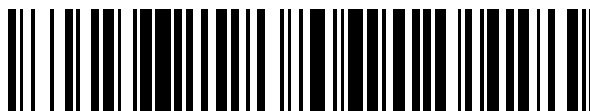


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 778 851**

51 Int. Cl.:

G16B 20/00 (2009.01)

C12Q 1/6886 (2008.01)

G16B 25/00 (2009.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **04.11.2005 PCT/US2005/040238**

87 Fecha y número de publicación internacional: **18.05.2006 WO06052862**

96 Fecha de presentación y número de la solicitud europea: **04.11.2005 E 05817266 (9)**

97 Fecha y número de publicación de la concesión europea: **04.03.2020 EP 1836629**

54 Título: **Predicción de la respuesta a la quimioterapia mediante el uso de marcadores de expresión génica**

30 Prioridad:

05.11.2004 US 625485 P

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

12.08.2020

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, B.;
BRYANT, JOHN, L.;
PAIK, SOONMYUNG y
SHAK, STEVEN**

74 Agente/Representante:

ISERN JARA, Jorge

ES 2 778 851 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

Predicción de la respuesta a la quimioterapia mediante el uso de marcadores de expresión génica

5 Campo de la invención

La presente invención proporciona información de expresión génica útil para predecir si es probable que los pacientes con cáncer tengan una respuesta beneficiosa a la respuesta al tratamiento con quimioterapia.

10 Descripción de la técnica relacionadaEstudios de expresión génica

15 Los oncólogos tienen una serie de opciones de tratamiento disponibles, que incluyen diferentes combinaciones de fármacos quimioterapéuticos que se caracterizan como "estándar de referencia" y una cantidad de fármacos que no portan una declaración en la etiqueta para el tratamiento de un cáncer en particular, pero para los cuales hay evidencia de eficacia en ese cáncer. La mejor probabilidad de un buen resultado del tratamiento requiere que los pacientes con mayor riesgo de enfermedad metastásica sean identificados y asignados al tratamiento óptimo disponible contra el cáncer. En particular, es importante determinar la probabilidad de respuesta del paciente a los fármacos terapéuticos

20 "estándar de referencia", como ciclofosfamida, metotrexato, 5-fluorouracilo, antraciclinas, taxanos y fármacos antiestrógenos, tal como el tamoxifeno, porque estos tienen una eficacia limitada y un espectro de efectos secundarios frecuentemente graves. La identificación de los pacientes con mayor o menor probabilidad de necesitar y responder a los fármacos disponibles podría aumentar el beneficio neto que estos fármacos tienen para ofrecer y disminuir la morbilidad y toxicidad netas, a través de una selección de pacientes más inteligente.

25 Actualmente, las pruebas de diagnóstico usadas en la práctica clínica son de un solo analito y, por lo tanto, no capturan el valor potencial de conocer las relaciones entre docenas de marcadores diferentes. Además, las pruebas de diagnóstico frecuentemente se basan en inmunohistoquímica, la cual no es cuantitativa. La inmunohistoquímica frecuentemente produce resultados diferentes en diferentes laboratorios, principalmente porque las interpretaciones son subjetivas. Las pruebas basadas en ARN, aunque son potencialmente altamente cuantitativas, no se han desarrollado debido a la percepción de que el ARN se destruye en muestras de tumor según se prepara habitualmente, es decir, se fijan en formalina y se incrustan en parafina (FPE), y porque es inconveniente obtener y almacenar muestras de tejido fresco de pacientes para análisis.

35 En las últimas dos décadas, la biología molecular y la bioquímica han revelado cientos de genes cuyas actividades influyen en el comportamiento de las células tumorales, su estado de diferenciación y su sensibilidad o resistencia a ciertos fármacos terapéuticos. Sin embargo, con algunas excepciones, el estado de estos genes no se ha explotado para el propósito de tomar decisiones clínicas habituales sobre tratamientos farmacológicos. En los últimos años, varios grupos han publicado estudios con respecto a la clasificación de varios tipos de cáncer mediante análisis de expresión

40 génica por micromatrices de miles de genes (véase, por ejemplo, Golub y otros, Science 286:531-537 (1999); Bhattacharjee y otros, Proc. Natl. Acad. Sci. USA 98:13790-13795 (2001); Chen-Hsiang y otros, Bioinformática 17 (Supl. 1):S316-S322 (2001); Ramaswamy y otros, Proc. Natl. Acad. Sci. USA 98:15149-15154 (2001); Martin y otros, Cancer Res. 60:2232-2238 (2000); West y otros, Proc. Natl. Acad. Sci. USA 98:11462-114 (2001); Sorlie y otros, Proc. Natl. Acad. Sci. USA 98:10869-10874 (2001); Yan y otros, Cancer Res. 61:8375-8380 (2001)). Sin embargo, estos estudios

45 aún no han arrojado pruebas de uso habitual en la práctica clínica, en gran parte debido a que los micromatrices requieren ARN de tejido fresco o congelado y tales muestras no están presentes en cantidad suficiente para permitir la validación clínica de las firmas moleculares identificadas.

50 En los últimos tres años, ha sido posible perfilar la expresión génica de cientos de genes en tejido embebido en parafina fijado con formalina (FPE) mediante el uso de la tecnología de RT-PCR. Se han descrito procedimientos que son sensibles, precisos y reproducibles (Cronin y otros, Am. J. Pathol. 164:35-42 (2004)). Debido a que existen miles de muestras de tejido clínico de FPE archivadas con registros clínicos asociados, tales como supervivencia, historial de tratamiento farmacológico, etc., la capacidad de analizar cuantitativamente la expresión génica en este tipo de tejido permite estudios clínicos rápidos que relacionan la expresión de ciertos genes con el pronóstico del paciente y la

55 probabilidad de respuesta a los tratamientos. El uso de datos generados mediante estudios clínicos anteriores permite obtener resultados rápidos porque los eventos clínicos son históricos. Por el contrario, por ejemplo, si se desea llevar a cabo un estudio de supervivencia en pacientes con cáncer recién reclutados, generalmente necesita esperarse durante muchos años para que se produzca un número de muertes estadísticamente suficiente.

