

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 651 849**

51 Int. Cl.:

G06F 19/00 (2011.01)

C12Q 1/68 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **30.06.2004 PCT/US2004/021163**

87 Fecha y número de publicación internacional: **27.01.2005 WO05008213**

96 Fecha de presentación y número de la solicitud europea: **30.06.2004 E 04777383 (3)**

97 Fecha y número de publicación de la concesión europea: **06.12.2017 EP 1644858**

54 Título: **Algoritmo del perfil de expresión y test para el pronóstico del cáncer**

30 Prioridad:

10.07.2003 US 486302 P
03.12.2003 US 526947 P

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
30.01.2018

73 Titular/es:

GENOMIC HEALTH, INC. (50.0%)
301 PENOBSCOT DRIVE
REDWOOD CITY, CA 94063, US y
NSABP FOUNDATION, INC. (50.0%)

72 Inventor/es:

BAKER, JOFFRE;
BRYANT, JOHN, L.;
PAIK, SOONMYUNG y
SHAK, STEVEN

74 Agente/Representante:

UNGRÍA LÓPEZ, Javier

ES 2 651 849 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

Algoritmo del perfil de expresión y test para el pronóstico del cáncer

Campo de la invención

5 La presente invención proporciona un test cuantitativo no invasivo para determinar el pronóstico de los pacientes con cáncer. El test se basa en mediciones de los niveles tumorales de determinados ARN mensajeros (ARNm) o los correspondientes productos de expresión génica. Estos niveles de ARNm o proteína se introducen en una fórmula polinómica (algoritmo) que produce una puntuación numérica, que indica el riesgo de recidiva (puntuación de recidiva) o la probabilidad de respuesta del paciente a la terapia (puntuación de respuesta).

10 Descripción de la técnica relacionada

15 Existe la necesidad de contar con tests clínicos que ayuden a los oncólogos a tomar las decisiones de tratamiento bien fundamentadas. Una de las decisiones más fundamentales a las que se enfrenta el oncólogo en su práctica diaria es si tratar o renunciar al tratamiento de un paciente concreto con agentes quimioterapéuticos. Por lo general, los agentes terapéuticos existentes para el cáncer tienen una eficacia moderada acompañada de una toxicidad sustancial. Por tanto, resulta altamente recomendable predeterminar qué pacientes (una vez reseca el tumor primario) es probable que padezcan una recidiva metastásica. Si existiera un método fiable para obtener esa información, se podría seleccionar a los pacientes de alto riesgo para la quimioterapia adyuvante y a los pacientes con pocas probabilidades de sufrir una recidiva del cáncer se les podría evitar una exposición innecesaria a los eventos adversos asociados a la quimioterapia. De forma similar, antes de someter a un paciente (antes o después de reseca un tumor primario) a un tratamiento concreto, sería recomendable saber si es probable que el paciente responda a este tratamiento. En el caso de pacientes que es poco probable que respondan a una determinada modalidad de tratamiento (por ejemplo, el tratamiento con determinados agentes quimioterapéuticos y/o radiología), se pueden diseñar y utilizar otros tratamientos, sin perder un valioso tiempo.

25 A pesar de que la bioquímica y la biología molecular modernas han revelado cientos de genes cuyas actividades influyen en la conducta y en el estado de diferenciación de las células tumorales, así como en su sensibilidad o resistencia a determinados fármacos terapéuticos, con algunas excepciones, el estado de estos genes no se ha explotado para tomar decisiones clínicas de forma rutinaria sobre los tratamientos con fármacos. Una excepción notable es el uso de la expresión de la proteína receptora de estrógenos (ER) en los carcinomas de mama para seleccionar a los pacientes aptos para el tratamiento con fármacos antiestrogénicos, como el tamoxifeno. Otro ejemplo excepcional es el uso de la expresión de proteína ErbB2 (Her2) en los carcinomas de mama para seleccionar a los pacientes aptos para el tratamiento con el anticuerpo anti-Her2, Herceptin® (Genentech, Inc., South San Francisco, CA).

35 La presente divulgación se refiere al cáncer, como el cáncer de mama, cuyo conocimiento biológico sigue siendo rudimentario. No cabe duda de que la clasificación del cáncer de mama en varios subgrupos, como el subgrupo ErbB2⁺, y subgrupos que se caracterizan por una expresión génica escasa o ausente del receptor de estrógenos (ER) y por algunos factores adicionales (Perou *et al.*, *Nature* 406:747-752 (2000)), no refleja la heterogeneidad celular y molecular del cáncer de mama y no permite la optimización del tratamiento y el cuidado del paciente. Lo mismo sucede con otros tipos de cáncer, que en muchos casos están menos estudiados y son más desconocidos que el cáncer de mama.

40 En estos momentos los tests de diagnóstico utilizados en la práctica clínica determinan un único analito y, por tanto, no aprovechan el valor potencial de conocer las relaciones existentes entre docenas de marcadores tumorales diferentes. Por otra parte, a menudo los test de diagnóstico no son cuantitativos y se basan en la inmunohistoquímica. Este método con frecuencia produce distintos resultados en laboratorios diferentes, en parte porque los reactivos no están estandarizados, y en parte porque las interpretaciones son subjetivas y no se pueden cuantificar fácilmente. Los test basados en ARN no se han utilizado con frecuencia debido al problema de la degradación del ARN con el paso del tiempo y al hecho de que resulta difícil obtener muestras de tejido fresco de los pacientes para su análisis. El tejido fijado y embebido en parafina se puede obtener más fácilmente y se han establecido métodos para detectar ARN en tejido fijado. Sin embargo, estos métodos típicamente no permiten el estudio de grandes cifras de genes obtenidas de pequeñas cantidades de material. Por tanto, tradicionalmente el tejido fijado rara vez se ha utilizado salvo para la detección inmunohistoquímica de proteínas.

45 En los últimos años, varios grupos han publicado estudios relativos a la clasificación de diversos tipos de cáncer mediante análisis de microarrays de expresión génica (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)). También se han documentado diversas clasificaciones del cáncer de mama humano basadas en patrones de expresión génica (Martin *et al.*, *Cancer Res.* 60:2232-2238 (2000); West *et al.*, *Proc. Natl. Acad. Sci. USA* 98:11462-11467 (2001)). La mayoría de estos estudios se centran en mejorar y perfeccionar la clasificación ya establecida de diversos tipos de cáncer, incluyendo el cáncer de mama. Algunos estudios identifican patrones de expresión génica que pueden ser pronósticos (Sorlie *et*

al., Proc. Natl. Acad. Sci. USA 98:10869-10874 (2001); Yan *et al., Cancer Res.* 61:8375-8380 (2001); Van De Vijver *et al. New England Journal of Medicine* 347: 1999-2009 (2002)), pero debido a las cifras insuficientes de pacientes estudiados, todavía no se encuentran lo bastante validados como para ser utilizados de forma generalizada en la práctica clínica.

5 Van't Veer *et al., Nature* 415:530-536 (2002) divulgan que los perfiles de expresión génica predicen el resultado clínico del cáncer de mama. Esteban *et al., Proc. Am. Soc. Clin. Oncol. Abstract* 3416 (2003) divulgan la expresión génica tumoral y el pronóstico en el cáncer de mama utilizando un ensayo multigénico RT-PCR de tejidos embebidos en parafina.

10 Cobleigh *et al., Proc. Am. Soc. Clin. Oncol* 22: Abstract 315 (2003) divulgan que la expresión génica tumoral predice la supervivencia libre de enfermedad a largo plazo en pacientes con cáncer de mama con 10 o más nódulos positivos utilizando un ensayo RT-PCR de alto rendimiento de tejidos tumorales embebidos en parafina.

15 Look *et al., J. Natl. Cancer Inst.* 94:116-128 (2002) divulgan el análisis combinado del impacto pronóstico del activador del plasminógeno tipo uroquinasa y su inhibidor PAI-1 en 8377 pacientes con cáncer de mama.

Paik *et al., New Engl. J. Med.* 351:2817-2826 (2004) divulgan un ensayo multigénico para predecir la recurrencia del beneficio clínico del tamoxifeno en un estudio de cáncer de mama positivo en receptor de estrógenos con nódulo negativo.

20 Paik *et al., J. Clin. Oncol.* 23:510 (2005) divulgan la expresión de 21 genes en un ensayo de Puntuación de Recidiva y el beneficio clínico del tamoxifeno en un estudio del cáncer de mama positivo en receptor de estrógenos con nódulo negativo.

Resumen de la invención

25 Esta invención proporciona un test basado en algoritmos para determinar la probabilidad de recidiva del cáncer de mama en pacientes positivos en receptores de estrógenos, negativos en nódulo linfático y tratados con tamoxifeno.

30 Las características del algoritmo que lo diferencian de otros métodos de pronóstico del cáncer de mama incluyen: 1) un set único de ARNm de ensayo (o los correspondientes productos de expresión génica) utilizado para determinar la probabilidad de recidiva; 2) determinadas ponderaciones utilizadas para combinar los datos de expresión en una fórmula; 3) umbrales utilizados para dividir a los pacientes en grupos con diferentes niveles de riesgo, como grupos de riesgo bajo, medio y alto. El algoritmo produce una puntuación de recidiva (RS) numérica.

35 El test requiere un ensayo de laboratorio para medir los niveles de los ARNm especificados o sus productos de expresión, pero puede utilizar cantidades muy pequeñas de muestras de tejido fresco o tejido congelado, o biopsia tumoral fijada y embebida en parafina que ya han sido tomadas necesariamente de pacientes y archivadas. Por tanto, el test puede ser no invasivo. También es compatible con varios métodos diferentes para recoger el tejido tumoral, por ejemplo, a través de una biopsia de base o la aspiración con aguja fina. El tejido tumoral puede ser diseccionado groseramente del tejido normal, aunque no necesariamente. Por otra parte, para cada miembro del conjunto de genes, la invención especifica secuencias de oligonucleótidos que se pueden utilizar en el test. Los niveles de ARNm están normalizados frente a un grupo especificado de genes de referencia. La invención abarca todas las subsecuencias del test y genes de referencia, incluyendo las secuencias 5' y 3' no traducidas y de intrones. Los niveles de proteínas se pueden normalizar de manera similar.

45 Por tanto, en un aspecto, la divulgación se refiere a un método de clasificación de un tumor en función de la probabilidad de recidiva del cáncer o respuesta al tratamiento en un sujeto mamífero que consiste en lo siguiente:

- a) someter una muestra biológica que contiene células cancerosas obtenidas de dicho sujeto para determinar el perfil de expresión génica o proteica;
- b) cuantificar el nivel de expresión génica o proteica de múltiples genes individuales, a fin de determinar un valor de expresión para cada gen;
- 50 c) crear subconjuntos de valores de expresión génica o proteica, de forma que cada subconjunto comprenda valores de expresión para genes o proteínas asociados por función biológica relacionada con el cáncer y/o por coexpresión;
- d) multiplicar el nivel de expresión de cada gen dentro de un subconjunto por un coeficiente que refleje su contribución relativa a la recidiva del cáncer o a la respuesta a la terapia dentro de dicho subconjunto y
- 55 añadir los productos de la multiplicación para obtener un término para dicho subconjunto;
- e) multiplicar el término de cada subconjunto por un factor que refleje su contribución a la recidiva del cáncer o la respuesta a la terapia; y
- f) producir la suma de términos para cada subconjunto multiplicado por dicho factor para obtener una puntuación de recidiva (RS) o una puntuación de respuesta a la terapia (RTS), donde la contribución de cada subconjunto que no muestra una correlación lineal con la recidiva del cáncer o la respuesta a la
- 60 terapia es incluida solo cuando supera un umbral predeterminado y donde a los subconjuntos en los que

ES 2 651 849 T3

una expresión aumentada de genes o proteínas incluidos reduce el riesgo de recidiva del cáncer se les asigna un valor negativo y a los subconjuntos en los que una expresión aumentada de genes o proteínas incluidos aumenta el riesgo de recidiva del cáncer se les asigna un valor positivo.

5 El algoritmo es adecuado para valorar el riesgo de recidiva del cáncer o respuesta a la terapia para cualquier tipo de cáncer, aunque se ejemplifica, sin limitación, para el cáncer de mama.

En el caso del cáncer de mama, los genes que contribuyen al algoritmo de recidiva incluyen 16 genes de ensayo y 5 genes de referencia.

Los genes/ARNm de ensayo especificados son los siguientes: Grb7; Her2; ER; PR; BC12; CEGP1; SURV; Ki-67; MYBL2; CCNB1; STK15; CTSL2; STMY3; GSTM1; BAG1; CD68.

10 Los genes/ARNm de referencia son: β -ACTIN; GAPDH; RPLPO; TFRC; GUS.

Para el cáncer de mama, el anterior conjunto de 16 genes de ensayo incluye 4 subconjuntos multigénicos.

15 Los genes de estos subconjuntos tienen en común tanto la expresión correlacionada como las categorías de función biológica: un subconjunto de proliferación comprende SURV, Ki-67, MYBL2, CCNB1 y STK15; un subconjunto del factor de crecimiento comprende Her2 y Grb7; un subconjunto del receptor de estrógenos comprende ER, PR, BC12, y CEGP1; un subconjunto de invasión comprende genes para las proteasas invasivas CTSL2 y STMY3.

Por tanto, en un aspecto particular, se proporciona un método para determinar la probabilidad de recidiva del cáncer de mama o de respuesta a la terapia hormonal en un sujeto mamífero que consiste en lo siguiente:

20 1. determinar los niveles de expresión de las transcripciones de ARN de GRB7, HER2, ER, PR, Bcl2, CEGP1, SURV, Ki.67, MYBL2, CCNB1, STK15, CTSL2, y STMY3, o sus productos de expresión, o los correspondientes genes sustitutos o sus productos de expresión, en una muestra biológica que contiene células tumorales obtenidas de dicho sujeto;

crear los siguientes subconjuntos de genes:

- 25 i) subconjunto del factor de crecimiento: GRB7 y HER2;
ii) subconjunto de diferenciación (receptor de estrógenos): ER, PR, Bcl2, y CEGP1;
iii) subconjunto de proliferación: SURV, Ki.67, MYBL2, CCNB1, y STK15; y
iv) subconjunto de invasión: CTSL2, y STMY3;

30 donde un gen de cualquiera de los subconjuntos (i)-(iv) puede ser sustituido por un gen sustituto que co-expresa conjuntamente con el mencionado gen en dicho tumor con un coeficiente de correlación de Pearson de $\geq 0,40$; y

c) calcular la puntuación de recidiva (RS) o la puntuación de respuesta a la terapia (RTS) para dicho sujeto ponderando las contribuciones de cada uno de los subconjuntos (i)-(iv) a la recidiva del cáncer de mama o a la respuesta a la terapia.

35 En una realización particular, el método comprende asimismo la determinación de los niveles de expresión de las transcripciones de ARN de CD68, GSTM1 y BAG1 o sus productos de expresión, o los correspondientes genes sustitutos o sus productos de expresión, e incluyendo la contribución de dichos genes o genes sustitutos a la recidiva del cáncer de mama o a la respuesta a la terapia para el cálculo de la RS o RTS, donde una RS o RTS más elevada representa una mayor probabilidad de recidiva del
40 cáncer de mama o una menor probabilidad de respuesta a la terapia, según corresponda.

