 1 usando RT-PCR cuantitativa.

1.4.2 Supervivencia del paciente.

La supervivencia sin recurrencia distante (DRFS) se basa en el tiempo (en años) desde la cirugía hasta la primera recurrencia distante. La enfermedad contralateral, otros segundos cánceres primarios y muertes antes de la recurrencia distante se considerarán acontecimientos de exclusión. Para el análisis principal, se ignora la recurrencia en mama homolateral, recurrencia en pared torácica local y la recurrencia regional, es decir, no se consideran ni como acontecimiento o ni como acontecimiento de exclusión.

1.4.3 Expresión génica.

Los niveles de expresión de los 21 genes enumerados en la tabla 1 se notifican como valores procedentes del ensayo de GHI. Se normalizaron los valores de expresión génica en relación con la media de los genes de referencia. Para cada gen relacionado con cáncer, se obtuvieron mediciones del ciclo umbral (C<sub>T</sub>) mediante RT-PCR, y se normalizaron en relación con un conjunto de cinco genes de referencia enumerados en la tabla 1. Se sabe que los genes de referencia son relativamente invariables en el cáncer de mama así como en diversas muestras y condiciones de procedimientos, haciéndolos útiles para normalizar efectos externos. Las mediciones de expresión normalizadas con respecto a la referencia normalmente oscilan entre 0 y 15, en las que un aumento de una unidad generalmente refleja un aumento de 2 veces en la cantidad de ARN.

1.4.4 Genes relacionados con cáncer y genes de referencia.

Tabla 1. Panel de expresión génica

Genes relacionados con cáncer	Número de registro	Genes de referencia	Número de registro
BAG1	NM_004323	ACTB	NM_001101
BCL2	NM_000633	GAPD	NM_002046
CCNB1	NM_031966	GUSB	NM_000181
CD68	NM_001251	RPLP0	NM_001002
SCUBE2	NM_020974	TRFC	NM_003234
CTSL2	NM_001333		
ESR1	NM_000125		
GRB7	NM_005310		
GSTM1	NM_000561		
ERBB2	NM_004448		
MKI67	NM_002417		
MYBL2	NM_002466		
PGR	NM_000926		
STK6	NM_003600		
MMP11	NM_005940		
BIRC5	NM_001168		

1.4.5 Puntuación de recurrencia

La puntuación de recurrencia (RS) en una escala desde 0 hasta 100 se deriva de las mediciones de expresión normalizadas con respecto a la referencia tal como sigue:

$$RS_u \text{ (puntuación de recurrencia no graduada)} = \begin{aligned} & 0,47 \times \text{puntuación umbral del grupo de GRB7} \\ & - 0,34 \times \text{puntuación del grupo de ESR1} \\ & + 1,04 \times \text{puntuación umbral del grupo de proliferación} \\ & + 0,10 \times \text{puntuación del grupo de invasión} \\ & + 0,05 \times CD68 \\ & - 0,08 \times GSTM1 \\ & - 0,07 \times BAG1 \end{aligned}$$

en la que:

$$\begin{aligned} \text{Puntuación del grupo de GRB7} &= 0,9 \times \text{GRB7} + 0,1 \times \text{ERBB2} \\ \text{Puntuación umbral del grupo de GRB7} &= \begin{cases} 8 & \text{si la puntuación del grupo de GRB7} < 8 \\ \text{puntuación del grupo de GRB7} & \text{en caso contrario} \end{cases} \\ \text{Puntuación del grupo de ESR1} &= (\text{ESR1} + \text{PGR} + \text{BCL2} + \text{SCUBE2})/4 \\ \text{Puntuación del grupo de proliferación} &= (\text{BIRC5} + \text{MKI67} + \text{CCNB1} + \text{STK6})/5 \\ \text{Puntuación umbral del grupo de proliferación} &= \begin{cases} 6,5 & \text{si la puntuación del grupo de proliferación} < 6,5 \\ \text{puntuación del grupo de proliferación} & \text{en caso contrario} \end{cases} \\ \text{Puntuación del grupo de invasión} &= (\text{CTSL2} + \text{MMP11})/2 \end{aligned}$$

5

La  $RS_u$  (puntuación de recurrencia no graduada) se vuelve a graduar entonces para que esté entre 0 y 100:

$$RS = \begin{cases} 0 & \text{si } 20 \times (RS_u - 6,7) < 0 \\ 20 \times (RS_u - 6,7) & \text{si } 0 \leq 20 \times (RS_u - 6,7) \leq 100 \\ 100 & \text{si } 20 \times (RS_u - 6,7) > 100 \end{cases}$$