60 Cáncer de mama

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

65 Actualmente, solo unas pocas pruebas moleculares se usan de manera habitual, clínicamente, en el cáncer de mama. Los ensayos inmunohistoquímicos para las proteínas receptor de estrógeno (ESR1) y receptor de progesterona (PGR)

se usan como base para la selección de pacientes para el tratamiento con fármacos antiestrógenos, tal como el tamoxifeno (TAM). Además, la inmunoquímica de ERBB2 (Her2) o la hibridación fluorescente in situ (que miden proteínas y ADN, respectivamente) se usan para seleccionar pacientes con fármacos antagonistas de Her2, tal como trastuzumab (Herceptin®; Genentech, Inc., South San Francisco, CA).

Debido a que las pruebas actuales para el pronóstico y la predicción de la respuesta a la quimioterapia son inadecuadas, las estrategias de tratamiento del cáncer de mama varían entre los oncólogos (Schott y Hayes, J. Clin. Oncol. PMID 15505274 (2004); Hayes, Breast 12: 543-9 (2003)). En general, los pacientes con ganglios linfáticos negativos cuyos tumores se encuentran positivos para ESR1 son tratados con un fármaco antiestrógeno, tal como TAM, y los pacientes cuyos tumores se encuentran negativos para ESR1 son tratados con quimioterapia. Frecuentemente, para los ESR1 positivos también se prescribe quimioterapia además de la terapia antiestrógenos, aceptando los efectos secundarios tóxicos de la quimioterapia para disminuir modestamente el riesgo de recurrencia del cáncer. Las toxicidades incluyen neuropatía, náuseas y otros síntomas gastrointestinales, pérdida de cabello y deterioro cognitivo. Se debe temer a la recurrencia porque el cáncer de mama recurrente generalmente es metastásico y responde poco al tratamiento. Claramente, existe la necesidad de identificar a aquellos pacientes que tienen un riesgo sustancial de recurrencia (es decir, proporcionar información de *pronóstico*) y si es probable que respondan a la quimioterapia (es decir, para proporcionar *información predictiva*). Del mismo modo, existe la necesidad de identificar a aquellos pacientes que no tienen un riesgo significativo de recurrencia, o que es poco probable que respondan a la quimioterapia, ya que estos pacientes deben evitar la exposición innecesaria a esos fármacos tóxicos.

Los factores pronósticos difieren de los factores predictivos del tratamiento en el cáncer de mama. Los factores pronósticos son aquellas variables relacionadas con la historia natural del cáncer de mama, que influyen en los rangos de recurrencia y el resultado de los pacientes una vez que han desarrollado cáncer de mama. Los parámetros clínicos que se han asociado con un peor pronóstico incluyen, por ejemplo, compromiso de los ganglios linfáticos, aumento del tamaño del tumor y tumores de alto grado. Los factores pronósticos se usan con frecuencia para clasificar a los pacientes en subgrupos con diferentes riesgos iniciales de recaída. Por el contrario, los factores predictivos del tratamiento son variables relacionadas con la probabilidad de una respuesta beneficiosa de un paciente individual a un tratamiento, como antiestrógenos o quimioterapia, independientemente del pronóstico.

El documento WO-A-03078662 divulga un procedimiento para predecir la probabilidad de una respuesta beneficiosa a la quimioterapia de un sujeto diagnosticado con cáncer.

Existe una gran necesidad de pruebas cuantitativas y precisas que predigan de manera confiable la probabilidad de que un paciente con cáncer, tal como un paciente con cáncer de mama, reciba cierto tipo de tratamiento. Tales pruebas ayudarían al médico a tomar decisiones de tratamiento inteligentes, adaptadas a las necesidades de un paciente en particular, basadas en un análisis de riesgo-beneficio bien fundado.

Breve descripción de las figuras

La Figura 1 muestra el beneficio absoluto de la quimioterapia como se determina por DRFS a los 10 años dentro de los grupos de pacientes del estudio 20 de NSABP identificados por la Puntuación de Recurrencia como de riesgo bajo, intermedio o alto.

La Figura 2 muestra el beneficio absoluto de la quimioterapia como se determina por DRFS a los 10 años dentro de los grupos de pacientes del estudio 20 de NSABP identificados por la Puntuación de Recurrencia como una variable continua.

Sumario de la invención

La invención se define en las reivindicaciones.

En un aspecto, la invención se refiere a un procedimiento para predecir la probabilidad de una respuesta beneficiosa a la quimioterapia de un sujeto diagnosticado con cáncer, que comprende

(a) determinar cuantitativamente, en una muestra biológica que comprende células cancerosas obtenidas de dicho sujeto, el valor de una o más de las siguientes variables:

- (i) Puntuación de Recurrencia,
- (ii) Puntuación del Grupo ESR1;
- (iii) Puntuación del Grupo de invasión;
- (iv) Puntuación Umbral del Grupo de Proliferación; y
- (v) El nivel de expresión del transcrito de ARN de al menos uno de MYBL2 y SCUBE2, o el producto de expresión correspondiente,

en donde

(b1) por cada unidad de un aumento en el valor de uno o más de (i), (iii), (iv), o el nivel de expresión del transcrito de ARN de MYBL2, o el producto de expresión correspondiente, se identifica que dicho sujeto tiene una probabilidad proporcionalmente aumentada de una respuesta beneficiosa a dicha quimioterapia; y