En un aspecto específico, la invención se refiere a un método para determinar la probabilidad de recidiva de cáncer de mama en un sujeto mamífero que consiste en lo siguiente:

- 45 a) determinar los niveles de expresión de GRB7, HER2, EstR1, PR, Bcl2, CEGP1, SURV, Ki.67, MYBL2, CCNB1, STK15, CTSL2, STMY3, CD68, GSTM1, y BAG1, o sus productos de expresión, en una muestra biológica que contiene células tumorales obtenidas de dicho sujeto; y
b) calcular la puntuación de recidiva (RS) a través de la ecuación siguiente:

$$RS = (0,23 \text{ a } 0,70) \times \text{umbralejeGRB7} - (0,17 \text{ a } 0,51) \times \text{ejeER} + (0,53 \text{ a } 1,56) \times \text{umbralejeprolif} + (0,07 \text{ a } 0,21) \times \text{ejeinvas} + (0,03 \text{ a } 0,15) \times \text{CD68} - (0,04 \text{ a } 0,25) \times \text{GSTM1} - (0,05 \text{ a } 0,22) \times \text{BAG1}$$

donde

- 50 (a) $\text{eje GRB7} = (0,45 \text{ a } 1,35) \times \text{GRB7} + (0,05 \text{ a } 0,15) \times \text{HER2}$
(b) si $\text{eje GRB7} < -2$, entonces $\text{umbral eje GRB7} = -2$, y si $\text{eje GRB7} \geq -2$, entonces $\text{umbral eje GRB7} = \text{eje GRB7}$;
(c) $\text{eje ER} = (\text{Est1} + \text{PR} + \text{Bcl2} + \text{CEGP1})/4$;
(d) $\text{ejeprolif} = (\text{SURV} + \text{Ki.67} + \text{MYBL2} + \text{CCNB1} + \text{STK15})/5$;
55 (e) si $\text{ejeprolif} < -3,5$, entonces $\text{umbralejeprolif} = -3,5$,
si $\text{ejeprolif} \geq -3,5$, entonces $\text{umbralejeprolif} = \text{ejeprolif}$; y

vi) $e_{\text{ejinvas}} = (\text{CTSL2} + \text{STMY3})/2$,

donde los términos para todos los genes individuales para los que no se indican rangos de forma específica pueden variar entre 0,5 y 1,5, y donde una RS más elevada representa una mayor probabilidad de recidiva del cáncer de mama.

5 El ensayo de laboratorio para determinar la expresión génica puede utilizar una variedad de plataformas diferentes. La realización preferible para el uso con tejidos fijados y embebidos en parafina es la reacción en cadena de polimerasa transcriptasa inversa (qRT-PCR). Sin embargo, también se pueden utilizar otras plataformas tecnológicas, incluyendo la espectrometría de masas y microarrays de ADN. En el caso de los
10 microarrays de ADN, las sondas de polinucleótidos pueden ser ADNc («arrays de ADNc») que tienen típicamente entre 500 y 5000 bases de longitud, aunque también se pueden utilizar ADNc más cortos o más largos y se consideran dentro del alcance de la presente invención. Alternativamente, los polinucleótidos pueden ser oligonucleótidos (microarrays de ADN), que tienen típicamente entre 20 y 80 bases de longitud, aunque también resultan adecuados oligonucleótidos más cortos o más largos y se consideran dentro del alcance de la invención. La superficie sólida puede ser, por ejemplo, vidrio o nylon,
15 o cualquier otra superficie sólida utilizada típicamente para preparar arrays, tales como microarrays, y es típicamente vidrio.

Breve descripción de los gráficos

La Tabla 1 recoge los genes de ensayo y de referencia, cuya expresión se utiliza para elaborar el algoritmo de recidiva del cáncer de mama. La tabla incluye números de acceso para los genes,
20 secuencias para los cebadores directos e inversos (designados «f» y «r», respectivamente) y sondas (designadas «p») utilizados para la amplificación PCR.

La Tabla 2 muestra las secuencias de amplicones de los mencionados genes.

Las Figuras 1A y B presentan resultados de un estudio clínico de 242 pacientes con cáncer de mama invasivo y negativo en nódulos linfáticos, para los que se disponía de datos de recidiva de 10 años. La
25 Figura 1A presenta la proporción de pacientes con recidiva real del cáncer (método Kaplan-Meier) frente a las puntuaciones de recidiva calculadas. Los pacientes están agrupados en 5 unidades de percentiles, salvo por los pacientes con puntuaciones <-4,5, que se han agrupado juntos.

La Figura 1B es similar a la Figura 1A. También presenta la proporción de pacientes con recidiva real del
30 cáncer, pero en este caso los datos de recidiva real están comparados gráficamente con los percentiles de las puntuaciones de recidiva. Los pacientes están agrupados en 5 unidades de percentiles, salvo por los pacientes del percentil 20 más bajo, que se han agrupado juntos.

La Figura 2 ilustra la relación entre la puntuación de recidiva (RS) y la tasa o probabilidad de recidiva
35 distante del cáncer de mama en 10 años. Los datos se toman de la población de pacientes NSABP B-14. Las líneas de puntos representan intervalos de confianza del 95%. Cada marca en el eje X denota un paciente que tiene la correspondiente RS.

Descripción detallada de la realización preferible

A. Definiciones

Salvo que se defina lo contrario, los términos técnicos y científicos usados en el presente tienen el
40 significado que entenderá una persona con los conocimientos habituales de la técnica a la que se refiere la presente invención. Singleton *et al.*, *Dictionary of Microbiology and Molecular Biology* 2ª ed., J. Wiley & Sons (New York, NY 1994), y March, *Advanced Organic Chemistry Reactions, Mechanisms and Structure* 4ª ed., John Wiley & Sons (New York, NY 1992), proporcionan a un experto en la técnica una guía general de muchos de los términos utilizados en la presente solicitud.

Un experto en la técnica reconocerá múltiples métodos y materiales similares o equivalentes a los
45 descritos en el presente que se podrían utilizar en la práctica de la presente invención. En efecto, la presente invención no se limita en modo alguno a los métodos y materiales descritos. A efectos de la presente invención, los siguientes términos se describen a continuación.

El término «microarray» se refiere a una disposición ordenada de elementos del array hibridables,
preferiblemente sondas de polinucleótidos, en un sustrato.

50 El término «polinucleótido», utilizado en singular o en plural, se refiere generalmente a cualquier polirribonucleótido o polidesoxirribonucleótido, que puede ser ARN o ADN sin modificar o ARN o ADN modificado. Así, por ejemplo, los polinucleótidos definidos en el presente incluyen, entre otros, ADN de
55 cadena simple y doble, ADN que incluye regiones de cadena simple y doble, ARN de cadena simple y doble, y ARN que incluye regiones de cadena simple y doble, moléculas híbridas que comprenden ADN y ARN que pueden ser de cadena simple o, más típicamente, de cadena doble o incluir regiones de cadena simple y de cadena doble. Por otra parte, el término «polinucleótido» a efectos del presente se refiere a regiones de cadena triple que comprenden ARN o ADN o tanto ARN como ADN. Las cadenas de estas regiones pueden ser de la misma molécula o de moléculas diferentes. Las regiones pueden incluir la totalidad de una o más de las moléculas, aunque más típicamente incluyen solamente una región de

algunas de las moléculas. Una de las moléculas de una región de triple hélice suele ser un oligonucleótido. El término «polinucleótido» incluye específicamente ADNc. El término incluye ADN (incluyendo ADNc) y ARN que contienen una o más bases modificadas. De este modo, los ADN o ARN con los esqueletos modificados por motivos de estabilidad o por otras razones son «polinucleótidos» tal y como se entiende ese término en el presente. Por otra parte, los ADN o ARN que comprenden bases inusuales, como inosina, o bases modificadas, como bases tritiadas, se incluyen en el término «polinucleótidos» a efecto del presente. En general, el término «polinucleótido» abarca todas las formas químicas, enzimáticas y/o metabólicamente modificadas de polinucleótidos no modificados, así como las formas químicas de ADN y ARN características de los virus y las células, incluyendo células simples y complejas.

El término «oligonucleótido» se refiere a un polinucleótido relativamente corto, incluyendo, entre otros, desoxirribonucleótidos de cadena simple, ribonucleótidos de cadena simple o doble, híbridos de ARN:ADN y ADN de cadena doble. Los oligonucleótidos, como oligonucleótidos de una sonda de ADN de cadena simple, se suelen sintetizar por métodos químicos, por ejemplo utilizando sintetizadores de oligonucleótidos automáticos que se pueden adquirir en el mercado. Sin embargo, los oligonucleótidos se pueden producir mediante muchos otros métodos, incluyendo técnicas mediadas por ADN recombinante *in vitro* y mediante la expresión de ADN en células y organismos.

Los términos «gen expresado diferencialmente», «expresión diferencial del gen» y sus sinónimos, que se utilizan de manera intercambiable, se refieren a un gen cuya expresión está activada a un nivel superior o inferior en un sujeto que padece una enfermedad, concretamente cáncer, como cáncer de mama, en comparación con su expresión en un sujeto normal o de control. Los términos también incluyen genes cuya expresión está activada hasta un nivel superior o inferior en diferentes etapas de la misma enfermedad. Se entiende asimismo que un gen expresado diferencialmente puede estar activado o inhibido a nivel de ácido nucleico o a nivel de proteína, o puede ser sometido a una unión alternativa para obtener un producto polipeptídico diferente. Estas diferencias se pueden poner de manifiesto a través de una variación de los niveles de ARNm, la expresión superficial, la secreción u otra división de un polipéptido, por ejemplo. La expresión diferencial del gen puede incluir una comparación de la expresión entre dos o más genes o sus productos génicos, o una comparación de los ratios de expresión entre dos o más genes o sus productos génicos, o incluso una comparación de dos productos del mismo gen procesados de forma diferente, que difieren entre sujetos normales y sujetos que padecen una enfermedad, concretamente cáncer, o entre diferentes etapas de la misma enfermedad. La expresión diferencial incluye diferencias tanto cuantitativas como cualitativas en el patrón de expresión temporal o celular de un gen o sus productos de expresión entre, por ejemplo, células normales y enfermas, o entre células que se han sometido a diferentes eventos patológicos o en diferentes etapas de una enfermedad. A efectos de la presente invención, la «expresión diferencial de un gen» se considera presente cuando se observa una diferencia de al menos dos veces, preferiblemente al menos cuatro veces, más preferiblemente al menos seis veces, y más preferiblemente al menos diez veces entre la expresión de un gen dado en sujetos normales y enfermos, o en diferentes etapas de la evolución de la enfermedad en un sujeto enfermo.

La expresión «amplificación génica» se refiere a un proceso por el que múltiples copias de un gen o fragmento de un gen se forman en una célula o línea celular concreta. La región duplicada (un tramo de ADN amplificado) se suele denominar «amplicón». Por lo general, la cantidad de ARN mensajero (ARNm) producida, es decir el nivel de expresión génica, también aumenta en proporción al número de copias producidas del gen expresado en cuestión.

El término «pronóstico» se utiliza en el presente para referirse a la predicción de la probabilidad de muerte atribuible al cáncer o de progresión, incluyendo recidiva, propagación metastásica y resistencia a los fármacos, de una enfermedad neoplásica, como el cáncer de mama.

El término «predicción» se utiliza a efectos del presente para referirse a la probabilidad de que un paciente responda favorable o desfavorablemente a un fármaco o conjunto de fármacos, y también al grado de esas respuestas, o de que un paciente sobreviva tras la extirpación quirúrgica del tumor primario y/o de quimioterapia durante un determinado periodo de tiempo sin recidiva del cáncer. Los métodos predictivos de la presente invención resultan herramientas valiosas para predecir si es probable que un paciente responda favorablemente a un régimen de tratamiento, como una intervención quirúrgica, quimioterapia con un determinado fármaco o combinación de fármacos, y/o terapia de radiación, o si la supervivencia a largo plazo del paciente, tras la cirugía y/o terminación de la quimioterapia u otras modalidades de tratamiento, resulta probable.

El término supervivencia «a largo plazo» se utiliza en el presente para referirse a una supervivencia de al menos cinco años, más preferiblemente al menos ocho años, y más preferiblemente al menos 10 años después de la cirugía u otro tratamiento.

El término «tumor» a efectos del presente, se refiere a todo crecimiento celular neoplásico y proliferación, sea maligno o benigno, y a todas las células y tejidos precancerosos y cancerosos.

Los términos «cáncer» y «canceroso» se refieren o describen la condición fisiológica de mamíferos que se caracteriza típicamente por un crecimiento celular anormal.

Entre los ejemplos de cáncer se incluyen el cáncer de mama, cáncer de colon, cáncer de pulmón, cáncer de próstata, cáncer hepatocelular, cáncer gástrico, cáncer pancreático, cáncer cervical, cáncer de ovario, cáncer de hígado, cáncer de vejiga, cáncer del tracto urinario, cáncer de tiroides, cáncer renal, carcinoma, melanoma y cáncer cerebral.

5 La «patología» del cáncer (tumor) incluye todos los fenómenos que ponen en peligro el bienestar del paciente. Estos incluyen, entre otros, crecimiento celular anómalo o incontrolable, metástasis, interferencia en el funcionamiento normal de células vecinas, liberación de citoquinas u otros productos de secreción a niveles anómalos, supresión o empeoramiento de respuesta inflamatoria o inmunológica, neoplasia, premalignidad, malignidad, invasión de órganos o tejidos adyacentes o distantes, tales como los nódulos linfáticos, etc.

10 La «rigurosidad» de las reacciones de hibridación puede ser determinada fácilmente por un experto en la técnica y generalmente es un cálculo empírico dependiente de la longitud de la sonda, la temperatura de lavado y la concentración de sal. En general, las sondas más largas requieren temperaturas más elevadas para la hibridación, mientras que las sondas más cortas necesitan temperaturas inferiores. La hibridación depende en general de la capacidad del ADN desnaturalizado para rehibridarse cuando hay cadenas complementarias presentes en un entorno por debajo de su temperatura de fusión. Cuanto más alto es el grado de homología deseado entre la sonda y la secuencia hibridable, más alta es la temperatura relativa que se puede utilizar. Como resultado, sucede que las temperaturas relativas más elevadas tenderían a crear unas condiciones de reacción más rigurosas y las bajas menos. Para obtener más información y una explicación de la rigurosidad de las reacciones de hibridación, véase Ausubel et al, Current Protocols in Molecular Biology, Wiley Interscience Publishers, (1995).

15 Las «condiciones rigurosas» o «condiciones de alta rigurosidad», a efectos del presente, típicamente: 1) emplean baja fuerza iónica y alta temperatura para lavar, por ejemplo, cloruro sódico/0,0015 M citrato sódico/0,1% dodecil sulfato de sodio a 50°C; 2) emplean durante la hibridación un agente desnaturalizante, como formamida, por ejemplo, 50% (v/v) formamida con 0,1 % albúmina de suero bovino/0,1% ficol/0,1% polivinilpirrolidona/50mM tampón de fosfato sódico a pH 6.5 con 750 mM cloruro sódico, 75 mM citrato sódico a 42°C; o 3) emplean 50% formamida, 5 x SSC (0,75 M NaCl, 0,075 M citrato sódico), 50 mM fosfato sódico (pH 6.8), 0,1% pirofosfato sódico, 5 x solución Denhard, ADN de esperma de salmón sometido a ultrasonidos DNA (50 µg/ml), 0,1 % SDS, y 10% sulfato de dextrano a 42°C, con lavados a 42°C en 0,2 x SSC (cloruro sódico/citrato sódico) y 50% formamida a 55°C, seguidos de un lavado de alta rigurosidad consistente en 0,1 x SSC que contiene EDTA a 55°C.

20 Las «condiciones moderadamente rigurosas» se pueden identificar tal y como describen Sambrook et al., *Molecular Cloning: A Laboratory Manual*, New York: Cold Spring Harbor Press, 1989, e incluyen el uso de solución de lavado y condiciones de hibridación (por ejemplo, temperatura, fuerza iónica y % SDS) menos rigurosas que las anteriormente descritas. Un ejemplo de condiciones moderadamente rigurosas es la incubación durante una noche a 37°C en una solución que comprende: 20% formamida, 5 x SSC (150 mM NaCl, 15 mM citrato trisódico), 50 mM fosfato sódico (pH 7.6), 5 x solución Denhard, 10% sulfato de dextrano, y 20 mg/ml de ADN de esperma de salmón fragmentado desnaturalizado, seguida del lavado de los filtros en 1 x SSC a unos 37-50°C. El experto en la técnica reconocerá cómo ajustar la temperatura, la fuerza iónica, etc., necesarias para adaptarse a factores como la longitud de la sonda y similares.