- 10 Basándose en un modelo de riesgos proporcionales de Cox, se evaluó la correlación entre la recurrencia de cáncer y la puntuación de recurrencia como una variable continua. La evaluación incluyó adicionalmente el grupo de proliferación, grupo de GRB7, grupo de ESR1, grupo de invasión, y cada uno de los dieciséis genes relacionados con cáncer como variables continuas.
- 15 Con el fin de identificar genes de pronóstico, el objetivo principal era explorar la relación entre la expresión génica y la supervivencia sin recurrencia distante (DRFS) y la supervivencia global (OS) en la rama de pacientes no tratados. La DRFS se basó en el tiempo (en años) desde la cirugía hasta la primera recurrencia distante en la que la enfermedad contralateral, otros segundos cánceres primarios y muertes antes de la recurrencia distante se consideran acontecimientos de exclusión y se ignoraron la recurrencia en mama homolateral, recurrencia en pared torácica local y las recurrencias regionales. Se utilizaron los modelos de riesgos proporcionales de Cox de efecto principal (D. R. Cox (1972) Regression Models and Life-Tables (with discussion). J Royal Statistical Soc. B, 34:187-220) para comparar la contribución adicional de la expresión génica más allá de las variables de pronóstico clínico convencionales, incluyendo edad, tamaño de tumor clínico y grado del tumor. Se utilizó una prueba para comparar el modelo reducido, que excluye la variable de expresión génica, frente al modelo completo de competencia que incluye la variable génica de interés, denominada prueba de razón de probabilidades (Ronald Fisher (1922) "On the Mathematical Foundations of Theoretical Statistics", Phil. Trans. Royal Soc, serie A, 222:326, 1922; Leonard Savage (1962), The Foundations of Statistical Inference (1962)) para identificar genes de pronóstico estadísticamente significativos.
- 20
- 25
- 30 Con el fin de identificar genes predictivos de tratamiento en cáncer de mama, el objetivo principal era explorar la relación entre la expresión génica y la DRFS y la OS en pacientes tratados. Para tales análisis, se utilizaron los datos de las ramas de pacientes tanto tratados como no tratados con el fin de diferenciar genes predictivos del tratamiento de genes puramente de pronóstico. Para identificar genes predictivos de tratamiento para la respuesta a tamoxifeno (TAM), se incluyeron tanto pacientes tratados con placebo como pacientes tratados con TAM del estudio B-14 de NSABP. En ambos estudios, se utilizaron modelos de riesgos proporcionales de Cox para examinar la interacción entre el efecto de tratamiento y la expresión génica. Existe una interacción entre el tratamiento y la expresión génica si el efecto de tratamiento depende del nivel de expresión génica; es decir, si la expresión génica es un factor predictivo de tratamiento. De nuevo, se usó la prueba de razón de probabilidades para identificar genes de tratamiento predictivos estadísticamente significativos comparando el modelo reducido que excluye la expresión génica mediante la interacción del tratamiento frente al modelo completo de competencia que incluye la expresión génica mediante la interacción del tratamiento.
- 35
- 40

Resultados

5 La tabla 2 notifica la razón de riesgo (H.R.) para la recurrencia frente a la variación en la expresión de genes y grupos de genes (ejes), del análisis de pacientes no tratados del ensayo B-14 de NSABP (rama de placebo).

Tabla 2. Análisis univariante de la rama de placebo de B14

Variable	Estimación	Valor de P (LR)	H.R.	CL inferior de HR	CL superior de HRU
<b>Grupo de prolif.</b>	0,424845	0,0005	1,529353	1,204024	1,942588
CCNB1_2	0,437596	0,0007	1,548979	1,204938	1,991252
BIRC5_2	0,288646	0,0009	1,334619	1,12784	1,57931
MYBL2_1	0,247787	0,0026	1,281188	1,090383	1,505381
PGR_6	-0,11734	0,0075	0,889281	0,8118	0,96829
STK6_2	0,349493	0,0079	1,418348	1,098354	1,831589
MKI67_2	0,238862	0,02	1,269803	1,03654	1,55556
GSTM1_1	-0,15	0,032	0,854961	0,742047	0,985056
GAPD_1	0,305394	0,044	1,35716	1,00396	1,834618
<b>Grupo de invasión</b>	0,253247	0,053	1,288201	0,994238	1,669079
RPLPO_2	-0,51811	0,056	0,595643	0,349143	1,016176
<b>Umbral de prolif.</b>	0,52227	0,061	1,685851	1,008598	2,817866
MMP11_3	0,13374	0,083	1,143096	0,982561	1,329859

10 Tal como se muestra en la tabla 2, trece variables (genes más grupos de genes) se correlacionaron con la H.R. de recurrencia con  $p < 0,1$ . Tal como se describió anteriormente, debido a que estas correlaciones están relacionadas con pacientes no tratados, estas variables, por tanto, son factores pronóstico estadísticamente significativos. Las variables de pronóstico son: grupo de proliferación; CCNB1; BIRC5; MYBL2; PGR; STK6; MKI67; GSTM1; GAPD; grupo de invasión; RPLPO; umbral de proliferación; MMP11. El grupo de proliferación, el grupo de invasión y el umbral de proliferación son componentes definidos del algoritmo de puntuación de recurrencia.

15 El aumento de la expresión de diez de estos factores pronóstico se correlaciona con el aumento de la H.R.: grupo de proliferación; CCNB1; BIRC5; MYBL2; STK6; MKI67; GAPD; grupo de invasión; umbral de proliferación; MMP11. Siete de los diez marcadores para un mal pronóstico son genes o conjuntos de genes que indican células en proliferación. Entre estos, el grupo de proliferación (tal como se define en el algoritmo de puntuación de recurrencia) es la variable principal con respecto al valor de P. El aumento de la expresión de tres de los trece factores pronóstico se correlaciona con la disminución de la H.R.: PGR, GSTM1 y RPLPO.