(b2) por cada unidad de un aumento en el valor de (ii) o el nivel de expresión del transcrito de ARN de SCUBE2, o el producto de expresión correspondiente, se identifica que dicho sujeto tiene una probabilidad proporcionalmente disminuida de una respuesta beneficiosa a la quimioterapia;

en donde

$$\text{Puntuación del Grupo ESR1} = (\text{ESR1} + \text{PGR} + \text{BCL2} + \text{SCUBE2})/4$$

$$\text{Puntuación del Grupo de Invasión} = (\text{CTSL2} + \text{MMP11})/2$$

$$\text{Puntuación del Grupo GRB7} = 0,9 \times \text{GRB7} + 0,1 \times \text{ERBB2};$$

La Puntuación Umbral del Grupo GRB7 es igual a 8 si la Puntuación del Grupo GRB7 es menor que 8 e igual a la Puntuación del Grupo GRB7 si la Puntuación del Grupo GRB7 es 8 o más

$$\text{Puntuación del Grupo de Proliferación} = (\text{BIRC5} + \text{MK167} + \text{MYBL2} + \text{CCNB1} + \text{STK6})/5;$$

La Puntuación Umbral del Grupo de Proliferación es igual a 6,5, si la Puntuación del Grupo de Proliferación es menor a 6,5; e igual a la puntuación del grupo de proliferación, si la puntuación del grupo de proliferación es 6,5 o más, y

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

en donde

$$RS_u = \begin{aligned} &0,47 \times \text{Puntuación Umbral del Grupo GRB7} \\ &- 0,34 \times \text{Puntuación del Grupo ESR1} \\ &+ 1,04 \times \text{Puntuación Umbral del Grupo de Proliferación} \\ &+ 0,10 \times \text{Puntuación del Grupo de Invasión} \\ &+ 0,05 \times \text{CD68} \\ &- 0,08 \times \text{GSTM1} \\ &- 0,07 \times \text{BAG1} \end{aligned}$$

donde los símbolos de genes en las ecuaciones representan los niveles de expresión de los transcritos de ARN de los genes respectivos o sus productos de expresión.

En otro caso, el sujeto es un paciente humano.

Los niveles de expresión de todos los genes incluidos en las variables (i) - (v), o sus productos de expresión, se normalizan en relación con los niveles de expresión de uno o más genes de referencia, o sus productos de expresión. Por ejemplo, los genes de referencia pueden seleccionarse del grupo que consiste en ACTB, GAPD, GUSB, RPLPO y TFRC. En otra realización, los niveles de expresión se normalizan en relación con la media de los niveles de expresión de ACTB, GAPD, GUSB, RPLPO y TFRC, o sus productos de expresión.

En una realización adicional, el valor cuantitativo de la probabilidad de una respuesta beneficiosa a la quimioterapia es directamente proporcional al valor de la variable o variables determinadas en un continuo.

El cáncer es cáncer de mama. El cáncer de mama incluye, sin limitación, cáncer de mama invasivo o cáncer de mama en estadio II o estadio III y cáncer de mama positivo para ESR1.

Cuando se determina que el paciente tiene una mayor probabilidad de una respuesta beneficiosa a la quimioterapia, un procedimiento divulgado incluye además una etapa para tratar al paciente con quimioterapia. La quimioterapia puede ser quimioterapia adyuvante o neoadyuvante e incluye la administración de cualquier fármaco quimioterapéutico que haya demostrado ser eficaz para el tratamiento del cáncer en particular. Por lo tanto, los fármacos quimioterapéuticos incluyen derivados de antraciclina, tales como doxorrubicina o adriamicina; derivados de taxano, tales como paclitaxel o

docetaxel; inhibidores de topoisomerasa, tales como camptotecina, toptecan, irinotecan, 20-S-camptotecina, 9-nitro-camptotecina, 9-amino-camptotecina o GI147211; e inhibidores de la biosíntesis de nucleótidos, tales como metotrexato y/o 5-fluorouracilo (5-FU).

5 El procedimiento de la invención puede comprender la determinación de al menos dos, o al menos tres, o al menos cuatro, o cinco de las variables enumeradas.

En una realización particular, el procedimiento de la invención comprende la determinación del nivel de expresión de uno o ambos de MYBL2 y SCUBE2, o sus productos de expresión.

10 La muestra biológica puede ser, por ejemplo, una muestra de tejido que comprende células cancerosas.

La muestra de tejido puede, sin limitación, fijarse, embeberse en parafina, estar fresca o congelada, y puede derivarse, por ejemplo, de una aguja fina, núcleo u otros tipos de biopsia. En una realización particular, la muestra de tejido se obtiene por aspiración con aguja fina, lavado bronquial o biopsia transbronquial.

15 En una realización adicional, la determinación de los niveles de expresión incluye RT-PCR cuantitativa.

En una realización diferente, la determinación de los niveles de expresión del producto de expresión de los genes enumerados incluye inmunohistoquímica.

20 En una realización adicional, los niveles de los productos de expresión génica se determinan mediante técnicas proteómicas.

En otra realización adicional, los niveles de expresión de los genes se determinan por RT-PCR cuantitativa, mediante el uso de secuencias de cebador y sonda basadas en una secuencia de genes diana.

25 En una realización específica, al menos una secuencia de gen diana es una secuencia basada en un intrón, cuya expresión se correlaciona con la expresión de una secuencia del exón del mismo gen.

30 El procedimiento de la presente invención puede incluir una etapa de crear un informe que resuma dicha probabilidad de respuesta beneficiosa, y opcionalmente una etapa de proporcionar el informe a un paciente diagnosticado con cáncer y/o al médico del paciente como un perfil genómico personalizado.