35 En el contexto de la presente invención, las referencias a «al menos uno», «al menos dos», «al menos cinco», etc. de los genes enumerados en cualquier conjunto de genes concreto significa una y cualquiera de las combinaciones de los genes enumerados o todas ellas.

40 El término cáncer «nódulo (linfático) negativo», como cáncer de mama «nódulo (linfático) negativo», se utiliza a efectos del presente para referirse a un cáncer que no se ha propagado a los nódulos linfáticos.

El término «determinación del perfil de expresión génica» se utiliza en el sentido más amplio e incluye métodos de cuantificación de los niveles de ARNm o proteína en una muestra biológica.

50 El término «terapia adyuvante» se emplea generalmente para referirse al tratamiento que se administra adicionalmente a un tratamiento primario (inicial). En el tratamiento del cáncer, el término «terapia adyuvante» se utiliza para referirse a quimioterapia, terapia hormonal y/o terapia de radiación tras la extirpación quirúrgica del tumor, con el objetivo principal de reducir el riesgo de recidiva del cáncer.

55 La «terapia neoadyuvante» es la terapia accesoria o adyuvante que se administra antes de la terapia primaria (principal). La terapia neoadyuvante incluye, por ejemplo, quimioterapia, terapia de radiación y terapia hormonal. Por tanto, la quimioterapia se puede administrar antes de la cirugía para reducir el tumor, al objeto de que la cirugía pueda ser más efectiva, o, en el caso de los tumores previamente inoperables, de que pueda resultar posible.

60 El término «función biológica relacionada con el cáncer» se utiliza en el presente para referirse a una actividad molecular que incide en el éxito del cáncer frente al huésped, incluyendo, entre otras, actividades que regulan la proliferación celular, muerte celular programada (apoptosis), diferenciación, invasión, metástasis, supresión tumoral, susceptibilidad a la vigilancia inmunológica, angiogénesis, mantenimiento o adquisición de inmortalidad.

Descripción detallada

La práctica de la presente divulgación empleará, a menos que se indique lo contrario, técnicas convencionales de biología molecular (incluyendo técnicas recombinantes), microbiología, biología celular y bioquímica, que son conocidas para un experto en la técnica. Estas técnicas se explican íntegramente en la bibliografía, tales como: *Molecular Cloning: A Laboratory Manual*, 2ª edición (Sambrook et al., 1989); *Oligonucleotide Synthesis* (M.J. Gait, ed., 1984); *Animal Cell Culture* (R.I. Freshney, ed., 1987); *Methods in Enzymology* (Academic Press, Inc.); *Handbook of Experimental Immunology*, 4ª edición (D.M. Weir & C.C. 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).

La presente divulgación proporciona un algoritmo para determinar la probabilidad de recidiva del cáncer o la respuesta a la terapia en pacientes con cáncer. El método se basa en 1) la identificación y agrupación de un conjunto de genes que pueden servir como marcadores de la recidiva del cáncer o la probabilidad de respuesta del paciente a una terapia concreta; 2) la asignación de determinadas ponderaciones a los grupos y los genes individuales que integran los grupos que reflejan su valor a la hora de predecir la recidiva del cáncer o la respuesta a la terapia que se utilizan para combinar los datos de expresión en una fórmula; y 3) la determinación de valores umbral utilizados para dividir a los pacientes en grupos con diversos grados de riesgo de recidiva del cáncer o con una probabilidad variable de respuesta al tratamiento, tales como grupos de riesgo bajo, medio y alto o grupos en los que la probabilidad de respuesta del paciente a un tratamiento concreto es baja, media o alta. El algoritmo produce una puntuación de recidiva (RS) o una puntuación de respuesta al tratamiento (RTS) numérica que se puede utilizar para tomar decisiones de tratamiento por lo que respecta a la terapia de los pacientes con cáncer.

El primer paso para generar los datos que se van a analizar a través del algoritmo de la presente invención consiste en determinar el perfil de expresión génico o proteico.

25 1. Técnicas para determinar el perfil de expresión

La presente divulgación requiere un ensayo para medir los niveles de genes especificados (ARNm) o sus productos de expresión (proteínas) en una muestra biológica que contiene células cancerosas. Típicamente, la muestra biológica es una muestra de tejido fresca o guardada obtenida de un tumor, por ejemplo mediante biopsia o aspiración, aunque también se pueden utilizar para el análisis de fluidos biológicos que contienen células cancerosas.

En su forma más común, la determinación del perfil de expresión génica implica la determinación de los niveles de ARNm en una muestra de tejido, como una muestra de biopsia tumoral embebida en parafina y fijada que ya ha sido tomada de un paciente y archivada. Por tanto, el test puede ser totalmente no invasivo. También es compatible con varios métodos diferentes para recoger el tejido tumoral, por ejemplo, a través de una biopsia central o una aspiración con aguja fina. El tejido tumoral puede ser diseccionado por completo del tejido normal, aunque no necesariamente.

Los métodos para determinar el perfil de expresión génica destinados a medir los niveles de ARNm se pueden dividir en dos grandes grupos: los métodos basados en el análisis de la hibridación de polinucleótidos y los métodos basados en la secuenciación de polinucleótidos. Los métodos más frecuentemente utilizados conocidos en la técnica para la cuantificación de la expresión de ARNm incluyen Northern blot y la hibridación *in situ* (Parker & Barnes, *Methods in Molecular Biology* 106:247-283 (1999)); ensayos de protección de ARNasa (Hod, *Biotechniques* 13:852-854 (1992)); y reacción de la cadena de polimerasa de transcripción inversa (RT-PCR) (Weis et al., *Trends in Genetics* 8:263-264 (1992)). Alternativamente, se han empleado anticuerpos que pueden reconocer dúplex específicos, incluyendo dúplex ADN, dúplex ARN y dúplex ADN-ARN híbrido o dúplex ADN-proteína. Los métodos representativos para el análisis de la expresión génica basados en la secuenciación incluyen el análisis en serie de la expresión génica (SAGE) y el análisis de expresión génica mediante secuenciación de firmas masivamente en paralelo (MPSS).

En todas estas técnicas, el primer paso consiste en aislar el ARNm de una muestra diana. El material de partida es típicamente ARN completo aislado de líneas celulares tumorales o tumores humanos y correspondientes a líneas celulares o tejidos normales, respectivamente. Por tanto, el ARN puede ser aislado de diversos tumores primarios, incluyendo tumores o líneas de células tumorales de mama, pulmón, colon, próstata, cerebro, hígado, riñón, páncreas, bazo, timo, testículo, ovario, útero, etc., con ADN recogido de donantes sanos. Si la fuente del ARNm es un tumor primario, el ARNm se puede extraer, por ejemplo, de muestras de tejido fijadas (por ejemplo, fijadas con formalina) y embebidas en parafina, congeladas o guardadas.

Los métodos generales para la extracción de ARNm son conocidos en la técnica y se divulgan en libros de texto estándar 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 divulgan, por ejemplo, en Rupp and Locker, *Lab Invest.* 56:A67 (1987), y De Andres et al., *BioTechniques* 18:42044 (1995). En concreto, el aislamiento de ARN se puede realizar utilizando un kit de purificación, una solución tampón y proteasa, comercializados por fabricantes como Qiagen, de conformidad con las

instrucciones del fabricante. Por ejemplo, el ARN completo de las células de un cultivo se puede aislar utilizando mini columnas Qiagen RNeasy. Otros kits de aislamiento de ARN disponibles en el mercado incluyen MasterPure™ Complete DNA and RNA Purification Kit (EPICENTRE®, Madison, WI), y Paraffin Block RNA Isolation Kit (Ambion, Inc.). El ARN completo de las muestras de tejido se puede aislar utilizando RNA Stat-60 (Tel-Test). El ARN preparado a partir de un tumor se puede aislar, por ejemplo, por centrifugación en gradiente de densidad de cloruro de cesio.

Aunque la práctica de la divulgación se ilustrará por referencia a las técnicas desarrolladas para determinar los niveles de ARNm en una muestra biológica (por ejemplo, de tejido), otras técnicas, como los métodos de análisis proteómico, también están incluidos en el amplio concepto de determinación del perfil génico y se encuentran dentro del ámbito de aplicación de la presente divulgación. En general, un método preferible para determinar el perfil de la expresión génica para su uso con tejido embebido en parafina es la reacción en cadena de la polimerasa con transcriptasa inversa cuantitativa (qRT-PCR); sin embargo, también se pueden utilizar otras plataformas tecnológicas, incluyendo la espectrometría de masas y microarrays de ADN.

En las siguientes secciones se debatirán más detalladamente varias técnicas de determinación del perfil de expresión génica que resultan representativas, aunque no exhaustivas.

2.PCR transcriptasa inversa (RT-PCR)

De las técnicas anteriormente enumeradas, el método cuantitativo más sensible y flexible es la qRT-PCR, que se puede utilizar para comparar niveles de ARNm en diferentes poblaciones de muestras, en tejidos normales y tumorales, con o sin tratamiento con fármacos, con el fin de caracterizar patrones de expresión génica, para diferenciar entre ARNm estrechamente relacionados y para analizar la estructura de ARN.

Dado que el ARN no puede servir de plantilla para la PCR, el primer paso para la determinación del perfil génico mediante RT-PCR es la transcripción inversa de la plantilla de ARN en ADNc, seguida de su amplificación exponencial en una reacción PCR. Las dos transcriptasas inversas utilizadas más habitualmente son la transcriptasa inversa del virus de la mieloblastosis aviar (AMV-RT) y la transcriptasa inversa del virus de la leucemia murina de Moloney (MMLV-RT). El paso de la transcripción inversa se inicia típicamente utilizando cebadores específicos, hexámeros aleatorios o cebadores oligo-dT, dependiendo de las circunstancias y del objetivo de la determinación del perfil de expresión. Por ejemplo, el ARN extraído se puede someter a transcripción inversa utilizando un kit GeneAmp RNA PCR (Perkin Elmer, CA, EE. UU.), siguiendo las instrucciones del fabricante. El ADNc obtenido se puede utilizar entonces como plantilla en la posterior reacción PCR.

A pesar de que el paso de la PCR puede utilizar diversas polimerasas de ADN dependientes de ADN termoestables, típicamente emplea ADN polimerasa Taq, que tiene una actividad nucleasa 5'-3' pero carece de una actividad endonucleasa 3'-5' de corrección de lectura. Por tanto, la PCR TaqMan® típicamente utiliza la actividad 5'-nucleasa de la polimerasa Taq o Tth para hibridar una sonda de hibridación unida a su amplicón diana, aunque se puede utilizar cualquier enzima con actividad nucleasa 5' equivalente. Se utilizan dos cebadores oligonucleótidos para generar un amplicón típico de una reacción PCR. Un tercer oligonucleótido, o sonda, está diseñado para detectar la secuencia de nucleótidos que se encuentra entre los dos cebadores de la PCR. La sonda no es extensible por la enzima ADN polimerasa Taq y es marcada con un colorante fluorescente reporter y un colorante fluorescente inhibidor. Cualquier emisión inducida por láser del colorante reporter es inhibida por el colorante inhibidor cuando los dos colorantes se encuentran juntos en la sonda. Durante la reacción de amplificación, la enzima ADN polimerasa Taq cliva la sonda de forma dependiente de la plantilla. Los fragmentos de la sonda resultantes se disocian en solución y la señal del colorante reporter liberado se libera del efecto inhibidor del segundo fluoróforo. Una molécula del colorante reporter se libera de cada nueva molécula sintetizada y la detección del colorante reporter no inhibido proporciona la base para la interpretación cuantitativa de los datos.

La RT-PCR TaqMan® se puede realizar utilizando equipos disponibles en el mercado, 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 preferible, el procedimiento de la nucleasa 5' se ejecuta en un dispositivo de PCR cuantitativa en tiempo real, como el sistema ABI PRISM 7700™ Sequence Detection System™. El sistema consta de un termociclador, un láser, un dispositivo de carga acoplada (CCD), una cámara y un ordenador. El sistema incluye un software para ejecutar el instrumento y para analizar los datos.

Los datos del ensayo de 5' nucleasa se expresan inicialmente en forma de ciclo umbral (Ct). Tal y como se ha expuesto anteriormente, los valores de fluorescencia se registran durante cada ciclo y representan la cantidad de producto amplificado hasta ese punto en la reacción de amplificación. El punto en el que la señal fluorescente se registra por primera vez como estadísticamente significativa es el ciclo umbral (Ct).

Para minimizar los errores y el efecto de la variación entre muestras, la RT-PCR se realiza habitualmente utilizando un estándar interno. El estándar interno ideal se expresa a un nivel constante entre los diferentes tejidos y no se ve afectado por el tratamiento experimental. Los ARN utilizados más frecuentemente para normalizar los patrones de la expresión génica son ARNm para los genes

constitutivos gliceraldehído-3-fosfato-deshidrogenasa (GAPDH) y β -actina.

Una variación más reciente de la técnica RT-PCR es la PCR cuantitativa en tiempo real, que mide la acumulación de producto de la PCR a través de una sonda fluorogénica doblemente marcada (por ejemplo, la sonda TaqMan®). La PCR en tiempo real es compatible tanto con la PCR competitiva
5 cuantitativa, donde el competidor interno para cada secuencia diana se utiliza para la normalización, como con la PCR comparativa cuantitativa, utilizando un gen de normalización incluido en la muestra, o un gen constitutivo para RT-PCR. Para más información, véase, por ejemplo, Held et al., *Genome Research* 6:986-994 (1996).

3. Microarrays

La expresión génica diferencial también se puede identificar o confirmar utilizando la técnica microarray. Por tanto, el perfil de expresión de los genes asociados al cáncer de mama se puede medir en tejido tumoral fresco o embebido en parafina, utilizando la tecnología del microarray. En este método, las secuencias de polinucleótidos de interés (incluyendo ADNc y oligonucleótidos) se colocan en placas o en arrays o en un sustrato de microchip. Las secuencias de arrays se hibridan a continuación con sondas de
10 ADN específicas de células o tejidos de interés. Como en el método de la RT-PCR, la fuente de ARNm es típicamente ARN completo aislado de líneas de células tumorales o tumores humanos y las correspondientes líneas celulares o tejidos normales. Por tanto, el ARN puede ser aislado de diversas líneas de células tumorales o tumores primarios. Si la fuente del ARNm es un tumor primario, el ARNm se puede extraer, por ejemplo, de muestras de tejido congelado o fijado (por ejemplo, fijado con formalina) y embebido en parafina guardado, que son preparadas y conservadas rutinariamente en la práctica clínica
15 diaria.

En una realización específica de la técnica microarray, los insertos amplificados por PCR de los clones de ADNc se aplican sobre un sustrato en un array denso. Preferiblemente se aplican al menos 10 000 secuencias de nucleótidos al sustrato. Los genes del microarray, inmovilizados sobre el microchip a
20 10 000 elementos cada uno, resultan adecuados para la hibridación en condiciones restrictivas. Las sondas de ADNc marcadas fluorescentemente se pueden generar mediante la incorporación de nucleótidos fluorescentes por transcripción inversa del ARN extraído de los tejidos de interés. Las sondas de ADNc marcadas aplicadas sobre el chip se hibridan con especificidad a cada punto de ADN del array. Tras el lavado restrictivo para eliminar sondas unidas de forma no específica, el chip se explora mediante
25 microscopía láser confocal u otro método de detección, como una cámara CCD. La cuantificación de la hibridación de cada elemento del array permite evaluar la correspondiente abundancia de ARNm. Con la fluorescencia de dos colores, las sondas de ADNc marcadas separadamente generadas a partir de dos fuentes de ARN se hibridan por pares en el array. La abundancia relativa de las transcripciones de las dos fuentes correspondientes a cada gen especificado se determina por tanto de forma simultánea. La escala
30 miniaturizada de la hibridación permite una evaluación rápida y cómoda del patrón de expresión de grandes cantidades de genes. Se ha demostrado que estos métodos tienen la sensibilidad necesaria para detectar transcripciones poco comunes, que se expresan en unas pocas copias por célula, y para detectar de forma reproducible al menos aproximadamente diferencias de 2x en los niveles de expresión (Schena et al., *Proc. Natl. Acad. Sci. USA* 93(2):106-149 (1996)).