25 Los resultados mostrados en la tabla 2 son significativos no sólo estadísticamente sino también en sus magnitudes de H.R.. Debe observarse que para cualquier marcador un cambio de dos veces en la H.R. de 1,0 cambia la H.R. en el 50%, y que cada valor de HR describe el impacto de un cambio de dos veces en la expresión de la variable de marcador en relación con la expresión promedio en la población de pacientes. Por tanto, por ejemplo, la tabla 2 muestra que para cada aumento de dos veces en la expresión del grupo de proliferación o CCNB1, la H.R. aumenta en aproximadamente el 50% (los límites de confianza del 95% abarcan del ~20% al ~90%).

30 La tabla 3 notifica el análisis de la interacción usando resultados tanto de la rama de placebo como de la de tratamiento con TAM de B14 de NSABP, llevado a cabo para identificar variables que predicen sensibilidad o resistencia a TAM.

Tabla 3. Análisis combinado de interacción de TAM y placebo de B14 (Int)

Variable	Estimación	Valor de P	H.R.	IC del 95% para H.R.	
IntESR1_1	-0,29602	0,000466	0,743775	0,630144	0,877876
IntSCUBE2_2	-0,21592	0,004183	0,805796	0,695108	0,934111
Grupo IntESR1	-0,27804	0,006669	0,757266	0,619458	0,925732
IntTFRC_3	0,664542	0,027624	1,943601	1,075949	3,510933
IntBCL2_2	-0,26181	0,071688	0,769655	0,578846	1,023362

Tal como se muestra, cinco variables (genes más grupos de genes) se correlacionan con la respuesta a TAM con  $p < 0,1$ . Estos son ESR1, SCUBE2, grupo de ESR1, TFRC y BCL2. El más significativo de estos marcadores es ESR1. Para cada incremento de dos veces en la expresión de ESR1, la H.R. para la recurrencia de pacientes tratados con TAM disminuye en aproximadamente el 25% (los límites de confianza del 95% abarcan del ~12% al ~37%). Estos datos pueden usarse como una indicación cuantitativa continua de la probabilidad de respuesta a TAM de un paciente. Esto se muestra gráficamente en la figura 1. Debe hacerse hincapié en que todos los pacientes de B-14 de NSABP se clasificaron como positivos para ER, basándose en un ensayo usado clínicamente, y por tanto candidatos para la prescripción de TAM. Los presentes datos demuestran que dentro de esta población “positiva para ER”, los pacientes experimentan niveles variables de beneficio partiendo de una base predecible como una función de su puntuación de ESR1.

Esta relación entre la expresión de ESR1 y la probabilidad de beneficio terapéutico de TAM también puede representarse como una función de categorías de expresión de ESR1 alta, intermedia o baja. La expresión de ESR1 puede representarse o bien en cuartiles, terciles o bien en otras divisiones. Por ejemplo, los datos obtenidos en este estudio de la población de pacientes de B14 de NSABP proporcionan los siguientes puntos de corte de  $C_T$  normalizados con respecto a la referencia para la división de la expresión de ESR1 en terciles:

Tabla 4. Puntos de corte de puntuación de ESR1 que clasifican la expresión de ESR1 en terciles, usando datos de ensayo de pacientes de B-14

Negativo para ER	Percentil 0-33	Percentil 33-67	Percentil 67-100
ESR1 < 6,5	[6,5, 10,375)	[10,375, 11,41)	ER ≥ 11,41

La figura 2, muestra el beneficio absoluto de TAM entre estos tres grupos de pacientes como una función de la expresión de ESR1 (las líneas horizontales a través de las barras representan los límites de confianza del 95%).

Tal como se muestra, TAM tiene un beneficio absoluto sustancial en los dos tercios de pacientes que expresan ESR1 a los niveles superiores, pero tiene mucho menos impacto en los pacientes en los terciles inferiores de la expresión de ESR1.

De manera similar, la relación entre la expresión de ESR1 y el beneficio de TAM puede representarse como una función de la división de la expresión de ESR1 en cuartiles. La figura 3 muestra el beneficio absoluto del tratamiento con TAM como una función de cuartiles de la expresión de ESR1 (las líneas horizontales a través de las barras representan los límites de confianza del 95%). Los puntos de corte para la expresión de ESR1, son  $C_T$  normalizados con respecto a la referencia derivados del análisis de la población de pacientes de B14 de NSABP. Tal como se muestra, los pacientes en el cuartil inferior de expresión de ESR1 experimentan poco beneficio de TAM.

Pueden usarse los datos de expresión de ESR1 conjuntamente con la puntuación de recurrencia para determinar simultáneamente si a un paciente se le debe prescribir TAM o TAM más quimioterapia. La tabla 6, que presenta datos del análisis de ambas secciones de B14 de NSABP, ilustra esto, así como el formato para expresar estos datos. Para ilustrar cómo podría usarse esta tabla, los pacientes de la categoría de bajo riesgo de recurrencia según se define mediante RS, que expresan ESR1 en el percentil 75 superior, son candidatos lógicos para tratarse con TAM solo. Los pacientes que tienen bajo riesgo de recurrencia, pero que están en el percentil 25 inferior en la expresión de ESR1 y para los que el riesgo de efectos secundarios de TAM es una preocupación particular, pueden considerarse de manera razonable pacientes para los que el tratamiento con TAM es menos apropiado. Los pacientes en riesgo alto que expresan ER por encima del percentil 25 son candidatos lógicos para el tratamiento tanto con TAM como con quimioterapia.