35 En otro aspecto, se divulga un procedimiento de preparación de un perfil genómico personalizado para un sujeto diagnosticado con cáncer, que comprende

(a) determinar cuantitativamente, en una muestra biológica que comprende células cancerosas obtenidas de dicho sujeto, el valor de una o más de las siguientes variables:

- 40 (i) Puntuación de Recurrencia,
 (ii) Puntuación del Grupo ESR1;
 (iii) Puntuación del Grupo de Invasión;
 (iv) Puntuación Umbral de Grupo de Proliferación; y
 45 (v) El nivel de expresión de los transcritos de ARN de al menos uno de MYBL2 y SCUBE2,

en donde

(b1) por cada unidad de un aumento en el valor de uno o más de (i), (iii), (iv), o el nivel de expresión del transcrito de ARN de MYBL2, o el producto de expresión correspondiente, se identifica que dicho sujeto tiene una probabilidad proporcionalmente aumentada de una respuesta beneficiosa a dicha quimioterapia;

50 (b2) por cada unidad de un aumento en el valor de (ii) o el nivel de expresión del transcrito de ARN de SCUBE2, o el producto de expresión correspondiente, se identifica que dicho sujeto tiene una probabilidad proporcionalmente disminuida de una respuesta beneficiosa a la quimioterapia; y

en donde

55
$$\text{Puntuación del Grupo ESR1} = (0,8 \times \text{ESR1} + 1,2 \times \text{PGR} + \text{BCL2} + \text{SCUBE2})/4$$

60
$$\text{Puntuación del Grupo de Invasión} = (\text{CTSL2} + \text{MMP11})/2$$

$$\text{Puntuación del Grupo GRB7} = 0,9 \times \text{GRB7} + 0,1 \times \text{ERBB2};$$

65 La Puntuación Umbral del Grupo GRB7 es igual a 8, si la Puntuación del Grupo GRB7 es menor que 8 e igual a la Puntuación del Grupo GRB7, si la Puntuación del Grupo GRB7 es 8 o más.

$$\text{Puntuación del Grupo de Proliferación} = (\text{BIRC5} + \text{MK167} + \text{MYBL2} + \text{CCNB1} + \text{STK6})/5;$$

La Puntuación Umbral del Grupo de proliferación es igual a 6,5, si la Puntuación del Grupo de Proliferación es menor a 6,5; y es idéntica a la Puntuación del Grupo de Proliferación, si la Puntuación del Grupo de Proliferación es 6,5 o más, y

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

en donde

$$RS_u = \begin{aligned} & 0,47 \times \text{Puntuación Umbral del Grupo GRB7} \\ & - 0,34 \times \text{Puntuación del Grupo ESR1} \\ & + 1,04 \times \text{Puntuación Umbral del Grupo de Proliferación} \\ & + 0,10 \times \text{Puntuación del Grupo de Invasión} \\ & + 0,05 \times \text{CD68} \\ & - 0,08 \times \text{GSTM1} \\ & - 0,07 \times \text{BAG1} \end{aligned}$$

donde los símbolos de genes en las ecuaciones representan los niveles de expresión de los transcritos de ARN de los genes respectivos, o sus productos de expresión, y las contribuciones individuales de los genes a las variables (i), (ii), (iii) y (iv) pueden ser ponderadas por un factor entre 0,5 a 1,5; y en donde cada gen individual o gen presente en cualquiera de dichas variables puede ser sustituido por otro gen que se coexpresa con dicho gen en dicho cáncer con un coeficiente de Pearson $\geq 0,5$; y (c) crear un informe que resuma los datos obtenidos por el análisis de expresión génica.

En una realización específica, si se determina un aumento en el valor de uno o más de (i), (iii), (iv), o el nivel de expresión del transcrito de ARN de MYBL2, o el producto de expresión correspondiente, el informe incluye una predicción de que el sujeto tiene una mayor probabilidad de una respuesta beneficiosa a la quimioterapia. En este caso, el procedimiento puede incluir además la etapa de tratar a dicho sujeto con un agente quimioterapéutico.

En otra realización más, si se determina un aumento en el valor de (ii) o el nivel de expresión del transcrito de ARN de SCUBE2, o el producto de expresión correspondiente, el informe incluye una predicción de que el sujeto tiene una probabilidad disminuida de una respuesta beneficiosa a la quimioterapia.

Descripción detallada de la realización preferente

A. Definiciones

A menos que se especifique de cualquier otra forma, los términos técnicos y científicos utilizados en la presente memoria tienen el mismo significado que el conocido comúnmente por aquellos con experiencia en la técnica a la que pertenece esta invención. Singleton y otros, Dictionary of Microbiology and Molecular Biology 2da edición, J. Wiley & Sons (Nueva York, NY 1994); y Webster's New World™ Medical Dictionary, 2da edición, Wiley Publishing Inc., 2003, proporcionan a un experto en la técnica una guía general de muchos de los términos usados en la presente solicitud. Para propósitos de la presente invención, se definen a continuación los siguientes términos.

El término "micromatriz" se refiere a una disposición ordenada de elementos de matriz hibridables, preferentemente sondas de polinucleótidos, en un sustrato.

El término "polinucleótido", cuando se usa en singular o plural, generalmente se refiere a cualquier polirribonucleótido o polidesoxirribonucleótido, que puede ser ARN o ADN no modificado o ARN o ADN modificado. Así, por ejemplo, los polinucleótidos como se definen en la presente memoria incluyen, sin limitación, ADN monocatenario y bicatenario, ADN que incluye regiones monocatenarias y bicatenarias, ARN monocatenario y bicatenario, y ARN que incluye regiones monocatenarias y bicatenarias, moléculas híbridas que comprenden ADN y ARN que pueden ser monocatenarias o, más típicamente, bicatenarias o incluyen regiones monocatenarias y bicatenarias. Además, el término "polinucleótido" como se usa en la presente memoria se refiere a regiones tricatenarias que comprenden ARN o ADN o tanto ARN como ADN. Las cadenas en tales regiones pueden ser de la misma molécula o de moléculas diferentes. Las regiones pueden incluir todas de una o más de las moléculas, pero más típicamente implican solo una región de algunas de las moléculas. Una de las moléculas de una región de triple hélice frecuentemente es un oligonucleótido. El término "polinucleótido" incluye específicamente los ADNc. El término incluye ADN (incluidos ADNc) y ARN que contienen una o más bases modificadas. Por lo tanto, las moléculas de ADN o ARN con cadenas principales modificadas para estabilidad o para otras razones son "polinucleótidos" como se pretende en la presente memoria. Además, las

5 moléculas de ADN o ARN que comprenden bases inusuales, como la inosina, o bases modificadas, como las bases tritriadas, se incluyen dentro del término "polinucleótidos" como se define en la presente memoria. En general, el término "polinucleótido" abarca todas las formas modificadas químicamente, enzimáticamente y/o metabólicamente de polinucleótidos no modificados, así como las formas químicas de ADN y ARN características de virus y células, incluidas las células simples y complejas.