Los análisis de microarrays se pueden realizar con cualquier equipo disponible en el mercado, siguiendo los protocolos del fabricante, por ejemplo utilizando la tecnología Affymetrix GenChip o la tecnología microarray de Incyte.

El desarrollo de métodos microarrays para el análisis a gran escala de la expresión génica hace posible la búsqueda sistemática de marcadores moleculares de la clasificación del cáncer y predecir el resultado en
45 diversos tipos de tumores.

4. 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 un gran número de transcripciones de genes, sin necesidad de proporcionar una sonda de hibridación individual para cada transcripción. Primero se genera una etiqueta de secuencia corta (aprox.
50 10-14 pb) que contiene información suficiente para identificar de forma única una transcripción, siempre que la etiqueta se obtenga de una posición única dentro de cada transcripción. A continuación, muchas transcripciones se unen para formar moléculas en serie largas, que pueden ser secuenciadas, revelando la identidad de las múltiples etiquetas simultáneamente. El patrón de expresión de cualquier población de transcripciones puede ser evaluado cuantitativamente determinando la abundancia de etiquetas
55 individuales e identificando el gen correspondiente a cada etiqueta. Para más información, véase, por ejemplo, Velculescu et al., *Science* 270:484-487 (1995); y Velculescu et al., *Cell* 88:243-51 (1997).

5. Análisis de expresión génica mediante secuenciación de firmas masivamente en paralelo (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 la clonación *in vitro* de millones de plantillas en microperlas separadas de 5 μ m de diámetro. Primero se construye mediante
60 clonación *in vitro* una biblioteca de microperlas de plantillas de ADN. A continuación se monta un array

plano de las microperlas que contienen la plantilla en una celda de flujo a una densidad elevada (típicamente superior a 3×10^6 microperlas/cm²). Los extremos libres de las plantillas clonadas en cada microperla son analizados simultáneamente, utilizando un método de secuenciación de firmas basado en la fluorescencia que no requiere la separación de los fragmentos de ADN. Se ha demostrado que este método proporciona de forma simultánea y precisa, en una única operación, cientos de miles de secuencias de firmas génicas de una librería de ADNc de levaduras.

6. Inmunohistoquímica

Los métodos de inmunohistoquímica también resultan adecuados para detectar los niveles de expresión de los marcadores pronósticos de la presente divulgación. Por tanto, los anticuerpos o antisueros, preferiblemente antisueros policlonales, y más preferiblemente anticuerpos monoclonales específicos para cada marcador se utilizan para detectar la expresión. Los anticuerpos se pueden detectar mediante el marcado directo de los propios anticuerpos, por ejemplo, con etiquetas radiactivas, etiquetas fluorescentes, etiquetas de haptenos, tales como biotina, o una enzima como peroxidasa de rábano o fosfatasa alcalina. Alternativamente, el anticuerpo primario no marcado se utiliza conjuntamente con un anticuerpo secundario marcado, que comprende antisueros, antisueros policlonales o un anticuerpo monoclonal específico para el anticuerpo primario. Los protocolos de inmunohistoquímica son kits bien conocidos en la técnica y están disponibles en el mercado.

7. Proteómica

El término «proteoma» se define como la totalidad de las proteínas presentes en una muestra (por ejemplo, tejido, organismo o cultivo celular) en un determinado punto en el tiempo. La proteómica incluye, entre otras cosas, el estudio de los cambios globales de la expresión proteica en una muestra (también denominada «expresión proteómica»). La proteómica incluye típicamente los pasos siguientes: 1) separación de proteínas individuales en una muestra mediante electroforesis bidimensional en gel (2-D PAGE); 2) identificación de las proteínas individuales recuperadas del gel, por ejemplo mediante espectrometría de masas o secuenciación N-terminal; y 3) análisis de los datos utilizando bioinformática. Los métodos de la proteómica son suplementos valiosos para otros métodos de determinación del perfil de expresión génica y pueden ser utilizados solos o en combinación con otros métodos para detectar los productos de los marcadores pronósticos de la presente invención.

8. Conjunto de genes del cáncer. Subsecuencias génicas sometidas a ensayo y aplicación clínica de los datos de expresión génica

Un aspecto importante de la presente divulgación consiste en utilizar la expresión medida de determinados genes, o sus productos de expresión, en el tejido canceroso, por ejemplo cáncer de mama, para proporcionar información pronóstica. Para ello es necesario corregir (normalizar) tanto las diferencias en la cantidad de ARN sometido a ensayo como la variabilidad de la calidad del ARN utilizado. Por tanto, el ensayo típicamente mide e incorpora la expresión de determinados genes de normalización, incluyendo genes constitutivos bien conocidos, tales como GAPDH y β -ACTINA. Alternativamente, la normalización se puede basar en la mediana o la media de la señal (Ct) de todos los genes sometidos a ensayo o un gran subconjunto de estos (enfoque de normalización global). Durante toda la divulgación, a menos que se indique lo contrario, la referencia a los niveles de expresión de un gen asume la expresión normalizada con respecto al conjunto de referencia aunque esto no siempre se indica de forma explícita.

9. Algoritmo para generar una puntuación de recidiva del cáncer

Cuando se utiliza la qRT-PCR para medir los niveles de ARNm, las cantidades de ARNm se expresan en unidades de Ct (ciclo umbral) (Held *et al.*, *Genome Research* 6:986-994 (1996)). La suma media de los Ct de ARNm de referencia se establece en cero y cada Ct de ARNm de ensayo medido se proporciona en relación con este punto cero. Por ejemplo, si para la muestra tumoral de un determinado paciente se concluye que la media de Ct de los 5 genes de referencia es 31 y que el Ct del gen de ensayo CRB7 es 35, el valor reportado para GRB7 es -4 (es decir, 31-35).

Como primer paso tras la determinación cuantitativa de los niveles de ARNm, los genes identificados en la muestra tumoral 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 «subconjunto de proliferación» (eje). Los genes que se sabe que están asociados con la invasión del cáncer constituirán el «subconjunto de invasión» (eje). Los genes asociados con la vía o vías principales receptoras del factor de crecimiento constituirán el «subconjunto del factor de crecimiento» (eje). Los genes que se sabe que participan en la activación o señalización a través del receptor de estrógeno (ER) constituirán el «subconjunto del receptor de estrógenos» (eje). La lista de subconjuntos no es, por supuesto, exhaustiva. Los subconjuntos (ejes) creados dependerán del cáncer concreto, es decir cáncer de mama, de próstata, de páncreas, de pulmón, etc. En general, los genes cuya expresión se sabe que está correlacionada o que se sabe que están implicados en la misma vía se agrupan en el mismo eje.

En el siguiente paso, el nivel tumoral medida de cada ARNm en un subconjunto se multiplica por un coeficiente que refleja su contribución relativa dentro del conjunto al riesgo de recidiva de cáncer y su producto se suma a los demás productos entre los niveles de ARNm del subconjunto y sus coeficientes,

ES 2 651 849 T3

para obtener un término, por ejemplo un término de proliferación, un término de invasión, un término del factor de crecimiento, etc. Por ejemplo, en el caso del cáncer de mama invasivo negativo en nódulos linfáticos, el término del factor de crecimiento es $(0,45 \text{ a } 1,35) \times \text{GRB7} + (0,05 \text{ a } 0,15) \times \text{Her2}$ (véase el ejemplo a continuación).

- 5 La contribución de cada término a la puntuación de recidiva total se pondera utilizando un coeficiente. Por ejemplo, en el caso del cáncer de mama invasivo negativo en nódulos linfáticos, el coeficiente del término del factor de crecimiento puede situarse entre 0,23 y 0,70.

- 10 Adicionalmente, para algunos términos, como los términos del factor de crecimiento y de proliferación, se realiza un paso adicional. Si la relación entre el término y el riesgo de recidiva es no lineal, se utiliza una transformación funcional no lineal del término, como un umbral. Por tanto, en el cáncer de mama invasivo negativo en nódulos linfáticos, cuando se concluye que el término del factor de crecimiento es < -2 , el valor se fija en -2 . Cuando se concluye que el término de proliferación es $< -3,5$, el valor se fija en $-3,5$.

La suma de los términos obtenidos proporciona la puntuación de recidiva (RS).

- 15 Una relación entre la puntuación de recidiva (RS) y el riesgo de recidiva se ha encontrado midiendo la expresión del ensayo y los genes de referencia en muestras tumorales biopsiadas de una población de pacientes negativos en nódulos linfáticos con cáncer de mama invasivo y aplicando el algoritmo.

- 20 Cabe señalar que la escala de RS generada por el algoritmo de la presente invención se puede ajustar de diversas maneras. Por tanto, a pesar de que la escala de RS específicamente descrita más arriba abarca efectivamente de $-4,5$ a $-2,0$, el rango se podría seleccionar de forma que la escala abarcase de 0 a 10, 0 a 50, o 0 a 100, por ejemplo.

Por ejemplo, en un enfoque de escalas concreto, la puntuación de recidiva a escala (SRS) se calcula en una escala de 0 a 100. Por comodidad, se añaden 10 unidades Ct a cada valor Ct medido y la RS no sometida a escala se calcula como se ha descrito anteriormente. Las puntuaciones de recidiva a escala (SRS) se calculan utilizando las ecuaciones que se muestran a continuación.

- 25 Puntuación GRB7 = $0,9 \times \text{GRB7} + 0,1 \times \text{HER2}$

$$\text{Puntuación umbral GRB7} = \begin{cases} 8 & \text{si puntuación GRB7} < 8 \\ \text{Puntuación GRB7 de lo contrario} & \end{cases}$$

$$\text{Puntuación ER} = (0,8 \times \text{Esrt1} + 1,2 \times \text{PR} + \text{Bcl2} + \text{CEGP1})/4$$

$$\text{Puntuación de proliferación} = (\text{sURV} + \text{Ki-67} + \text{MYBL2} + \text{CCNB1} + \text{STK15})/5$$

- 30 Puntuación umbral prolif. = $\begin{cases} 6,5 & \text{si puntuación proliferación} < 6,5 \\ \text{Puntuación proliferación} & \text{de lo contrario} \end{cases}$

$$\text{Puntuación invasión} = (\text{CTSL2} + \text{STYM3})/2$$

$$\text{SRS} = 0 \text{ si } 20 \times (\text{RS} - 6,7) < 0$$

$$100 \text{ si } 20 \times (\text{RS} - 6,7) > 100$$

- 35 $20 \times (\text{RS} - 6,7)$ de lo contrario

donde

$$\text{RS} = 0,47 \times \text{puntuación umbral grupo GRB7}$$

$$-0,34 \times \text{puntuación grupo ER}$$

- 40 $+1,04 \times \text{puntuación umbral grupo proliferación}$

$$+0,10 \times \text{puntuación grupo invasión}$$

$$+0,05 \times \text{CD68}$$

$$-0,08 \times \text{GSTM1}$$

$$-0,07 \times \text{antigua puntuación BAG1}$$

- 45 Pacientes asignados a diversas categorías de riesgo utilizando las siguientes puntuaciones de recidiva:

Categoría de riesgo	Puntuación de recidiva
Bajo riesgo de recidiva	$RS < 18$
Riesgo intermedio	$18 \leq RS < 31$
Alto riesgo de recidiva	$RS \geq 31$

Este enfoque a escala se utilizó para calcular las puntuaciones de recidiva mostradas en la Figura 2.

10. Uso de las puntuaciones de recidiva y respuesta al tratamiento

5 Las puntuaciones de recidiva (RS) o las puntuaciones de respuesta al tratamiento (RTS) determinadas por el algoritmo de la presente divulgación proporcionan herramientas valiosas al médico responsable de tomar las decisiones de tratamiento críticas. Por tanto, si la RS de un paciente particular es baja, el médico podría decidir que la quimioterapia tras la extirpación quirúrgica del cáncer, por ejemplo un cáncer de mama, no resulta necesaria para garantizar la supervivencia a largo plazo del paciente. Como resultado, el paciente no se verá expuesto a los efectos secundarios, a menudo muy graves, del tratamiento quimioterapéutico estándar. Si, por otra parte, se determina una RS elevada, esta información puede ser utilizada para decidir qué opción de tratamiento quimioterapéutico o de otro tipo (como terapia de radiación) resulta más adecuada para tratar al paciente. De forma similar, si la puntuación RTS de un paciente para un tratamiento específico es baja, se utilizarán otras modalidades de tratamiento más efectivas para combatir el cáncer en ese paciente particular.

10 A modo de ejemplo, las opciones quimioterapéuticas estándar actuales para el tratamiento del cáncer de mama incluyen la administración de antraciclinas más ciclofosfamida (AC); o AC más un taxano (ACT); o C más metotrexato más 5-fluorouracilo (CMF); o la administración de cualquiera de estos fármacos en monoterapia o en cualquier combinación. Otros fármacos quimioterapéuticos utilizados a menudo en el tratamiento del cáncer, por ejemplo cáncer de mama, incluyen, entre otros, Herceptin®, vinblastina, actinomicina D, etopósido, cisplatino, y doxorubicina. La determinación de la RS para un determinado paciente a través del algoritmo de la presente invención permitirá al médico adaptar el tratamiento del paciente, de forma que el paciente tenga más oportunidades de supervivencia a largo plazo al tiempo que se minimizan los efectos secundarios no deseados.

15 Por tanto, los pacientes con bajo riesgo de recidiva del cáncer son típicamente asignados a un tratamiento hormonal en monoterapia, o tratamiento hormonal y un régimen de tratamiento quimioterapéutico menos tóxico, por ejemplo utilizando Herceptin®. Por otra parte, los pacientes cuyo riesgo de recidiva del cáncer determinado es medio o alto se someten típicamente a quimioterapia más agresiva, como, por ejemplo, antraciclina y/o regímenes de tratamiento a base de taxano.

20 El valor de RS también se puede utilizar para decidir si tratar al paciente con fármacos terapéuticos que en estos momentos no forman parte del protocolo de tratamiento de primera línea para un cáncer concreto, como, por ejemplo, inhibidores de EGFR y/u otras opciones de tratamiento, como la radiación en monoterapia, o antes o después del tratamiento quimioterapéutico.

En los siguientes ejemplos proporcionados exclusivamente a título ilustrativo se proporcionarán más detalles de la invención.

35 EJEMPLO 1

Algoritmo para predecir la recidiva del cáncer de mama

40 El algoritmo se basa en las mediciones de los niveles tumorales de 16 marcadores de ARNm, que se seleccionan en función de dos criterios principales: 1) correlación univariante con la recidiva del cáncer de mama en el estudio clínico que se describe a continuación, y 2) implicación en la patología tumoral indicada por la bibliografía científica publicada.

45 Los marcadores seleccionados indican funciones de diversas conductas celulares y vías en la recidiva del cáncer: por ejemplo, Her2 del factor de crecimiento, receptor de estrógenos, proliferación e invasión. En consonancia con los actuales conocimientos de la patología molecular del cáncer de mama, los niveles aumentados de ARNm que representan genes en los ejes del factor de crecimiento, proliferación e invasión están correlacionados con el riesgo incrementado de recidiva, mientras que los niveles aumentados de ARNm que representan genes en el eje del receptor de estrógenos están correlacionados con un riesgo reducido de recidiva. Los genes relevantes en estas vías no solamente están vinculados por la función biológica, sino que también están vinculados por coexpresión, tal y como indican los coeficientes de correlación de Pearson (datos no mostrados). Para el cálculo del coeficiente de correlación de Pearson, véase, por ejemplo, K. Pearson y A. Lee, *Biometrika* 2:357 (1902).