Tabla 5. Distribución de pacientes de NSABP en categorías de expresión de ESR1 y RS

Grupo de RS	Percentil 75-100 de ER	Percentil 50-75 de ER	Percentil 35-50 de ER	Percentil 0-25 de ER	Negativo para ER	Totales de fila
Riesgo bajo	101	89	82	41	0	313
Riesgo intermedio	29	45	37	42	1	154
Riesgo alto	24	23	32	71	28	178
Todos los grupos	1554	157	151	154	29	645

(El número dentro de cada celda es el número de pacientes de B-14 sometidos a ensayo en la categoría indicada).

Basándose en los siguientes puntos de corte de ESR1 normalizados con respecto a la referencia:

Tabla 6

Negativo para ER	Percentil 0-25	Percentil 25-50	Percentil 50-75	Percentil 75-100
ER < 6,5	[6,5, 10,065)	[10,065, 10,855)	[10,855, 11,805)	ER ≥ 11,805

Las siguientes ilustraciones indican cómo puede usarse la tabla 5 para tomar decisiones acerca del tratamiento de un paciente con antiestrógenos y/o quimioterapia.

5

**Terapia hormonal recomendada**

Grupo RS	Percentil 75-100 de RE	Percentil 50-75 de RE	Percentil 25-50 de RE	Percentil 0-25 de RE	Neg para RE	Totales de filas
Riesgo bajo	101	89	82	41	0	313
Riesgo inter.	29	45	37	42	1	154
Riesgo alto	24	23	32	71	28	178
todos los grupos	154	157	151	154	29	645

**Terapia hormonal recomendada, considerar quimioterapia**

Grupo RS	Percentil 75-100 de RE	Percentil 50-75 de RE	Percentil 25-50 de RE	Percentil 0-25 de RE	Neg para RE	Totales de filas
Riesgo bajo	101	89	82	41	0	313
Riesgo inter.	29	45	37	42	1	154
Riesgo alto	24	23	32	71	28	178
todos los grupos	154	157	151	154	29	645

**Terapia hormonal recomendada + quimioterapia**

Grupo RS	Percentil 75-100 de RE	Percentil 50-75 de RE	Percentil 25-50 de RE	Percentil 0-25 de RE	Neg para RE	Totales de filas
Riesgo bajo	101	89	82	41	0	313
Riesgo inter.	29	45	37	42	1	154
Riesgo alto	24	23	32	71	28	178
Todos los grupos	154	157	151	154	29	645

Un experto en la técnica reconocerá numerosos métodos y materiales similares o equivalentes a los descritos en el presente documento, que podrían usarse en la práctica de la presente invención. En efecto, la presente invención no se limita en modo alguno a los métodos y materiales descritos. Aunque la presente invención se ha descrito con referencia a lo que se consideran que son las realizaciones específicas, debe entenderse que la invención no se limita a tales realizaciones. Por el contrario, la invención pretende cubrir diversas modificaciones incluidas dentro del alcance de las reivindicaciones adjuntas. Por ejemplo, aunque la descripción se ilustra identificando genes y grupos de genes útiles para predecir la respuesta beneficiosa de un paciente con cáncer de mama al tratamiento con TAM, métodos similares para determinar la respuesta del paciente al tratamiento con otros fármacos antiestrógenos están específicamente dentro del alcance en el presente documento.