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

15 El término "expresión génica" describe la conversión de la información de la secuencia génica del ADN en ARN transcrito (el transcrito de ARN no empalmado inicial o el ARNm maduro) o el producto proteico codificado. La expresión génica se puede controlar midiendo los niveles absolutos de ARN o de los productos proteicos del gen o sus subsecuencias.

20 El término "sobrexpresión" con respecto a un transcrito de ARN se usa para referirse al nivel del transcrito determinado por la normalización con respecto al nivel de ARNm de referencia, que podrían ser todos los transcritos medidos en la muestra o un conjunto de referencia particular de los ARNm.

25 La frase "amplificación de genes" se refiere a un procedimiento mediante el cual se forman múltiples copias de un gen o fragmento del gen en una célula o línea celular particular. La región duplicada (un tramo de ADN amplificado) frecuentemente se denomina "amplicón". Por lo general, la cantidad de ARN mensajero (ARNm) producido, *es decir*, el nivel de expresión génica, también aumenta en la proporción del número de copias realizadas del gen particular expresado.

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

35 El término "pronóstico" se usa en la presente memoria para referirse a la probabilidad de muerte atribuible al cáncer o a la progresión del cáncer, incluida la recurrencia y diseminación metastásica de una enfermedad neoplásica, como el cáncer de mama, durante la historia natural de la enfermedad. Los factores pronósticos son aquellas variables relacionadas con la historia natural de una enfermedad neoplásica, tal como el cáncer de mama, que influyen en las tasas de recurrencia y el resultado de la enfermedad una vez que el paciente desarrolló la enfermedad neoplásica, tal como el cáncer de mama. En este contexto, "resultado natural" significa resultado en ausencia de tratamiento adicional. Por ejemplo, en el caso del cáncer de mama, "resultado natural" significa el resultado después de la resección quirúrgica del tumor, en ausencia de tratamiento adicional (tal como, tratamiento de quimioterapia o radioterapia). Los factores pronósticos se usan con frecuencia para clasificar a los pacientes en subgrupos con diferentes riesgos basales, tales como los riesgos iniciales de recaída.

40 El término "predicción" se usa en la presente memoria para referirse a la probabilidad de que un paciente responda de manera favorable o desfavorable a un fármaco o conjunto de fármacos, y también al alcance de esas respuestas. Por lo tanto, los factores predictivos del tratamiento son aquellas variables relacionadas con la respuesta de un paciente individual a un tratamiento específico, independientemente del pronóstico. Los procedimientos predictivos de la presente invención pueden usarse clínicamente para tomar decisiones de tratamiento al elegir las modalidades de tratamiento más apropiadas para cualquier paciente en particular. Los procedimientos predictivos de la presente invención son instrumentos valiosos para predecir si es probable que un paciente responda favorablemente a un régimen de tratamiento, tal como la terapia antiestrógenos, tal como el tratamiento TAM solo o en combinación con quimioterapia y/o radioterapia.

45 El término "respuesta beneficiosa" significa una mejora en cualquier medida del estado del paciente, incluidas las medidas comúnmente usadas en la técnica, tales como la supervivencia general, la supervivencia a largo plazo, la supervivencia sin recurrencia y la supervivencia sin recurrencia distante. La supervivencia libre de recurrencia (RFS) se refiere al tiempo (en años) desde la cirugía hasta la primera recurrencia local, regional o distante. La supervivencia libre de recurrencia distante (DFRS) se refiere al tiempo (en años) desde la cirugía hasta la primera recurrencia distante. La recurrencia se refiere a RFS y/o DFRS como lo demuestra su uso particular. El cálculo de estas medidas en la práctica puede variar de un estudio a otro dependiendo de la definición de eventos que se censuren o no se consideren. El término supervivencia "a largo plazo" se usa en la presente memoria para referirse a la supervivencia durante al menos

3 años, con mayor preferencia durante al menos 8 años, con la máxima preferencia durante al menos 10 años después de la cirugía u otro tratamiento.

5 El término "tumor", como se usa en la presente memoria, se refiere a todo crecimiento y proliferación de células neoplásicas, ya sean malignas o benignas, y a todas las células y tejidos precancerosos y cancerosos.

10 Los términos "cáncer" y "canceroso" se refieren o describen la condición fisiológica en mamíferos que se caracteriza típicamente por un crecimiento celular desregulado. Los ejemplos de cáncer incluyen, pero no se limitan a, cáncer de mama, cáncer de ovario, cáncer de colon, cáncer de pulmón, cáncer de próstata, cáncer hepatocelular, cáncer gástrico, cáncer de páncreas, cáncer de cuello uterino, cáncer de hígado, cáncer de vejiga, cáncer del tracto urinario, cáncer de tiroides, cáncer renal, carcinoma, melanoma y cáncer de cerebro.