50 Se construyó un algoritmo ponderando las contribuciones de determinados genes coexpresados en las anteriores vías. El modelo fue optimizado iterativamente mediante selección para una mejor correlación con el riesgo de recidiva. La adecuación del modelo incluyó la selección de los mejores genes que

representan los anteriores ejes funcionales y de coexpresión y la selección para unos coeficientes óptimos de la ecuación.

5 Por otra parte, se incluyeron otros genes que no estaban estrechamente correlacionados en expresión con ninguno de los genes anteriores ni entre sí, pero que también se concluyó que contribuyen independientemente a predecir la recidiva del cáncer. En vista de la multitud de evidencias recogidas en la bibliografía biomédica de que los macrófagos tumorales ofrecen un mal pronóstico del cáncer (M. Orre y P.A Rogers *Gynecol. Oncol.* 73: 47-50 [1999]; L.M. Coussens *et al. Cell* 103: 481-90 [2000]; S. Huang *et al. J. Natl. Cancer Inst.* 94:1134-42 [2002]; M.A. Cobleigh *et al. Proceedings of A.S.C.O.* 22: Abstract 3415 [2003]), el marcador del macrófago CD68 también se incluyó en el conjunto de genes de ensayo. Por 10 último, se identificó un modelo que incluía los genes: Grb7; Her2; ER; PR; Bcl2; CEGP1; SURV; Ki-67; MYBL2; CCNB1; STK15; CTSL2; STMY3; GSTM1; BAG1; CD68.

El algoritmo proporciona una puntuación del riesgo de recidiva (RS). Concretamente, la RS se obtiene como sigue:

1. Definir el eje GRB7

15 $ejeGRB7 = (0,45 \text{ a } 1,35) \times GRB7 + (0,05 \text{ a } 0,15) \times HER2$

2. Definir el umbral del eje GRB7
si $ejeGRB7 < -2$

entonces $umbralGRB7GT = -2$,

20 por lo demás, $umbralGRB7GT = ejeGRB7$

3. Definir el eje ER

$ejeEr = (EstR1 + PR + Bcl2 + CEGP1)/4$, donde las contribuciones individuales de los genes enumerados se pueden ponderar por un factor que oscila entre 0,5 y 1,5, inclusive.

4. Definir el eje de proliferación, $ejeprolif$

25 $ejeprolif = (SURV + Ki.67 + MYBL2 + CCNB1 + STK15)/5$, donde las contribuciones individuales de los genes enumerados se pueden ponderar por un factor entre 0,5 y 1,5, inclusive.

5. Definir el umbral del eje de proliferación, $umbralejeprolif$

si $ejeprolif < -3,5$

entonces $umbralejeprolif = -3,5$,

30 por lo demás, $umbralejeprolif = ejeprolif$

6. Definir el eje de invasión

$ejeinvas = (CTSL2 + STMY3)/2$, donde las contribuciones individuales de los genes enumeradas se pueden ponderar por un factor entre 0,5 y 1,5, inclusive.

7. Calcular la puntuación de recidiva (RS)

35 $RS = (0,23 \text{ a } 0,70) \times umbralGRB7GT - (0,17 \text{ a } 0,51) \times ejeER + (0,52 \text{ a } 1,56) \times umbralejeprolif + (0,07 \text{ a } 0,21) \times ejeinvas + (0,03 \text{ a } 0,15) \times CD68 - (0,04 \text{ a } 0,25) \times GSTM1 - (0,05 \text{ a } 0,22) \times BAG1$.

En una realización particular, la RS se calcula como sigue:

$RS = 0,47 \times umbralGRB7GT - 0,34 \times ejeER + 1,04 \times umbralejeprolif + 0,14 \times ejeinvas + 0,11 \times CD68 - 0,17 \times GSTM1 - 0,15 \times BAG1$.

40 Un experto en la técnica entenderá que se pueden sustituir otros genes por los genes de ensayo anteriormente especificados. En términos generales, cualquier gen que se co-expresa con un gen del panel de ensayo de 16-20 genes con un coeficiente de correlación de Pearson ≥ 0.40 puede ser sustituido por ese gen en el algoritmo. Para sustituir el gen X por el gen correlacionado Y, se determina el rango de valores para el gen X en la población de pacientes, el rango de valores para el gen Y en la población de 45 pacientes, y se aplica una transformación lineal para adaptar los valores del gen Y a los correspondientes valores del gen X. Por ejemplo, si el rango de valores para el gen X en la población de pacientes es 0 a 10 y el rango de valores para el gen Y en la población de pacientes es 0 a 5, la transformación adecuada consiste en multiplicar los valores del gen Y por 2.

50 A modo de ejemplo, los siguientes genes que se encuentran en el amplicón de Her2 son algunos de los genes que pueden ser sustituidos en el subconjunto de genes del factor de crecimiento: Q9BRT3; TCAP; PNMT; ML64; IPPD; Q9H7G1; Q9HBS1; Q9Y220; PSMD3; y CSF3

A modo de ejemplo, algunos de los genes que pueden ser sustituidos en el subconjunto de genes de proliferación son: C20.orf1, TOP2A, CDC20, KNSL2, MELK, TK1, NEK2,

LMNB1, PTTG1, BUB1, CCNE2, FLJ20354, MCM2, RAD54L, PRO2000, PCNA, Chk1, NME1, TS, FOXM1, AD024, y HNRPAB.

5 A modo de ejemplo, algunos de los genes que pueden ser sustituidos en el subconjunto del receptor de estrógenos (ER) son: HNF3A, ErbB3, GAT A3, BECN1, IGF1R, AKT2, DHPS, BAG1, hENT1, TOP2B, MDM2, CCND1, DKFZp586M0723, NPD009, B.Catenin, IRS1, Bclx, RBM5, PTEN, A.Catenin, KRT18, ZNF217, ITGA7, GSN, MTA1, G.Catenin, DR5, RAD51C, BAD, TP53BP1, RIZ1, IGFBP2, RUNX1, PPM1D, TFF3, S100A8, P28, SFRS5, y IGFBP2.

A modo de ejemplo, algunos de los genes que pueden ser sustituidos en el eje de invasión son: upa, COL1A1, COL1A2, y TIMP2.

10 Algunos de los genes que pueden ser sustituidos por CD68 son: CTSB, CD18, CTSL, HLA.DPB1, MMP9.

Algunos de los genes que pueden ser sustituidos por GSTM1 son: GSTM3.2, MYH11.1, GSN.3, ID1.1.

Algunos de los genes que pueden ser sustituidos por BAG1 son: Bcl2.2, GAT A3.3,

DHPS.3, HNF3A.1.

15 Los valores de RS determinados se pueden utilizar para orientar la decisión binaria (sí o no) de tratar a los pacientes con cáncer de mama negativo en nódulos linfáticos con quimioterapia adyuvante. Para este propósito se puede definir un umbral que separe a los pacientes de alto riesgo y de bajo riesgo. Como se ha señalado anteriormente, puede resultar más informativo definir tres categorías: riesgo alto, moderado y bajo, que requiere dos puntos de corte de la RS. Los puntos de corte se pueden seleccionar para
20 pacientes de subgrupos de categorías de riesgo deseadas aplicando el algoritmo a determinados datos de estudios clínicos. Por ejemplo, es posible que los oncólogos deseen evitar tratar a los pacientes con cáncer de mama negativo en nódulos linfáticos que tienen menos de un 10% de riesgo de recidiva en el plazo de 10 años.

25 La aplicación del algoritmo a los datos del ensayo clínico descrito en el Ejemplo 2 indica que los pacientes con una RS <-3,9 tienen menos de un 10% de riesgo de recidiva en 10 años, mientras que los pacientes con una RS > -3,5 tienen más de un 39% de riesgo de recidiva en 10 años. Los pacientes con valores de RS intermedios se pueden considerar en la categoría de riesgo moderado.

Los oncólogos pueden recibir los resultados de la RS en una cifra continua que relaciona la puntuación de recidiva con la supervivencia de Kaplan Meier a 10 años.

EJEMPLO 2

30 Estudio de la expresión génica en 242 tumores de mama malignos

Se diseñó y realizó un estudio de expresión génica para determinar los valores de la puntuación de recidiva (RS) en una población de pacientes con cáncer de mama invasivo nódulo negativo y explorar la correlación entre los valores de RS y la supervivencia libre de enfermedad. El estudio utilizó bloques tumorales fijados y embebidos en parafina guardados como fuente de ARN y comparó los historiales de
35 pacientes archivados.

Diseño del estudio:

40 Se realizaron ensayos moleculares con tejidos de tumores primarios de mama fijados con formalina y embebidos en parafina obtenidos de 252 pacientes individuales a los que se les había diagnosticado un cáncer de mama invasivo. Todos los pacientes eran nódulos negativos, ER positivos y recibían tratamiento con tamoxifeno. La edad media era de 52 años y el tamaño clínico medio del tumor era de 2 cm. La media de seguimiento era de 10,9 años. A 1 de enero de 2003, 41 pacientes tenían recidiva de la enfermedad local o distante o muerte por cáncer de mama. Los pacientes fueron incluidos en el estudio únicamente si la evaluación histopatológica, realizada tal y como se describe en la sección de Materiales y métodos, indicaba unas cantidades adecuadas de tejido tumoral y una patología homogénea.

45 Materiales y métodos:

50 Cada bloque de tumor representativo se caracterizó mediante histopatología estándar para el diagnóstico, la valoración semicuantitativa de la cantidad de tumor y grado del tumor. Cuando el área del tumor era inferior al 70% de la sección, el área del tumor se disecó por completo y se tomó tejido de seis secciones (10 micrones). De lo contrario, se preparó un total de tres secciones (también de 10 micrones de grosor
55 cada una). Las secciones se colocaron en dos tubos de microcentrifuga de la marca Costar (tubos de 1,7 mL de polipropileno transparente). Si se obtuvo más de un bloque de tumor como parte del procedimiento quirúrgico, el bloque más representativo de la patología se utilizó para el análisis. Análisis de expresión génica; el ARNm se extrajo y purificó a partir de muestras de tejido fijadas y embebidas en parafina, y se preparó para el análisis de expresión génica como se ha descrito anteriormente. Los ensayos moleculares de la expresión génica cuantitativa se realizaron mediante RT-PCR, utilizando el sistema ABI PRISM 7900™ Sequence Detection System™ (Perkin-Elmer-Applied Biosystems, Foster City, CA, EE. UU.). ABI PRISM 7900™ consta de un termociclador, un láser, un dispositivo de carga acoplada (CCD), una cámara y un ordenador.

El sistema amplifica muestras en un formato de 384 pocillos en un termociclador. El sistema incluye un software para ejecutar el instrumento y para analizar los datos.

Resultados:

5 Los valores de expresión génica normalizada se obtuvieron de las anteriores muestras de pacientes y se utilizaron para calcular la RS para cada paciente, que se contrastaron con los respectivos datos de supervivencia a 10 años de Kaplan-Meier, tal y como se muestra en las Figuras 1A y B. Como se muestra, los pacientes con valores de RS inferior a aproximadamente -3,75 tienen un 90% de probabilidades de supervivencia libre de recidiva a 10 años. Las tasas de recidiva aumentan de forma pronunciada a unos valores de RS > -3,3. A una RS aproximada de -3,0, el riesgo de recidiva a 10 años es aproximadamente del 30% y a una RS aproximada de -2,0, el riesgo de recidiva a 10 años es > 50%). La Figura 1B revela que aproximadamente el 70% de los pacientes se encuentran en la categoría de más bajo riesgo.

Los resultados demuestran por tanto una estrecha correspondencia entre los valores de RS determinados por el ensayo y el algoritmo descritos y el riesgo de recidiva del cáncer de mama.

EJEMPLO 3

15 Validación del algoritmo de la invención

Se realizó un estudio de validación clínica prospectivo para examinar el rendimiento de un ensayo multigénico por RT-PCR para extraer y cuantificar el ARN de tejido tumoral fijado y embebido en parafina en pacientes incluidos en el grupo del tamoxifeno en monoterapia del estudio *National Surgical Adjuvant Breast and Bowel Project (NSABP) Study B-14: A Laboratory Study To Clinically Validate Whether Genomic Tumor Expression Profiles Can Define The Likelihood Of Recurrence In Patients With Primary Invasive Breast Cancer, Negative Axillary Nodes And Estrogen-Receptor Positive Tumors*. El NSABP Study B-14 se realizó para valorar la eficacia del tratamiento con tamoxifeno posterior a la mastectomía en pacientes con cáncer de mama primario nódulo negativo e incluyó un total de 2892 pacientes aptos que recibieron tratamiento durante cinco años. El protocolo se realizó en forma de estudio aleatorizado a doble ciego con un conjunto de pacientes a los que se les administró tamoxifeno a una dosis de 10 mg dos veces al día y un conjunto de control de pacientes a los que se les administraron comprimidos de placebo físicamente idénticos. Se realizaron comparaciones del ensayo multigénico con medidas moleculares e histopatológicas estándar de la clasificación tumoral.

30 Las puntuaciones de recidiva específicas obtenidas de los pacientes del ensayo multigénico se obtuvieron para un total de 668 pacientes aptos del grupo del tamoxifeno en monoterapia del NSABP Study B-14 (tiempo de seguimiento medio: 14,1 años). Los análisis estadísticos revelaron que las puntuaciones de recidiva específicas de los pacientes resultantes (Figura 2) mantienen una correlación significativa con la probabilidad de recidiva distante y proporcionan información significativa más allá de las medidas clínicas estándar de la clasificación tumoral, incluyendo la edad del paciente en la cirugía, el tamaño clínico del tumor y el grado del tumor.

Se concluyó que el ensayo y el algoritmo de la presente divulgación proporcionan medidas reproducibles e informativas de la probabilidad de recidiva distante del tejido tumoral recogido en el momento de la cirugía para los pacientes con cáncer de mama primario nódulo negativo y ER positivo tratados con tamoxifeno. La puntuación de recidiva generada del paciente por el método de la presente invención proporciona información significativa ($P < 0,0001$) más allá de las medidas histopatológicas y moleculares estándar de la clasificación tumoral normalmente utilizada en la práctica clínica, incluyendo la edad, el tamaño clínico del tumor y el grado del tumor. A diferencia de otras medidas clínicas habituales, la puntuación de recidiva del paciente es altamente reproducible y a diferencia de otros tests moleculares aprovecha simultáneamente la información de múltiples marcadores genómicos, incluyendo ER, PR y HER2.

Aunque la presente invención se ha descrito por referencia a las que se consideran realizaciones específicas, se entenderá que la divulgación no se limita a estas realizaciones. Por el contrario, la divulgación pretende abarcar diversas modificaciones y equivalentes incluidos en el espíritu y el alcance de las reivindicaciones adjuntas. Por ejemplo, a pesar de que la divulgación se centra en la identificación de diversos genes y conjuntos de genes asociados al cáncer de mama, y en el pronóstico personalizado del cáncer de mama, otros genes y conjuntos de genes similares relativos a otros tipos de cáncer se incluyen específicamente en el ámbito de aplicación del presente.