10

15

Tabla 7

Reactivo	Gen	Registro	Oligo	Secuencia	Longitud
Directo	ACTB	NM_001101	S0034/B-acti.f2	CAGCAGATGTGGATCAGCAAG	21
Inverso	ACTB	NM_001101	S0036/B-acti.r2	GCATTTGCGGTGGACGAT	18
Sonda	ACTB	NM_001101	S4730/B-acti.p2	AGGAGTATGACGAGTCCGGCCCC	23
Directo	BAG1	NM_004323	S1386/BAG1.f2	CGTTGTCAGCACTTGGGAATAACA	23
Inverso	BAG1	NM_004323	S1387/BAG1.r2	GTTCAACCTCTTCCTGTGGACTGT	24
Sonda	BAG1	NM_004323	S4731/BAG1.p2	CCCAATTAACATGACCCGGCAACCAT	26
Directo	BCL2	NM_000633	S0043/Bcl2.f2	CAGATGGACCTAGTACCCACTGAGA	25
Inverso	BCL2	NM_000633	S0045/Bcl2.r2	CCTATGATTTAAGGGCATTTTTCC	24
Sonda	BCL2	NM_000633	S4732/Bcl2.p2	TTCCACGCCGAAGGACAGCGAT	22
Directo	CCNB1	NM_031966	S1720/CCNB1.f2	TTCAGGTTGTTGCAGGAGAC	20
Inverso	CCNB1	NM_031966	S1721/CCNB1.r2	CATCTTCTGGGCACACAAT	20
Sonda	CCNB1	NM_031966	S4733/CCNB1.p2	TGTCTCCATTATTGATCGGTTTCATGCA	27
Directo	CD68	NM_001251	S0067/CD68.f2	TGGTCCCAGCCCTGTGT	18
Inverso	CD68	NM_001251	S0069/CD68.r2	CTCCTCCACCCTGGGTTGT	19
Sonda	CD68	NM_001251	S4734/CD68.p2	CTCCAAGCCCAGATTGAGATTCGAGTCA	28
Directo	SCUBE2	NM_020974	S1494/SCUBE2.f2	TGACAATCAGCACACCTGCAT	21
Inverso	SCUBE2	NM_020974	S1495/SCUBE2.r2	TGTGACTACAGCCGTGATCCTTA	23
Sonda	SCUBE2	NM_020974	S4735/SCUBE2.p2	CAGGCCCTCTCCGAGCGGT	20
Directo	CTSL2	NM_001333	S4354/CTSL2.f1	TGTCTCACTGAGCGAGCAGAA	21
Inverso	CTSL2	NM_001333	S4355/CTSL2.r1	ACCATTGCAGCCCTGATTG	19
Sonda	CTSL2	NM_001333	S4356/CTSL2.p1	CTTGAGGACGCGAACAGTCCACCA	24
Directo	ESR1	NM_000125	S0115/EstR1.f1	CGTGGTGCCCTCTATGAC	19
Inverso	ESR1	NM_000125	S0117/EstR1.r1	GGCTAGTGGGCGCATGTAG	19
Sonda	ESR1	NM_000125	S4737/EstR1.p1	CTGGAGATGCTGGACGCC	19
Directo	GAPD	NM_002046	S0374/GAPD.f1	ATCCACCCATGGCAAATC	20
Inverso	GAPD	NM_002046	S0375/GAPD.r1	GATGGGATTTCCATTGATGACA	22
Sonda	GAPD	NM_002046	S4738/GAPD.p1	CCGTTCTCAGCCTTGACGGTGC	22
Directo	GRB7	NM_005310	S0130/GRB7.f2	CCATCTGCATCCATCTTGT	20
Inverso	GRB7	NM_005310	S0132/GRB7.r2	GGCCACCAGGTATTATCTG	20
Sonda	GRB7	NM_005310	S4726/GRB7.p2	CTCCCACCCTTGAGAAGTGCCT	23
Directo	GSTM1	NM_000561	S2026/GSTM1.r1	GGCCCAGCTTGAATTTTCA	20
Inverso	GSTM1	NM_000561	S2027/GSTM1.f1	AAGCTATGAGGAAAAGAAGTACACGAT	27
Sonda	GSTM1	NM_000561	S4739/GSTM1.p1	TCAGCCACTGGCTTCTGTGATAATCAGGAG	30
Directo	GUSB	NM_000181	S0139/GUS.f1	CCCCTCAGTAGCCAAGTCA	20
Inverso	GUSB	NM_000181	S0141/GUS.r1	CACGCAGGTGGTATCAGTCT	20
Sonda	GUSB	NM_000181	S4740/GUS.p1	TCAAGTAAACGGGCTGTTTTCCAAACA	27
Directo	ERBB2	NM_004448	S0142/HER2.f3	CGGTGTGAGAAGTGCAGCAA	20
Inverso	ERBB2	NM_004448	S0144/HER2.r3	CCTCTCGCAAGTGCTCCAT	19
Sonda	ERBB2	NM_004448	S4729/HER2.p3	CCAGACCATAGCACACTCGGGCAC	24
Directo	MKI67	NM_002417	S0436/MKI67.f2	CGGACTTTGGGTGCGACTT	19
Inverso	MKI67	NM_002417	S0437/MKI67.r2	TTACAACCTTCCACTGGGACGAT	24
Sonda	MKI67	NM_002417	S4741/MKI67.p2	CCACTTGTGCAACCACCGCTCGT	23
Directo	MYBL2	NM_002466	S3270/MYBL2.f1	GCCGAGATCGCCAAGATG	18
Inverso	MYBL2	NM_002466	S3271/MYBL2.r1	CTTTTGATGGTAGAGTCCAGTGATTC	27
Sonda	MYBL2	NM_002466	S4742/MYBL2.p1	CAGCATTGTCTGTCCTCCCTGGCA	24
Directo	PGR	NM_000926	S1336/PR.f6	GCATCAGGCTGTCATTATGG	20
Inverso	PGR	NM_000926	S1337/PR.r6	AGTAGTTGTGCTGCCCTTCC	20
Sonda	PGR	NM_000926	S4743/PR.p6	TGTCCTTACCTGTGGGAGCTGTAAGGTC	28
Directo	RPLP0	NM_001002	S0256/RPLP0.f2	CCATTCTATCATCAACGGGTACAA	24
Inverso	RPLP0	NM_001002	S0258/RPLP0.r2	TCAGCAAGTGGGAAGGTGTAATC	23

ES 2 384 107 T3

Sonda	RPLP0	NM_001002	S4744/RPLP0.p2	TCTCCACAGACAAGGCCAGGACTCG	25
Directo	STK6	NM_003600	S0794/STK6.f2	CATCTTCCAGGAGGACCACT	20
Inverso	STK6	NM_003600	S0795/STK6.r2	TCCGACCTTCAATCATTCA	20
Sonda	STK6	NM_003600	S4745/STK6.p2	CTCTGTGGCACCCCTGGACTACCTG	24
Directo	MMP11	NM_005940	S2067/MMP11.f3	CCTGGAGGCTGCAACATACC	20
Inverso	MMP11	NM_005940	S2068/MMP11.r3	TACAATGGCTTTGGAGGATAGCA	23
Sonda	MMP11	NM_005940	S4746/MMP11.p3	ATCCTCCTGAAGCCCTTTTCGCAGC	25
Directo	BIRC5	NM_001168	S0259/BIRC5.f2	TGTTTTGATTCCCAGGCTTA	20
Inverso	BIRC5	NM_001168	S0261/BIRC5.r2	CAAAGCTGTCAGCTTAGCAAAG	24
Sonda	BIRC5	NM_001168	S4747/BIRC5.p2	TGCCTTCTCCTCCCTCACTTCTCACCT	28
Directo	TFRC	NM_003234	S1352/TFRC.f3	GCCAACTGCTTTCATTTGTG	20
Inverso	TFRC	NM_003234	S1353/TFRC.r3	ACTCAGGCCCATTTCCTTTA	20
Sonda	TFRC	NM_003234	S4748/TFRC.p3	AGGGATCTGAACCAATACAGAGCAGACA	28