15 La "patología" del cáncer incluye todos los fenómenos que comprometen el bienestar del paciente. Esto incluye, sin limitación, crecimiento celular anormal o incontrolable, metástasis, interferencia con el funcionamiento normal de las células vecinas, liberación de citocinas u otros productos secretorios a niveles anormales, supresión o agravamiento de la respuesta inflamatoria o inmunológica, neoplasia, premalignidad, malignidad, invasión de tejidos u órganos circundantes o distantes, como los ganglios linfáticos, etc.

20 En el contexto de la presente invención, la referencia a "al menos uno", "al menos dos", "al menos tres", "al menos cuatro", "al menos cinco", etc. de los genes enumerados en cualquier conjunto de un gen particular significa cualquiera o cualquiera y todas las combinaciones de los genes enumerados.

25 El término cáncer "negativo en ganglios", tal como cáncer de mama "negativo en ganglios", se usa en la presente memoria para referirse al cáncer que no se ha diseminado a los ganglios linfáticos drenantes.

Los términos "empalme" y "empalme de ARN" se usan indistintamente y se refieren al procesamiento de ARN que elimina intrones y une exones para producir ARNm maduro con una secuencia de codificación continua, que se mueve hacia el citoplasma de una célula eucariota.

30 En teoría, el término "exón" se refiere a cualquier segmento de un gen interrumpido que está representado en el producto de ARN maduro (B. Lewin. Genes IV Cell Press, Cambridge Mass. 1990). En teoría, el término "intrón" se refiere a cualquier segmento de ADN que se transcribe pero se elimina del interior del transcrito al unir los exones a ambos lados del mismo. Operativamente, las secuencias de exón ocurren en la secuencia de ARNm de un gen como se define por los números SEQ ID de referencia. Operativamente, las secuencias de intrones son las secuencias que intervienen dentro del ADN genómico de un gen, entre corchetes por las secuencias de exones y que tienen secuencias de consenso de empalme GT y AG en sus límites 5' y 3'.

B. Descripción detallada

40 La práctica de la presente invención emplea, a menos que se indique lo contrario, técnicas convencionales de análisis estadístico, biología molecular (incluidas las técnicas recombinantes), microbiología, biología celular y bioquímica, que están dentro la experiencia de la técnica. Tales técnicas se explican completamente en la literatura, tales como, por ejemplo, "Molecular Cloning: A Laboratory Manual", 2ª edición (Sambrook y otros, 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 y M.P. Calos, eds., 1987); "Current Protocols in Molecular Biology" (F.M. Ausubel y otros, eds., 1987); "Statistical Methods and Scientific Inference", 3 ediciones (R. A. Fisher., 1956/59/74) y "PCR: The Polymerase Chain Reaction", (Mullis y otros, Eds., 1994).

50 B.1. Descripción general de la invención

55 En los últimos dos años, Genomic Health, Inc y colaboradores (Esteban y otros, Proc Am Soc Clin Oncol 22: página 850, 2003 (resumen 3416); Soule y otros, Proc Am Soc Clin Oncol 22: página 862, 2003 (resumen 3466); Cobleigh y otros Soc Clin Oncol 22: página 850, 2003 (resumen 3415); Cronin y otros, Am JPathol 164(1):35-42 (2004)) informaron varios estudios clínicos exploratorios de la expresión génica en el cáncer de mama temprano, con el objetivo de encontrar una firma molecular para el riesgo de recurrencia. Estos estudios usaron RT-PCR cuantitativa para evaluar 250 marcadores de genes candidatos en muestras de tejido congelado, embebidas en parafina (FPE) que tienen registros clínicos vinculados. Se realizó un análisis en los tres estudios para examinar si se podían identificar genes que estuvieran relacionados de manera consistente con el riesgo de recurrencia en un grupo diverso de pacientes. En base a estos resultados univariantes, se diseñaron y analizaron modelos de múltiples genes en los tres estudios. Se desarrolló un único ensayo de múltiples genes, que consta de 16 genes relacionados con el cáncer y 5 genes de referencia, para ser probado prospectivamente en estudios de validación clínica. Se generó un algoritmo llamado Puntuación de Recurrencia (RS), que utiliza las mediciones de estas 21 especies de ARNm e informa el riesgo de recurrencia en una escala de 100 puntos.

65 Para probar la validez clínica de esta prueba y el algoritmo de Puntuación de Recurrencia, se llevó a cabo un ensayo clínico a ciegas con criterios de valoración identificados prospectivamente. Este ensayo de validación se centró en

pacientes tratados con TAM solo en los brazos aleatorizados y de registro de la población del ensayo clínico del Estudio B-14 de NSABP (Fisher B, Costantino JP, Redmond CK y otros: Cáncer de endometrio en pacientes con cáncer de mama tratado: Hallazgos del estudio B-14 del Proyecto Nacional de Terapia Quirúrgica y Adyuvante de Mama e Intestino (NSABP). J Natl Cancer Inst 86:527-537 (1994)). Genomic Health, Inc. y NSABP llevaron a cabo el ensayo de RT-PCR de 21 genes en 668 muestras de tejido de cáncer de mama derivadas de estos pacientes y calcularon una Puntuación de Recurrencia para cada paciente.

Los puntos de corte preespecificados de la Puntuación de Recurrencia clasificaron a los pacientes en una de tres categorías: bajo riesgo, riesgo intermedio y alto riesgo de recurrencia de la enfermedad a distancia. La proporción de 668 pacientes categorizados como de riesgo bajo, intermedio y alto por el ensayo de RT-PCR fue de 51 %, 23 % y 27 %, respectivamente. Las estimaciones de Kaplan-Meier y los intervalos de confianza del 95 % para las tasas de recurrencia a distancia a los 10 años fueron de 6,8 % (4,0 %, 9,6 %), 14,3 % (8,3 %, 20,3 %) 30,5 % (23,6 %, 37,4 %), para los grupos de bajo, intermedio y alto riesgo respectivamente; la tasa para el grupo de bajo riesgo fue significativamente menor que la tasa para el grupo de alto riesgo ($p < 0,001$). En un modelo de Cox multivariante que relaciona la recurrencia distante con la puntuación de recurrencia, la edad y el tamaño del tumor, la Puntuación de Recurrencia proporciona un poder predictivo significativo ($p < 0,001$) que va más allá de la edad y el tamaño del tumor. Este estudio validó la Puntuación de Recurrencia como un potente predictor de recurrencia a distancia en pacientes tratados con tamoxifeno, sin ganglios implicados y que tienen tumores que son positivos para ESR1 (Paik y otros Breast Cancer Research and Treatment 82, Suplemento 1: página S10, 2003 (Resumen 16).