Tabla 1

Gen	Núm. acceso	Nombre	Secuencia artif.	Secuencia	Longitud
B-actina	NM_001101	S0034/B-acti.f2	SEC. ID. N°: 1	CAGCAGATGTGGATCAGCAAG	21
B-actina	NM_001101	S0036/B-acti.r2	SEC. ID. N°: 2	GCATTTGCGGTGGACGAT	18
B-actina	NM_001101	S4730/B-acti.p2	SEC. ID. N°: 3	AGGAGTATGACGAGTCCGGCCCC	23
BAG1	NM_004323	S1386/BAG1.f2	SEC. ID. N°: 4	CGTTGTCAGCACTTGGGAATAAA	23

ES 2 651 849 T3

BAG1	NM_004323	S1387/BAG1.r2	SEC. ID. Nº: 5	GTTCAACCTCTTCCTGTGGACTGT	24
BAG1	NM_004323	S4731/BAG1.p2	SEC. ID. Nº: 6	CCCAATTAACATGACCCGGCAACCAT	26
Bcl2	NM_000633	S0043/Bcl2.f2	SEC. ID. Nº: 7	CAGATGGACCTAGTACCCACTGAGA	25
Bcl2	NM_000633	S0045/Bcl2.r2	SEC. ID. Nº: 8	CCTATGATTTAAGGGCATTITTTCC	24
Bcl2	NM_000633	S4732/Bcl2.p2	SEC. ID. Nº: 9	TTCCACGCCGAAGGACAGCGAT	22
CCNB1	NM_031966	S1720/CCNB1.f2	SEC. ID. Nº: 10	TTCAGGTTGTTGCAGGAGAC	20
CCNB1	NM_031966	S1721/CCNB1.r2	SEC. ID. Nº: 11	CATCTTCTTGGGCACACAAT	20
CCNB1	NM_031966	S4733/CCNB1.p2	SEC. ID. Nº: 12	TGTCTCCATTATTGATCGGTTTCATGCA	27
CD68	NM_001251	S0067/CD68.f2	SEC. ID. Nº: 13	TGGTTCCCAGCCCTGTGT	18
CD68	NM_001251	S0069/CD68.r2	SEC. ID. Nº: 14	CTCCTCCACCCTGGGTTGT	19
CD68	NM_001251	S4734/CD68.p2	SEC. ID. Nº: 15	CTCCAAGCCCAGATTTCAGATTTCGAGTCA	28
CEGP1	NM_020974	S1494/CEGP1.f2	SEC. ID. Nº: 16	TGACAATCAGCACACCTGCAT	21
CEGP1	NM_020974	S1495/CEGP1.r2	SEC. ID. Nº: 17	TGTGACTACAGCCGTGATCCTTA	23
CEGP1	NM_020974	S4735/CEGP1.p2	SEC. ID. Nº: 18	CAGGCCCTCTCCGAGCGGT	20
CTSL2	NM_001333	S4354/CTSL2.f1	SEC. ID. Nº: 19	TGTCTCACTGAGCGAGCAGAA	21
CTSL2	NM_001333	S4355/CTSL2.r1	SEC. ID. Nº: 20	ACCATTGCAGCCCTGATTG	19
CTSL2	NM_001333	S4356/CTSL2.p1	SEC. ID. Nº: 21	CTTGAGGACGCGAACAGTCCACCA	24
EstR1	NM_000125	S0115/EstR1.f1	SEC. ID. Nº: 22	CGTGGTGCCCTCTATGAC	19
EstR1	NM_000125	S0117/EstR1.r1	SEC. ID. Nº: 23	GGCTAGTGGGCGCATGTAG	19
EstR1	NM_000125	S4737/EstR1.p1	SEC. ID. Nº: 24	CTGGAGATGCTGGACGCC	19
GAPDH	NM_002046	S0374/GAPDH.f1	SEC. ID. Nº: 25	ATTCCACCCATGGCAAATTC	20
GAPDH	NM_002046	S0375/GAPDH.r1	SEC. ID. Nº: 26	GATGGGATTTCATTGATGACA	22
GAPDH	NM_002046	S4738/GAPDH.p1	SEC. ID. Nº: 27	CCGTTCTCAGCCTTGACGGTGC	22
GRB7	NM_005310	S0130/GRB7.f2	SEC. ID. Nº: 28	ccatctgcatccatctgtt	20
GRB7	NM_005310	S0132/GRB7.r2	SEC. ID. Nº: 29	ggccaccagggtattatctg	20
GRB7	NM_005310	S4726/GRB7.p2	SEC. ID. Nº: 30	ctccccacccttgagaagtgcct	23
GSTM1	NM_000561	S2026/GSTM1.r1	SEC. ID. Nº: 31	GGCCCAGCTTGAATTTTTCA	20
GSTM1	NM_000561	S2027/GSTM1.f1	SEC. ID. Nº: 32	AAGCTATGAGGAAAAGAAGTACACGAT	27
GSTM1	NM_000561	S4739/GSTM1.p1	SEC. ID. Nº: 33	TCAGCCACTGGCTTCTGTCATAATCAGG	30
GUS	NM_000181	S0139/GUS.f1	SEC. ID. Nº: 34	ĈĈCACTCAGTAGCCAAGTCA	20
GUS	NM_000181	S0141/GUS.r1	SEC. ID. Nº: 35	CACGCAGGTGGTATCAGTCT	20
GUS	NM_000181	S4740/GUS.p1	SEC. ID. Nº: 36	TCAAGTAAACGGGCTGTTTTCCAAACA	27
HER2	NM_004448	S0142/HER2.f3	SEC. ID. Nº: 37	CGGTGTGAGAAGTGCAGCAA	20
HER2	NM_004448	S0144/HER2.r3	SEC. ID. Nº: 38	CCTCTCGCAAAGTGCTCCAT	19
HER2	NM_004448	S4729/HER2.p3	SEC. ID. Nº: 39	CCAGACCATAGCACACTCGGGCAC	24
Ki-67	NM_002417	S0436/Ki-67.f2	SEC. ID. Nº: 40	CGGACTTTGGGTGCGACTT	19
Ki-67	NM_002417	S0437/Ki-67.r2	SEC. ID. Nº: 41	TTACAACCTCTTCCACTGGGACGAT	24
Ki-67	NM_002417	S4741/Ki-67.p2	SEC. ID. Nº: 42	CCACTTGTGGAACCACCGCTCGT	23
MYBL2	NM_002466	S3270/MYBL2.f1	SEC. ID. Nº: 43	GCCGAGATCGCCAAGATG	18
MYBL2	NM_002466	S3271/MYBL2.r1	SEC. ID. Nº: 44	CTTTTGATGGTAGAGTTCAGTGATTC	27
MYBL2	NM_002466	S4742/MYBL2.p1	SEC. ID. Nº: 45	CAGCATTGTCTGTCCTCCCTGGCA	24
PR	NM_000926	S1336/PR.f6	SEC. ID. Nº: 46	GCATCAGGCTGTCATTATGG	20
PR	NM_000926	S1337/PR.r6	SEC. ID. Nº: 47	AGTAGTTGTGCTGCCCTTCC	20
PR	NM_000926	S4743/PR.p6	SEC. ID. Nº: 48	TGTCCTTACCTGTGGGAGCTGTAAGGTC	28
RPLPO	NM_001002	S0256/RPLPO.f2	SEC. ID. Nº: 49	CCATTCTATCATCAACGGGTACAA	24
RPLPO	NM_001002	S0258/RPLPO.r2	SEC. ID. Nº: 50	TCAGCAAGTGGGAAGGTGTAATC	23
RPLPO	NM_001002	S4744/RPLPO.p2	SEC. ID. Nº: 51	TCTCCACAGACAAGGCCAGGACTCG	25
STK15	NM_003600	S0794/STK15.f2	SEC. ID. Nº: 52	CATCTTCCAGGAGGACCACT	20
STK15	NM_003600	S0795/STK15.r2	SEC. ID. Nº: 53	TCCGACCTCAATCATTCA	20
STK15	NM_003600	S4745/STK15.p2	SEC. ID. Nº: 54	CTCTGTGGCACCCCTGGACTACCTG	24
STMY3	NM_005940	S2067/STMY3.f3	SEC. ID. Nº: 55	CCTGGAGGCTGCAACATACC	20

ES 2 651 849 T3

STMY3	NM_005940	S2068/STMY3.r3	SEC. ID. Nº: 56	TACAATGGCTTTGGAGGATAGCA	23
STMY3	NM_005940	S4746/STMY3.p3	SEC. ID. Nº: 57	ATCCTCCTGAAGCCCTTTCGCAGC	25
SURV	NM_001168	S0259/SURV.f2	SEC. ID. Nº: 58	TGTTTTGATTCCCGGGCTTA	20
SURV	NM_001168	S0261/SURV.r2	SEC. ID. Nº: 59	CAAAGCTGTCAGCTCTAGCAAAG	24
SURV	NM_001168	S4747/SURV.p2	SEC. ID. Nº: 60	TGCCTTCTTCCCTCCCTCACTTCTCACCT	28
TFRC	NM_003234	S1352/TFRC.f3	SEC. ID. Nº: 61	GCCAAGTCTTTTCAATTTGTG	20
TFRC	NM_003234	S1353/TFRC.r3	SEC. ID. Nº: 62	ACTCAGGCCCATTTCTTTA	20
TFRC	NM_003234	S4748/TFRC.p3	SEC. ID. Nº: 63	AGGGATCTGAACCAATACAGAGCAGAC	28

A

Tabla 2

Gen	Núm. acceso	Secuencia artif.	Secuencia
B-actina	NM_001101	SEC. ID. Nº: 64	CAGCAGATGTGGATCAGCAAGCAGGAGTATGACGAGTCCGGCCCTCCATCGTCCACCGCAATGC
BAG1	NM_004323	SEC. ID. Nº: 65	CGTTGTCCAGCAGCTGGAATACAAGATGGTTGCCGGGTCATGTAATTGGGAAAAAGAACAGTCCACAGGAAGAGGTTGAAC
Bcl2	NM_000633	SEC. ID. Nº: 66	CAGATGGACCTAGTACCCACTGAGATTTCCACGCCAAGGACAGCGATGGGAAAAATGCCCTAAATCATAGG
CCNB1	NM_031966	SEC. ID. Nº: 67	TTCAAGTTGTTGCCAGGAGACCATGTACATGACTGTCTCCATTATTGATCGGTTTCATGCAGATAATTGTGTGCCAAGAAGATG
CD68	NM_001251	SEC. ID. Nº: 68	TGGTTCACGACCTGTGTCCACTCCAAGCCAGATTTCAGATTCGAGTGCATGTACACAACCCAGGGTGGAGAG
CEGP1	NM_020974	SEC. ID. Nº: 69	TGACAATCAGCACACCTGCATTCCAGCTCGGAAGAGGGCCGAGCTGCATGAATAAGGATCACGGCTGTAGTCACA
CTSL2	NM_001333	SEC. ID. Nº: 70	TGTCTCACTGAGCGAGCAGAATCTGGTGGACTGTCGCGTCCCAAGGCAATCAGGGCTGCAATGGT
EstR1	NM_000125	SEC. ID. Nº: 71	CGTGGTCCCTCTATGACCTGCTGCTGGAGATGCTGGACGCCACCCGCTACATGCCCCACTAGCC
GAPDH	NM_002046	SEC. ID. Nº: 72	ATTCACCCATGGCAAATCCATGGCACCGTCAAGGCTGAGAAACGGGAAGCTTGTATCAATGGAAATCCCATC
GRB7	NM_005310	SEC. ID. Nº: 73	CCATGTGCATCCATCTTGTGGCTCCACCCCTTGAGAAAGTGCCTCAGATAATACCCCTGGTGGCC
GSTM1	NM_000561	SEC. ID. Nº: 74	AAGCTATGAGGAAAAAGTACAGATGGGGACGCTCCTGATTATGACAGAAGCCAGTGGCTGAATGAAAAATCAAGCTGGGCC
GUS	NM_000181	SEC. ID. Nº: 75	CCCACTCAGTACCAAGTCAATGTTGGAAAAAGCCCGTTTACTTGAGCAAGACTGATACCACCTGCGTG
HER2	NM_004448	SEC. ID. Nº: 76	CGGTGTGAGAAGTGCAGCAAGCCCTGTGCCGAGTGTGCTATGGTCTGGGCATGGGACTTGGGAGAGG
Ki-67	NM_002417	SEC. ID. Nº: 77	CGGACTTTGGGTGCGACTTGACGAGCGTGGTTCGACAAGTGGCCCTGCGGGCCGGATCGTCCAGTGGAAAGAGTTGTAA
MYBL2	NM_002466	SEC. ID. Nº: 78	GCCGAGATCGCCAAGATGTTGCCAGGAGGACAGACAATGCTGTGAAGAATCACTGGAACCTTACCATCAAAAG
PR	NM_000926	SEC. ID. Nº: 79	GCATCAGGCTGTCTATTGTTGCTTACCTGTGGAGCTGTAAGGTCTTCTTTAAGAGGGCAATGGAAGGGCAGCACAACACT
RPLPO	NM_001002	SEC. ID. Nº: 80	CCATTCATCATCAACGGGTACAACAGTCCCTGGCCTTGTCTGTGAGACGGATTACACCTTCCACTTGTCTGA
STK15	NM_003600	SEC. ID. Nº: 81	CATCTCCAGGAGGACCCTCTCTGTGGCACCCTGGACTACCTGCCCTGAAATGATTGAAGGTCCGA
STMY3	NM_005940	SEC. ID. Nº: 82	CCTGGAGGCTGCAACATACCTCAATCCTGTCCAGGCCGATCCTCTGAAGCCCTTTTCGACGACTGCTATCCTCCAAAGCCATTGTA
SURV	NM_001168	SEC. ID. Nº: 83	TGTTTTGATCCCGGCTTACCAGGTGAGAAGTGAAGGAGGAAGGCAGTGTCCCTTTTCTAGAGCTGACAGCTTTG
TFRC	NM_003234	SEC. ID. Nº: 84	GCCAAGTCTTTCATTTGTGAGGATCTGAACCAATACAGAGCAGACATAAAGAAATGGGCTGAGT

LISTADO DE SECUENCIAS

<110> BAKER, Joffre
 5 BRYANT, Dohn L.
 PAIK, Soonmyung
 SHAK, Steven

 <120> Algoritmo del perfil de expresión y test para el pronóstico del cáncer
 10
 <130> 39740-0013 PCT

 <140> por asignar
 <141> 2004-06-30
 15
 <150> US 60/486 302
 <151> 10/07/2003

 <150> US 60/526 947
 20 <151> 03/12/2003

 <160> 84

 <170> FastSEQ para windows versión 4.0
 25
 <210> 1
 <211> 21
 <212> ADN
 <213> Secuencia artificial
 30
 <220>
 <223> Cebador-sonda PCR

 <400> 1
 35 cagcagatgt ggatcagcaa g 21
 <210> 2
 <211> 18
 <212> ADN
 <213> Secuencia artificial

	<220>	
	<223> Cebador-sonda PCR	
5	<400> 2 gcatttgcgg tggacgat	18
	<210> 3	
	<211> 23	
	<212> ADN	
10	<213> Secuencia artificial	
	<220>	
	<223> Cebador-sonda PCR	
15	<400> 3 aggagtatga cgagtccggc ccc	23
	<210> 4	
	<211> 23	
	<212> ADN	
20	<213> Secuencia artificial	
	<220>	
	<223> Cebador-sonda PCR	
25	<400> 4 cgttgtcagc acttggata caa	23
	<210> 5	
	<211> 24	
	<212> ADN	
30	<213> Secuencia artificial	
	<220>	
	<223> Cebador-sonda PCR	
35	<400> 5 gttcaacctc ttctgtgga ctgt	24
	<210> 6	
	<211> 26	
	<212> ADN	
	<213> Secuencia artificial	