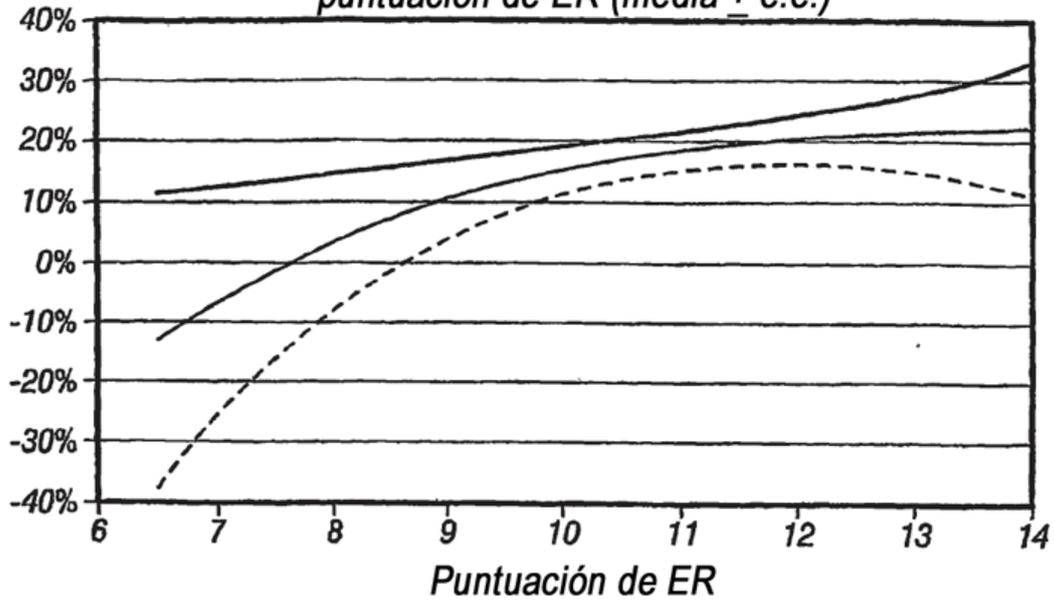
Tabla 8

Gen	LocusLink	Secuencia
ACTB	NM_001101	CAGCAGATGTGGATCAGCAAGAGAGATGACAGAGTCCGGCCCTCCATCGTCACCGCAATGC
BAG1	NM_004323	CGTTGTCAGCACITGGAAATACAAGATGTTCCCGGGTCAATGTTATTGGGMAAAGACAGTCCACAGGAAGAGTTGAAC
BCL2	NM_000633	CAGATGGACCTAGTACCCACTGAGATTTCACAGCCGAGGACAGCGATGGGMAAATCCCTTAAATCATAGG
CCNB1	NM_031966	TTCAGGTTGTTGGCAGGAGACCATGTACATGACTGTCTCCATTATTGATCGTTATCGCAGAAATATTTGTGTGCCCAAGAGATG
CD68	NM_001251	TGTTCCAGCCCTGTGTCCACCTCCAGCCAGATTCCAGATTCGAGTCTATACACAAACCAGGGTGGAGAG
SCUBE2	NM_020974	TGACAATCAGCACACCTGCATTCCCGCTCGAAGAGGGCCCTGAGTGCATGATGATAGGATACGGCTGTGTAGTCACA
CTSL2	NM_001333	TGTCCTCACTGAGCGAGCAGAACTCTGGTGGACTGTTCGGCTCCTCAAGGCCAATCAGGGCTGCAATGGT.
ESR1	NM_000125	CGTGGTCCCTCTATGACCTCTGCTGAGATGCTGGACGCCACCCCTTACATGCCCCACTAGCC
GAPD	NM_002046	ATTCCACCCATGGCAATTCATGGCACCGTCAAGGCTGAGAACGGGAAGCTTGTCAATGGAATCCCATC
GRB7	NM_005310	CCATGTGCATCGATCTGTTGGGCTCCCAACCCTTGAGAGTCCCTCAGATAATACCTGGTGGCC
GSTM1	NM_000961	AAGCTATGAGGMAAAGATCACAGATGGGGACCTCTGATATATACAGAAAGCCAGTGGCTGAATGMAAATTCAGCTGGGCC
GUS9	NM_000181	CCCACTCAGTAGCCCAAGTACAAATGTTGGAAACAGCCGTTTACTTGAAGAACTGATACCACCTGCCGTG
ERBB2	NM_004446	CGGTGTGAGAAAGTCCAGCAAGCCCTGTGCCCGAGTGTCTATGGTCTGGGCACTGGAGCACTTGGCAGAGG
MYBL2	NM_002417	CGGACTTTGGGTGGACTTGAAGAGCGGTGTTCCAGCAAGTGGCCCTTCCGGCCCGGATCCCGAGTGGAAAGATTGTAA
PGR	NM_000926	GCCGAGATCGCCAGATGTTCCAGGAGSACACACAAATGCTGTGAAGAACTACTGGAACTTACCATCAAAG
RPLP0	NM_001002	GCATCAGGCTGTCAATATGTTGCTTACCTGTGGAGCTGTAAGGTCTTTTMAAGGGGCAATGGAAGGGCAGCACACTACT
STK6	NM_003600	CCATTCTATCATAACGGGTACAAACGAGTCTGGCCCTTGTCTGTGGAGACGGATTACACCTTCCCCTTCTCTGA
MMP11	NM_005940	CATCTCCAGGAGACCACTCTCTGTGGCACCCCTGGACTACCTGCCCTCCCTGAAATGAAAGTGGGA
BIRC5	NM_001168	CCTGGAGGCTGCACATACACTCAATCCTGTCCAGCCGGATCCCTCCTGAAGCCCTTTTGGCAGCACTGCTATCCTCCAAAGCCATTGTA
TFRC	NM_003234	TGTTTTGATCCCGGGCTTACCAGGTGAGAAATGAGGGAGAGAGCCAGTGTCCCTTTTGTAGAGCTGACAGCTTTG
		GCCAACTGCTTTTCATTTGTGAGGSAATCTGAACCAATAACAAGCAGACATMAAGGAAATGGGCTGGAGT