Al ampliar los resultados de estos hallazgos y al usar los resultados del Estudio B-20 de NSABP, la presente invención proporciona genes y conjuntos de genes útiles para predecir la respuesta a la quimioterapia de pacientes con cáncer de mama. Además, la invención proporciona una prueba clínicamente validada, predictiva de la respuesta de la paciente con cáncer de mama a la quimioterapia, mediante el uso de ARN de múltiples genes y análisis.

En particular, los presentes inventores identificaron un conjunto de genes: BCL2; SCUBE2; CCNB1; CTSL2; ESR1; MMP11; MYBL2; PGR; STK6; BIRC5 y MMP11, GSTM1, CD68; BOLSA1; GRB7; ERBB2, que son útiles para predecir si es probable que un paciente con cáncer, tal como un paciente con cáncer de mama, muestre una respuesta beneficiosa a la quimioterapia. Algunos de estos genes son predictivos individualmente, mientras que otros se usan como parte de ciertos grupos de genes, usados como variables en los procedimientos de la presente invención.

Por lo tanto, las variables independientes usadas en los procedimientos predictivos de la presente invención incluyen uno o más de (i) Puntuación de Recurrencia, (ii) Puntuación del Grupo ESR1; (iii) Puntuación del Grupo de Invasión; (iv) Puntuación Umbral del Grupo de Proliferación; y (v) el nivel de expresión del transcrito de ARN de al menos uno de los genes MYBL2 y SCUBE2, en donde

(b1) por cada unidad de un aumento en el valor de uno o más de (i), (iii), (iv), o el nivel de expresión del transcrito de ARN de MYBL2, o el producto de expresión correspondiente, se identifica al paciente con una probabilidad proporcionalmente mayor de tener una respuesta beneficiosa a la quimioterapia;

(b2) por cada unidad de aumento en el valor de (ii) o el nivel de expresión del transcrito de ARN de SCUBE2, o el producto de expresión correspondiente, se identifica al paciente que tiene una probabilidad proporcionalmente disminuida de una respuesta beneficiosa a la quimioterapia;

En las variables anteriores:

$$\text{Puntuación del Grupo ESR1} = (\text{ESR1} + \text{PGR} + \text{BCL2} + \text{SCUBE2})/4$$

$$\text{Puntuación del Grupo de Invasión} = (\text{CTSL2} + \text{MMP11})/2$$

$$\text{Puntuación del Grupo de Proliferación} = (\text{BIRC5} + \text{MK167} + \text{MYBL2} + \text{CCNB1} + \text{STK6})/5;$$

La Puntuación Umbral del Grupo de Proliferación es igual a 6,5, si la puntuación del Grupo de Proliferación es menor a 6,5; y es idéntico a la Puntuación del Grupo de Proliferación, si la puntuación del Grupo de Proliferación es 6,5 o más, y la puntuación de recurrencia (RS):

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

en donde

$$\text{Puntuación del Grupo GRB7} = 0,9 \times \text{GRB7} + 0,1 \times \text{ERBB2};$$

La Puntuación Umbral del grupo GRB7 es igual a 6,5, si la Puntuación del Grupo GRB7 es menor a 6,5; y es idéntico a la Puntuación de Recurrencia del Grupo GRB7, si la Puntuación de Recurrencia del Grupo GRB7 es 6,5 o más,

y

$$RS_u = \begin{aligned} &0,47 \times \text{Puntuación Umbral del Grupo GRB7} \\ &- 0,34 \times \text{Puntuación del Grupo ESR1} \\ &+ 1,04 \times \text{Puntuación Umbral del Grupo de Proliferación} \\ &+ 0,10 \times \text{Puntuación del Grupo de Invasión} \\ &+ 0,05 \times \text{CD68} \\ &- 0,08 \times \text{GSTM1} \\ &- 0,07 \times \text{BAG1} \end{aligned}$$

donde los símbolos de genes en las ecuaciones representan los niveles de expresión de los transcritos de ARN de los genes respectivos, o sus productos de expresión.

En diversas realizaciones de las invenciones, están disponibles diversos enfoques tecnológicos para la determinación de los niveles de expresión de los genes divulgados, que incluyen, sin limitación, RT-PCR, micromatrices, análisis en serie de la expresión génica (SAGE) y análisis de expresión génica mediante secuenciación paralela masiva de la firma (MPSS), que se analizan en detalle a continuación. En realizaciones particulares, el nivel de expresión de cada gen puede determinarse en relación con diversas características de los productos de expresión del gen que incluyen exones, intrones, epítotos proteicos y actividad proteica.

B.2 Perfil de expresión génica

En general, los procedimientos de creación de perfiles de expresión génica se pueden dividir en dos grandes grupos: procedimientos basados en el análisis de hibridación de polinucleótidos y procedimientos basados en la secuenciación de polinucleótidos. Los procedimientos más comúnmente usados conocidos en la técnica para la cuantificación de la expresión de ARNm en una muestra incluyen hibridación northern e hibridación *in situ* (Parker y Barnes, *Methods in Molecular Biology* 106:247-283 (1999)); Ensayos de protección de ARNasa (Hod, *Biotechniques* 13:852-854 (1992)); y la reacción en cadena de la polimerasa con transcriptasa inversa (RT-PCR) (Weis y otros, *Trends in Genetics* 8:263-264 (1992)). Alternativamente, se pueden emplear anticuerpos que puedan reconocer pares específicos, incluidos pares de ADN, pares de ARN y pares híbridos de ADN-ARN o pares de ADN-proteína. Los procedimientos representativos para el análisis de expresión génica basados en secuenciación incluyen el análisis en serie de la expresión génica (SAGE) y el análisis de expresión génica mediante secuenciación paralela masiva de la firma (MPSS).