<220>

<223> Cebador-sonda PCR

5 <400> 6
cccaattaac atgacccggc aaccat

26

<210> 7

<211> 25

<212> ADN

10 <213> Secuencia artificial

<220>

<223> Cebador-sonda PCR

15 <400> 7
cagatggacc tagtaccac tgaga

25

<210> 8

<211> 24

<212> ADN

20 <213> Secuencia artificial

<220>

<223> Cebador-sonda PCR

25 <400> 8
cctatgattt aagggcattt ttcc

24

<210> 9

<211> 22

<212> ADN

30 <213> Secuencia artificial

<220>

<223> Cebador-sonda PCR

35 <400> 9
ttccacgccg aaggacagcg at

22

<210> 10

<211> 20

<212> ADN

	<213> Secuencia artificial	
	<220>	
	<223> Cebador-sonda PCR	
5	<400> 10 ttcaggttgt tgcaggagac	20
	<210> 11	
	<211> 20	
	<212> ADN	
10	<213> Secuencia artificial	
	<220>	
	<223> Cebador-sonda PCR	
15	<400> 11 catcttcttg ggcacacaat	20
	<210> 12	
	<211> 27	
	<212> ADN	
20	<213> Secuencia artificial	
	<220>	
	<223> Cebador-sonda PCR	
25	<400> 12 tgtctccatt attgatcggg tcatgca	27
	<210> 13	
	<211> 18	
	<212> ADN	
30	<213> Secuencia artificial	
	<220>	
	<223> Cebador-sonda PCR	
35	<400> 13 tggttcccag ccctgtgt	18
	<210> 14	
	<211> 19	
	<212> ADN	

	<213> Secuencia artificial	
	<220>	
	<223> Cebador-sonda PCR	
5	<400> 14 ctcctccacc ctqggttgt	19
	<210> 15	
	<211> 28	
	<212> ADN	
10	<213> Secuencia artificial	
	<220>	
	<223> Cebador-sonda PCR	
15	<400> 15 ctccaagccc agattcagat tcgagtca	28
	<210> 16	
	<211> 21	
	<212> ADN	
	<213> Secuencia artificial	
20	<220>	
	<223> Cebador-sonda PCR	
	<400> 16	
25	tgacaatcag cacacctgca t	21
	<210> 17	
	<211> 23	
	<212> ADN	
30	<213> Secuencia artificial	
	<220>	
	<223> Cebador-sonda PCR	
35	<400> 17 tgtgactaca gccgtgatcc tta	23
	<210> 18	
	<211> 20	
	<212> ADN	
40	<213> Secuencia artificial	

	<220>	
	<223> Cebador-sonda PCR	
5	<400> 18 caggccctct tccgagcgg	20
	<210> 19	
	<211> 21	
	<212> ADN	
10	<213> Secuencia artificial	
	<220>	
	<223> Cebador-sonda PCR	
15	<400> 19 tgtctcactg agcgagcaga a	21
	<210> 20	
	<211> 19	
	<212> ADN	
20	<213> Secuencia artificial	
	<220>	
	<223> Cebador-sonda PCR	
25	<400> 20 accattgcag ccctgattg	19
	<210> 21	
	<211> 24	
	<212> ADN	
<213> Secuencia artificial		
30	<220>	
	<223> Cebador-sonda PCR	
35	<400> 21 cttgaggacg cgaacagtcc acca	24
	<210> 22	
	<211> 19	
	<212> ADN	
	<213> Secuencia artificial	

<220>

<223> Cebador-sonda PCR

<400> 22
cgtggtgccc ctctatgac

19

5

<210> 23

<211> 19

<212> ADN

<213> Secuencia artificial

10

<220>

<223> Cebador-sonda PCR

<400> 23
ggctagtggg cgcatgtag

19

15

<210> 24

<211> 19

<212> ADN

<213> Secuencia artificial

20

<220>

<223> Cebador-sonda PCR

<400> 24
ctggagatgc tggacgccc

19

25

<210> 25

<211> 20

<212> ADN

<213> Secuencia artificial

30

<220>

<223> Cebador-sonda PCR

<400> 25
attccaccca tggcaaattc

20

35

<210> 26

<211> 22

<212> ADN

<213> Secuencia artificial

	<220>	
	<223> Cebador-sonda PCR	
5	<400> 26 gatgggattt ccattgatga ca	22
	<210> 27	
	<211> 22	
	<212> ADN	
10	<213> Secuencia artificial	
	<220>	
	<223> Cebador-sonda PCR	
15	<400> 27 ccgttctcag ccttgacggt gc	22
	<210> 28	
	<211> 20	
	<212> ADN	
	<213> Secuencia artificial	
20	<220>	
	<223> Cebador-sonda PCR	
25	<400> 28 ccatctgcat ccatcttggt	20
	<210> 29	
	<211> 20	
	<212> ADN	
	<213> Secuencia artificial	
30	<220>	
	<223> Cebador-sonda PCR	
35	<400> 29 ggccaccagg gtattatctg	20
	<210> 30	
	<211> 23	
	<212> ADN	
	<213> Secuencia artificial	

<220>

<223> Cebador-sonda PCR

5 <400> 30
ctccccaccc ttgagaagtg cct 23

<210> 31

<211> 20

<212> ADN

10 <213> Secuencia artificial

<220>

<223> Cebador-sonda PCR

15 <400> 31
ggcccagctt gaatttttca 20

<210> 32

<211> 27

<212> ADN

20 <213> Secuencia artificial

<220>

<223> Cebador-sonda PCR

25 <400> 32
aagctatgag gaaaagaagt acacgat 27
<210> 33

<211> 30

<212> ADN

<213> Secuencia artificial

30

<220>

<223> Cebador-sonda PCR

35 <400> 33
tcagccactg gcttctgtca taatcaggag 30

<210> 34

<211> 20

<212> ADN

<213> Secuencia artificial

<220>
 <223> Cebador-sonda PCR

5 <400> 34 20
 cccactcagt agccaagtca

 <210> 35
 <211> 20
 <212> ADN
 <213> Secuencia artificial

10 <220>
 <22B> Cebador-sonda PCR

 <400> 35 20
 cacgcaggtg gtatcagtct

15 <210> 36
 <211> 27
 <212> ADN
 <213> Secuencia artificial

20 <220>
 <223> Cebador-sonda PCR

 <400> 36 27
 tcaagtaaac gggctgtttt ccaaaca

 <210> 37
 <211> 20

25 <212> ADN
 <213> Secuencia artificial

 <220>
 <223> Cebador-sonda PCR

30 <400> 37 20
 cggtgtgaga agtgcagcaa

 <210> 38
 <211> 19

35 <212> ADN
 <213> Secuencia artificial

 <220>

ES 2 651 849 T3

<223> Cebador-sonda PCR

<400> 38
cctctcgcaa gtgctccat 19

5 <210> 39
<211> 24
<212> ADN
<213> Secuencia artificial

10 <220>
<223> Cebador-sonda PCR

<400> 39
ccagaccata gcacactcgg gcac 24

15 <210> 40
<211> 19
<212> ADN
<213> Secuencia artificial

20 <220>
<223> Cebador-sonda PCR

<400> 40
cggactttgg gtgcgactt 19

<210> 41
25 <211> 24
<212> ADN
<213> Secuencia artificial

<220>
30 <223> Cebador-sonda PCR

<400> 41
ttacaactct tccactggga cgat 24

<210> 42
35 <211> 23
<212> ADN
<213> Secuencia artificial

<220>

ES 2 651 849 T3

<223> Cebador-sonda PCR

<400> 42
ccacttgatcg aaccaccgct cgt
<210> 43

23

5 <211> 18
<212> ADN
<213> Secuencia artificial

<220>

10 <223> Cebador-sonda PCR

<400> 43
gccgagatcg ccaagatg

18

15 <210> 44
<211> 27
<212> ADN
<213> Secuencia artificial

<220>

20 <223> Cebador-sonda PCR

<400> 44
cttttgatgg tagagttcca gtgattc

27

25 <210> 45
<211> 24
<212> ADN
<213> Secuencia artificial

<220>

30 <223> Cebador-sonda PCR

<400> 45
cagcattgac tgcctccct ggca

24

35 <210> 46
<211> 20
<212> ADN
<213> Secuencia artificial

<220>

<223> Cebador-sonda PCR

<400> 46
gcatcaggct gtcattatgg 20

5 <210> 47
<211> 20
<212> ADN
<213> Secuencia artificial

10 <220>
<223> Cebador-sonda PCR

<400> 47
agtagttgtg ctgcccttcc 20

15 <210> 48
<211> 28
<212> ADN
<213> Secuencia artificial

20 <220>
<223> Cebador-sonda PCR

<400> 48
tgtccttacc tgtgggagct gtaaggtc 28

<210> 49
25 <211> 24
<212> ADN
<213> Secuencia artificial

<220>
30 <223> Cebador-sonda PCR

<400> 49
ccattctatc atcaacgggt acaa 24

<210> 50
<211> 23
35 <212> ADN
<213> Secuencia artificial

<220>

<223> Cebador-sonda PCR

<400> 50
tcaqcaadtq ggaaggtgta atc 23

5 <210> 51
<211> 25
<212> ADN
<213> Secuencia artificial

10 <220>
<223> Cebador-sonda PCR

<400> 51
tctccacaga caaggccagg actcg 25

15 <210> 52
<211> 20
<212> ADN
<213> Secuencia artificial

20 <220>
<223> Cebador-sonda PCR

<400> 52
catcttccag gaggaccact 20

<210> 53
25 <211> 20
<212> ADN
<213> Secuencia artificial
<220>
<223> Cebador-sonda PCR

30 <400> 53
tccgaccttc aatcatttca 20

<210> 54
<211> 24
<212> ADN
35 <213> Secuencia artificial

<220>
<223> Cebador-sonda PCR

	<400> 54 ctctgtggca ccctggacta cctg	24
	<210> 55 <211> 20	
5	<212> ADN <213> Secuencia artificial	
	<220> <223> Cebador-sonda PCR	
10	<400> 55 cctggaggct gcaacatacc	20
	<210> 56 <211> 23	
15	<212> ADN <213> Secuencia artificial	
	<220> <223> Cebador-sonda PCR	
20	<400> 56 tacaatggct ttggaggata gca	23
	<210> 57 <211> 25 <212> ADN	
25	<213> Secuencia artificial	
	<220> <223> Cebador-sonda PCR	
	<400> 57 atcctcctga agcccttttc gcagc	25
30	<210> 58 <211> 20 <212> ADN	
	<213> Secuencia artificial	
35	<220> <223> Cebador-sonda PCR	

	<400> 58 tgttttgatt cccgggctta	20
	<210> 59	
	<211> 24	
	<212> ADN	
5	<213> Secuencia artificial	
	<220>	
	<22B> Cebador-sonda PCR	
10	<400> 59 caaagctgtc agctctagca aaag	24
	<210> 60	
	<211> 28	
	<212> ADN	
	<213> Secuencia artificial	
15	<220>	
	<223> Cebador-sonda PCR	
20	<400> 60 tgcctttcttc ctccctcact tctcacct	28
	<210> 61	
	<211> 20	
	<212> ADN	
	<213> Secuencia artificial	
25	<220>	
	<223> Cebador-sonda PCR	
30	<400> 61 gccaaactgct ttcatttgtg	20
	<210> 62	
	<211> 20	
	<212> ADN	
	<213> Secuencia artificial	
35	<220>	
	<223> Cebador-sonda PCR	

<400> 62
 actcaggccc atttccttta 20
 <210> 63
 <211> 28
 <212> ADN
 5 <213> Secuencia artificial

 <220>
 <223> Cebador-sonda PCR

 <400> 63
 10 agggatctga accaatacag agcagaca 28

 <210> 64
 <211> 66
 <212> ADN
 15 <213> Secuencia artificial

 <220>
 <223> Amplificación por PCR

 <400> 64
 20 cagcagatgt ggatcagcaa gcaggagtat gacgagtccg gccctccat cgtccaccgc 60
 aaatgc 66

 <210> 65
 <211> 81
 <212> ADN
 25 <213> Secuencia artificial

 <220>
 <223> Amplificación por PCR

 <400> 65
 30 cgttgtcagc acttggata caagatggtt gccgggtcat gttaattggg aaaaagaaca 60
 gtccacagga agaggttgaa c 81

 <210> 66
 <211> 73
 <212> ADN
 35 <213> Secuencia artificial

 <220>
 <223> Amplificación por PCR

<400> 66
cagatggacc tagtaccac tgagatttcc acgccgaagg acagcgatgg gaaaaatgcc 60
cttaaatacat agg 73

<210> 67

5 <211> 84

<212> ADN

<213> Secuencia artificial

<220>

10 <223> Amplificación por PCR

<400> 67
ttcaggttgt tgcaggagac catgtacatg actgtctcca ttattgatcg gttcatgcag 60
aataattgtg tgcccaagaa gatg 84

<210> 68

15 <211> 74

<212> ADN

<213> Secuencia artificial

<220>

20 <223> Amplificación por PCR

<400> 68
tggttcccag ccctgtgtcc acctccaagc ccagattcag attcgagtca tgtacacaac 60
ccaggggtgga ggag 74

25 <210> 69

<211> 77

<212> ADN

<213> Secuencia artificial

30 <220>

<223> Amplificación por PCR

<400> 69
tgacaatcag cacacctgca ttcaccgctc ggaagagggc ctgagctgca tgaataagga 60
tcacggctgt agtcaca 77

35

<210> 70

<211> 67

ES 2 651 849 T3

<212> ADN

<213> Secuencia artificial

<220>

5 <223> Amplificación por PCR

<400> 70
tgtctcactg agcgagcaga atctggtgga ctgttcgcgt cctcaaggca atcagggctg 60
caatggt 67

10 <210> 71

<211> 68

<212> ADN

<213> Secuencia artificial

15 <220>

<223> Amplificación por PCR

<400> 71
cgtggtgccc ctctatgacc tgctgctgga gatgctggac gccaccgcc tacatgcgcc 60
cactagcc 68

20 <210> 72

<211> 74

<212> ADN

<213> Secuencia artificial

25 <220>

<223> Amplificación por PCR

<400> 72
attccaccca tggcaaattc catggcaccg tcaaggctga gaacgggaag cttgtcatca 60
atggaatcc catc 74

<210> 73

30 <211> 67

<212> ADN

<213> Secuencia artificial

<220>

35 <223> Amplificación por PCR

ES 2 651 849 T3

<400> 73
ccatctgcat ccatcttggt tgggctcccc acccttgaga agtgcctcag ataataccct 60
ggtggcc 67

<210> 74

<211> 86

<212> ADN

5 <213> Secuencia artificial

<220>

<223> Amplificación por PCR

<400> 74
aagctatgag gaaaagaagt acacgatggg ggacgctcct gattatgaca gaagccagtg 60
gctgaatgaa aaattcaagc tgggccc 86

10

<210> 75

<211> 73

<212> ADN

15 <213> Secuencia artificial

<220>

<223> Amplificación por PCR

<400> 75
cccactcagt agccaagtca caatgtttgg aaaacagccc gtttacttga gcaagactga 60
taccacctgc gtg 73

20

<210> 76

<211> 70

<212> ADN

<213> Secuencia artificial

25

<220>

<223> Amplificación por PCR

<400> 76
cgggtgtgaga agtgcagcaa gccctgtgcc cgagtgtgct atggtctggg catggagcac 60
ttgcgagagg 70

30

<210> 77

<211> 80

<212> ADN

<213> Secuencia artificial

<220>

35 <223> Amplificación por PCR

ES 2 651 849 T3

<400> 77
cggacttttg gtgcgacttg acgagcgggt gttcgacaag tggccttgcg ggccggatcg 60
tcccagtgga agagttgtaa 80

<210> 78

<211> 74

<212> ADN

5 <213> Secuencia artificial

<220>

<223> Amplificación por PCR

<400> 78
gccgagatcg ccaagatggt gccagggagg acagacaatg ctgtgaagaa tcaactggaac 60
tctaccatca aaag 74

10

<210> 79

<211> 85

<212> ADN

<213> Secuencia artificial

15

<220>

<223> Amplificación por PCR

<400> 79
gcatcaggct gtcattatgg tgtccttacc tgtgggagct gtaaggcttt cttaagagg 60
gcaatggaag ggcagcacia ctact 85

20

<210> 80

<211> 75

<212> ADN

<213> Secuencia artificial

25

<220>

<223> Amplificación por PCR

<400> 80
ccattctatc atcaacgggt acaaacgagt cctggccttg tctgtggaga cggattacac 60
cttcccactt gctga 75