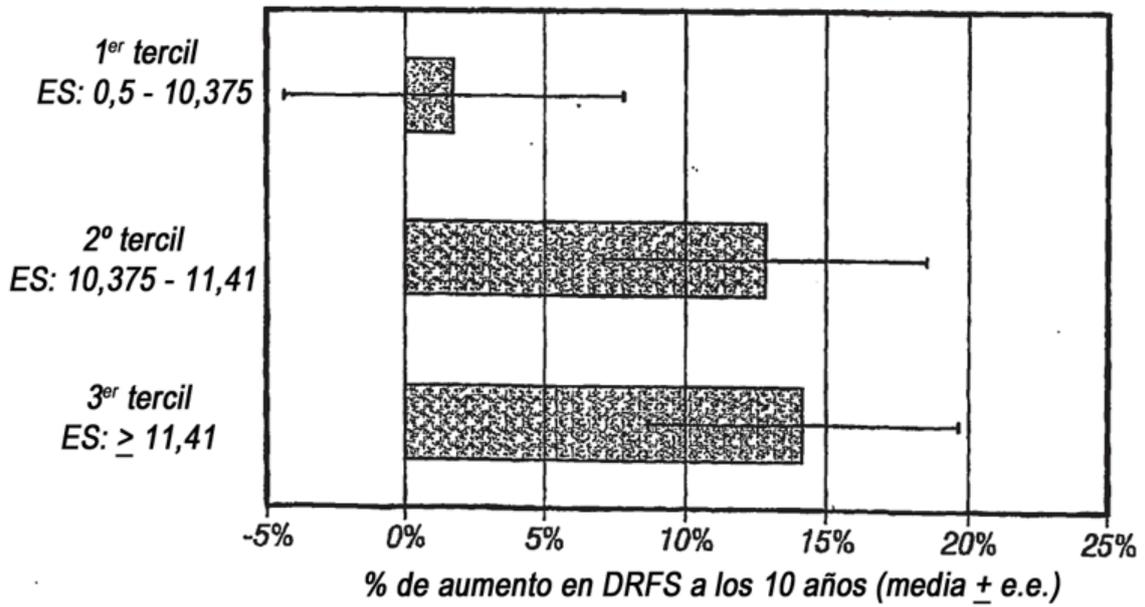
## REIVINDICACIONES

1. Un método para la determinación cuantitativa de la probabilidad de una respuesta beneficiosa en un paciente con cáncer de mama positivo para receptor de estrógenos 1 (ESR1) al tratamiento con un fármaco antiestrógeno, que comprende determinar cuantitativamente, en una muestra biológica que comprende células de cáncer de mama obtenidas de dicho paciente, una puntuación del grupo de ESR1 que se basa en los niveles de expresión de ESR1, PGR, BCL2 y SCUBE2, o productos de expresión de los mismos, en el que dicha puntuación del grupo de ESR1 se representa como una variable continua o como intervalos de expresión y en el que se usan aumentos en dichos niveles de expresión como una indicación cuantitativa de que dicho paciente tiene una probabilidad aumentada de una respuesta beneficiosa al tratamiento con un fármaco antiestrógeno
2. Un método según la reivindicación 1, en el que el fármaco antiestrógeno se selecciona del grupo que consiste en tamoxifeno, toremifeno, anastrozol y acetato de megasterol.
3. Un método según la reivindicación 1, en el que el fármaco antiestrógeno es un antagonista de la unión de estrógenos al receptor de estrógenos.
4. Un método según la reivindicación 2 ó 3, en el que el fármaco antiestrógeno es tamoxifeno.
5. Un método según la reivindicación 2, en el que el fármaco antiestrógeno es anastrozol.
6. Un método según la reivindicación 1, que comprende además la etapa de preparar un informe para dicho paciente, que incluye una recomendación acerca de si dicho paciente debe recibir el tratamiento con dicho fármaco antiestrógeno solo, quimioterapia sola o quimioterapia más fármaco antiestrógeno.
7. Un método según la reivindicación 6, en el que dicho tratamiento con antiestrógeno solo, quimioterapia sola o quimioterapia más antiestrógeno se recomienda consultando un gráfico o una tabla a lo largo de uno de cuyos ejes se presenta el nivel de expresión de ESR1, o bien como una variable continua o bien en intervalos de expresión de ESR1, y a lo largo de cuyo otro eje se presenta la probabilidad de recurrencia de cáncer, o bien como una variable continua o bien en intervalos de riesgo.
8. Un método según la reivindicación 7, en el que dichos intervalos de expresión de ESR1 son terciles dentro de una población de valores de ESR1 de pacientes con cáncer de mama, y se dividen en intervalos de ESR1 bajo, intermedio y alto.
9. Un método según la reivindicación 8, en el que dichos intervalos de riesgo son intervalos de riesgo bajo, intermedio y alto.
10. Un método según la reivindicación 9, en el que a dichos pacientes en los intervalos de ESR1 alto y de riesgo bajo se les recomienda el tratamiento con un fármaco antiestrógeno solo.
11. Un método según la reivindicación 9, en el que a dichos pacientes en los intervalos de ESR1 bajo y de riesgo alto se les recomienda el tratamiento con quimioterapia sola.
12. Un método según la reivindicación 9, en el que a dichos pacientes en los intervalos de ESR1 intermedio y de riesgo alto se les recomienda tanto el tratamiento con fármaco antiestrógeno como el tratamiento con quimioterapia.
13. Un método según la reivindicación 1, en el que dicha variable es el nivel de expresión del transcrito de ARN de ESR1, o su producto de expresión.
14. Un método según la reivindicación 1, que comprende además la etapa de determinar una puntuación de recurrencia para dicho paciente, que predice la probabilidad de recurrencia de cáncer en el sujeto.
15. Un método según la reivindicación 14, en el que el fármaco antiestrógeno se selecciona del grupo que consiste en tamoxifeno, toremifeno, anastrozol y acetato de megasterol.
16. Un método según la reivindicación 14, en el que el fármaco antiestrógeno es un antagonista de la unión de estrógenos al receptor de estrógenos.
17. Un método según la reivindicación 15 ó 16, en el que el fármaco antiestrógeno es tamoxifeno.
18. Un método según la reivindicación 15, en el que el fármaco antiestrógeno es anastrozol.
19. Un método según la reivindicación 14, que comprende además la etapa de preparar un informe para dicho paciente, que incluye una recomendación acerca de si dicho paciente debe recibir tratamiento con dicho fármaco antiestrógeno solo, quimioterapia sola o quimioterapia más fármaco antiestrógeno.

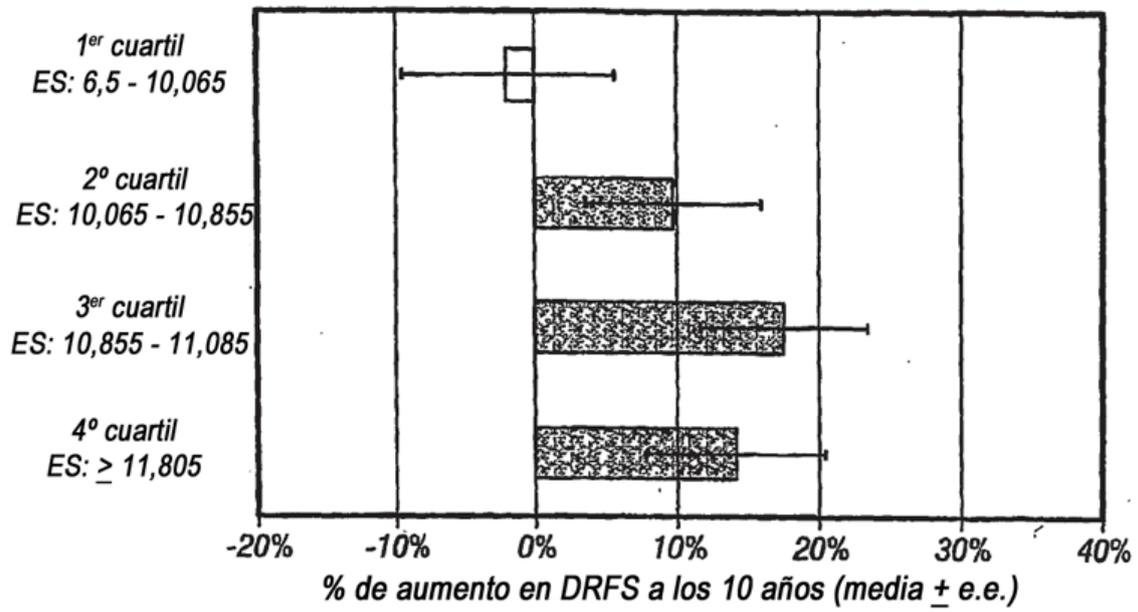
*Beneficio absoluto de tamoxifeno como función de la puntuación de ER (media  $\pm$  e.e.)*



**FIG. 1**



**FIG. 2**



**FIG. 3**