30

<210> 81

<211> 69

<212> ADN

<213> Secuencia artificial

<220>

35

<223> Amplificación por PCR

ES 2 651 849 T3

<400> 81
catcttccag gaggaccact ctctgtggca ccctggacta cctgccccct gaaatgattg 60
aaggtcggg 69

<210> 82

<211> 90

5 <212> ADN

<213> Secuencia artificial

<220>

<223> Amplificación por PCR

10

<400> 82
cctggaggct gcaacatacc tcaatcctgt ccagggccgg atcctcctga agcccttttc 60
gcagcactgc tctcctccaa agccattgta 90

<210> 83

<211> 80

15 <212> ADN

<213> Secuencia artificial

<220>

<223> Amplificación por PCR

20

<400> 83
tgttttgatt cccgggctta ccaggtgaga agtgaggag gaagaaggca gtgtcccttt 60
tgctagagct gacagctttg 80

<210> 84

<211> 68

<212> ADN

25 <213> Secuencia artificial

<220>

<223> Amplificación por PCR

<400> 84
gccaaactgct ttcatttgtg agggatctga accaatacag agcagacata aaggaaatgg 60
gcctgagt 68

30

35

REIVINDICACIONES

1. Un método para determinar la probabilidad de recidiva del cáncer de mama en un paciente humano con cáncer de mama positivo en receptores de estrógenos y negativo en nódulo linfático, tratado con tamoxifeno, donde el método consiste en lo siguiente:
- 5 a) determinar los niveles de expresión de las transcripciones de ARN de GRB7, HER2, ER, PR, Bcl2, CEGP1, SURV, Ki.67, MYBL2, CCNB1, STK15, CTSL2, STMY3, CD68, GSTM1, y BAG1, o sus productos de expresión, en una muestra biológica que contiene células tumorales obtenidas del paciente;
- b) crear los siguientes subconjuntos de genes:
- (a) subconjunto del factor de crecimiento: GRB7 y HER2;
- 10 (b) subconjunto del receptor de estrógenos: ER, PR, Bcl2, y CEGP1;
- (c) subconjunto de proliferación: SURV, Ki.67, MYBL2, CCNB1, y STK15; y
- (d) subconjunto de invasión: CTSL2, y STMY3;
- c) calcular una puntuación de recidiva (RS) para el paciente a través de la ecuación siguiente:
- 15
$$RS = (0,23 \text{ a } 0,70) \times \text{umbrajeGRB7} - (0,17 \text{ a } 0,55) \times \text{ejeER} + (0,52 \text{ a } 1,56) \times \text{umbrajejepro lif} + (0,07 \text{ a } 0,21) \times \text{ejeinvas} + (0,03 \text{ a } 0,15) \times \text{CD68} - (0,04 \text{ a } 0,25) \times \text{GSTM1} - (0,05 \text{ a } 0,22) \times \text{BAG1}$$
- donde
- i) $\text{eje GRB7} = (0,45 \text{ a } 1,35) \times \text{GRB7} + (0,05 \text{ a } 0,15) \times \text{HER2}$;
- ii) si el eje GRB7 < -2, entonces el umbral del eje GRB7 = -2, y
- si el eje GRB7 > -2, entonces el umbral del eje GRB7 = eje GRB7;
- 20 iii) $\text{eje ER} = (\text{ER} + \text{PR} + \text{Bcl2} + \text{CEGP1})/4$;
- iv) $\text{ejejepro lif} = (\text{SURV} + \text{Ki. 67} + \text{MYBL2} + \text{CCNB1} + \text{STK15})/5$;
- v) si $\text{ejejepro lif} < -3,5$, entonces $\text{umbrajejepro lif} = -3,5$, si $\text{ejejepro lif} > -3,5$, entonces $\text{umbrajejepro lif} = \text{ejejepro lif}$; y
- vi) $\text{ejeinvas} = (\text{CTSL2} + \text{STMY3})/2$,
- 25 donde las contribuciones individuales de los genes de (3), (4) y (6) están ponderadas por un factor de 0,5 a 1,5, y
- d) crear un informe que proporciona la RS, donde una RS más elevada indica una mayor probabilidad de recidiva del cáncer de mama en dicho paciente.
- 30 2. Un método como el reivindicado en la reivindicación 1, donde la expresión de los genes GRB7, HER2, ER, PR, CEGP1, BCL2, SURV, KI67, MYBL2, CCNB1, STK15, CTSL2, STMY3, CD68, GSTM1, y BAG1 se mide determinando los niveles de expresión proteica.
3. Un método como el reivindicado en la reivindicación 1, donde el nivel aumentado de la transcripción de ARN de CD68, o su producto de expresión, está asociado con un mayor riesgo de recidiva del cáncer de mama y tiene asignado un valor positivo.
- 35 4. Un método como el reivindicado en la reivindicación 1, donde los niveles aumentados de las transcripciones de ARN de GSTM1 y BAG1, o sus productos de expresión, están asociados con un menor riesgo de recidiva del cáncer de mama y tienen asignado un valor negativo.
5. Un método como el reivindicado en la reivindicación 1, donde los niveles aumentados de las transcripciones de ARN de los genes de los subconjuntos génicos (i), (iii) y (iv), o sus productos de
- 40 expresión, están asociados con un mayor riesgo de recidiva del cáncer de mama y tienen asignado un valor positivo.
6. Un método como el reivindicado en la reivindicación 1, donde los niveles aumentados de las transcripciones de ARN de los genes del subconjunto génico (ii), o sus productos de expresión, están asociados con un menor riesgo de recidiva del cáncer de mama y tienen asignado un valor negativo.
- 45 7. Un método como el reivindicado en la reivindicación 1, donde la muestra biológica es un tejido tumoral fresco.
8. Un método como el reivindicado en la reivindicación 1, donde la muestra biológica es un aspirado extraído con aguja fina.
9. Un método como el reivindicado en la reivindicación 1, donde la muestra biológica es líquido
- 50 peritoneal.
10. Un método como el reivindicado en la reivindicación 1, donde la muestra biológica es lavado ductal.

11. Un método como el reivindicado en la reivindicación 1, donde la muestra biológica es líquido pleural.
12. Un método como el reivindicado en la reivindicación 1, donde la muestra biológica es una muestra tumoral.
13. Un método como el reivindicado en la reivindicación 13, donde la muestra tumoral se obtiene de un tejido tumoral fijado y embebido en parafina.
14. Un método como el reivindicado en la reivindicación 14, donde el tejido se ha obtenido por biopsia.
15. Un método como el reivindicado en la reivindicación 14, donde los niveles de expresión de las transcripciones de ARN se determinan por la reacción en cadena de polimerasa transcriptasa inversa.
16. Un método como el reivindicado en la reivindicación 14, donde el ARN está fragmentado.
17. Un método como el reivindicado en la reivindicación 1, donde los niveles de expresión de las transcripciones de ARN están determinados por:
- a) proporcionar un polinucleótido complementario a una secuencia de ARN intrónica dentro del gen diana cuyo nivel de expresión se pretende determinar, donde la expresión de la secuencia de ARN intrónica está correlacionada con la expresión de una secuencia de ARNm exónica dentro de dicho gen;
- b) hibridar dicho polinucleótido con dicha secuencia de ARN intrónica para formar un complejo polinucleótido-ARN intrónico; y detectar el complejo polinucleótido-ARN intrónico.
18. Un método como el reivindicado en la reivindicación 1, que comprende también el paso de analizar los datos utilizando las curvas de supervivencia de Kaplan-Meier.
19. Un método como el reivindicado en la reivindicación 1, donde el informe incluye una recomendación de tratamiento para el paciente.
20. Un método para determinar una opción de tratamiento para un paciente humano con cáncer de mama positivo en receptores de estrógenos y negativo en nódulos linfáticos, tratado con tamoxifeno, que consiste en lo siguiente:
- a) determinar los niveles de expresión de las transcripciones de ARN de GRB7, HER2, EstR1, PR, Bcl2, CEGP1, SURV, Ki.67, MYBL2, CCNB1, STK15, CTSL2, STMY3, CD68, GSTM1, y BAG1, o sus productos de expresión, en una muestra biológica que contiene células tumorales obtenidas del sujeto;
- b) crear los siguientes subconjuntos multigénicos:
- i) GRB7 y HER2;
- ii) EstR1, PR, Bcl2, y CEGP1;
- iii) SURV, Ki.67, MYBL2, CCNB1, y STK15
- iv) CTSL2, y STMY3;
- c) calcular una puntuación de recidiva (RS) para el paciente a través de la ecuación siguiente:
- $$RS = (0,23 \text{ a } 0,70) \times \text{umbralejeGRB7} - (0,17 \text{ a } 0,55) \times \text{ejeER} + (0,52 \text{ a } 1,56) \times \text{umbralejeprolif} + (0,07 \text{ a } 0,21) \times \text{ejeinvas} + (0,03 \text{ a } 0,15) \times \text{CD68} - (0,04 \text{ a } 0,25) \times \text{GSTM1} - (0,05 \text{ a } 0,22) \times \text{BAG1}$$
- donde
- i) $\text{eje GRB7} = (0,45 \text{ a } 1,35) \times \text{GRB7} + (0,05 \text{ a } 0,15) \times \text{HER2}$;
- ii) si $\text{eje GRB7} < -2$, entonces $\text{umbral eje GRB7} = -2$, y si $\text{eje GRB7} > -2$, entonces $\text{umbral eje GRB7} = \text{eje GRB7}$;
- iii) $\text{eje ER} = (\text{ER} + \text{PR} + \text{Bcl2} + \text{CEGP1})/4$;
- iv) $\text{ejeprolif} = (\text{SURV} + \text{Ki. 67} + \text{MYBL2} + \text{CCNB1} + \text{STK15})/5$;
- v) si $\text{ejeprolif} < -3,5$, entonces $\text{umbralejeprolif} = -3,5$, si $\text{ejeprolif} > -3,5$, entonces $\text{umbralejeprolif} = \text{ejeprolif}$;
- vi) $\text{ejeinvas} = (\text{CTSL2} + \text{STMY3})/2$,
- donde las contribuciones individuales de los genes de (iii), (iv) y (vi) están ponderados por un factor de 0,5 a 1,5, y
- d) clasificar el sujeto en un grupo de bajo, medio o alto riesgo; y
- e) recomendar ausencia de tratamiento o terapia hormonal solamente o terapia hormonal más otros tratamientos menos tóxicos si el sujeto se encuentra en el grupo de bajo riesgo; recomendar tratamiento quimioterapéutico si el sujeto se encuentra en el grupo de riesgo medio o alto.
21. Un método como el reivindicado en la reivindicación 1, donde la puntuación de recidiva se calcula

ES 2 651 849 T3

utilizando la siguiente ecuación:

$$RS = 0,47 \times \text{umbralGRB7GT} - 0,34 \times \text{ejeER} + 1,04 \times \text{umbralejeprolif} + 0,14 \times \text{ejeinvas} + 0,11 \times \text{CD68} - 0,17 \times \text{GSTM1} - 0,15 \times \text{BAG1}$$

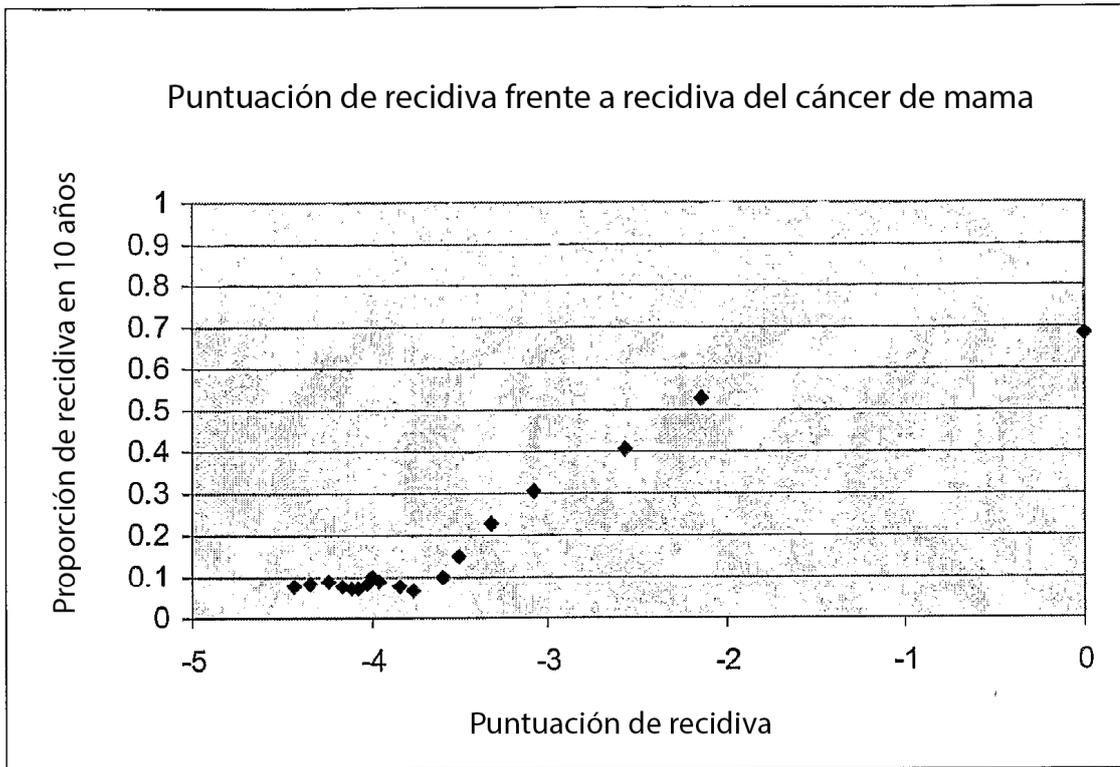


FIGURA 1A

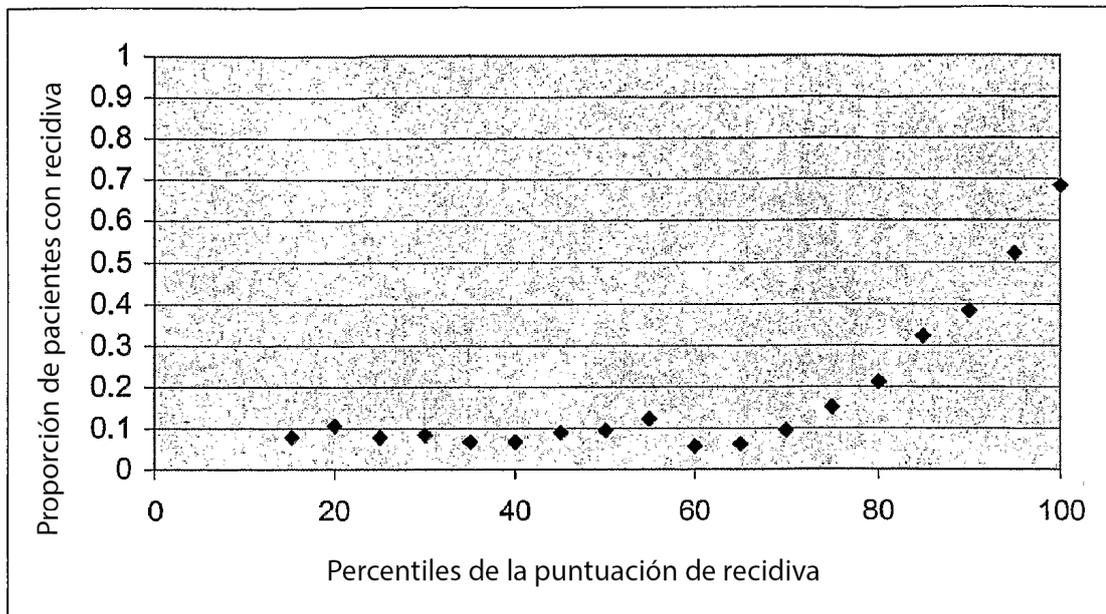


FIGURA 1B

FIGURA 2