Dos procesos biológicos que comúnmente están implicados en la tumorigénesis incluyen a la amplificación de genes y a la metilación del ADN. Ambos procesos dan como resultado la expresión aberrante de genes importantes en la formación o progresión tumoral. Por lo tanto, los procedimientos que monitorean la amplificación de genes y la metilación del ADN pueden considerarse procedimientos sustitutos para el perfil de expresión génica.

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

También se ha demostrado que la metilación del ADN es una alteración común en el cáncer que conduce a una expresión elevada o disminuida de un amplio espectro de genes (Jones, P.A. *Cancer Res.* 65:2463 (1996)). En general, la hipometilación de las islas CpG en las regiones promotoras y los elementos reguladores da como resultado una mayor expresión génica, incluidas las de muchos oncogenes (Hanada, M., y otros, *Blood* 82:1820 (1993), Feinberg, A.P.

y Vogelstein, B. *Nature* 301:89 (1983)). Debido a que la metilación del ADN se correlaciona con el nivel de expresión génica específica en muchos tipos de cáncer, sirve como un sustituto útil para el perfil de expresión de tumores (Toyota, M. y otros, *Blood* 97: 2823 (2001), Adorjan, P. y otros. *Nucl. Ácidos. Res.* 10:e21 (2002)).

5 PCR de transcriptasa inversa (RT-PCR)

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

La primera etapa es el aislamiento del ARNm de una muestra diana. El material de partida es típicamente ARN total aislado de tejidos de seres humanos o líneas celulares. Por lo tanto, el ARN puede aislarse de una variedad de tumores primarios, que incluyen mama, pulmón, colon, próstata, cerebro, hígado, riñón, páncreas, bazo, timo, testículo, ovario, útero, etc., o líneas celulares tumorales. Si la fuente de ARNm es un tumor primario, el ARNm se puede extraer, por ejemplo, de muestras de tejido congeladas o archivadas embebidas en parafina y fijadas (por ejemplo, fijadas con formalina).

Los procedimientos generales para la extracción de ARNm son bien conocidos en la técnica y son divulgados en libros de texto de norma de biología molecular, que incluyen Ausubel y otros, *Current Protocols of Molecular Biology*, John Wiley y Sons (1997). Los procedimientos para la extracción de ARN de tejidos embebidos en parafina se divulgan, por ejemplo, en Rupp y Locker, *Lab Invest.* 56:A (1987) y De Andrés y otros, *BioTechniques* 18:42044 (1995). En particular, el aislamiento de ARN puede realizarse mediante el uso de un kit de purificación, un conjunto de tampones y una proteasa de fabricantes comerciales, como Qiagen, de acuerdo con las instrucciones del fabricante. Por ejemplo, el ARN total de las células en cultivo puede aislarse mediante el uso de mini columnas del kit Qiagen RNeasy. Otros kits de aislamiento de ARN disponibles comercialmente incluyen el Kit completo de purificación de ADN y ARN MasterPure™ (EPICENTRE®, Madison, WI) y el Kit de aislamiento de ARN de bloque de parafina (Ambion, Inc.). El ARN total de las muestras de tejido puede aislarse mediante el uso de ARN Stat-60 (Tel-Test). El ARN preparado a partir de tumor puede aislarse, por ejemplo, mediante centrifugación en gradiente de densidad con cloruro de cesio.

Como el ARN no puede servir como molde para la PCR, el primer paso en el perfil de expresión génica mediante la RT-PCR es la transcripción inversa del molde de ARN en ADNc, seguido de su amplificación exponencial en una reacción de PCR. Las dos transcriptasas inversas más usadas comúnmente 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 transcripción inversa, se ceba típicamente mediante el uso de cebadores específicos, hexámeros aleatorios o cebadores oligo-dT, dependiendo de las circunstancias y el objetivo del perfil de expresión. Por ejemplo, el ARN extraído puede transcribirse de forma inversa utilizando un kit de PCR de ARN GeneAmp (Perkin Elmer, CA, Estados Unidos), siguiendo las instrucciones del fabricante. El ADNc derivado se puede usar como molde en la reacción de PCR posterior.

Aunque la etapa de PCR puede usar una variedad de ADN polimerasas termoestables dependientes de ADN, típicamente emplea la ADN polimerasa Taq, que tiene una actividad nucleasa 5'-3' pero carece de una actividad endonucleasa correctora 3'-5'. Por lo tanto, la PCR TaqMan® usa típicamente la actividad nucleasa 5' de la polimerasa Taq o Tth para hidrolizar una sonda de hibridación unida a su amplicón objetivo, pero se puede usar cualquier enzima con actividad nucleasa 5' equivalente. Se usan dos cebadores oligonucleotídicos para generar un amplicón típico de una reacción de PCR. Un tercer oligonucleótido, o sonda, está diseñado para detectar la secuencia de nucleótidos ubicada entre los dos cebadores de PCR. La sonda no es extensible por la enzima Taq ADN polimerasa, y está marcada con un colorante fluorescente indicador y un colorante fluorescente apagador. Cualquier emisión inducida por el láser del colorante indicador se apaga mediante el colorante de enfriamiento cuando los dos colorantes se encuentran muy juntos, como lo están en la sonda. Durante la reacción de amplificación, la enzima Taq ADN polimerasa escinde la sonda de manera dependiente del molde. Los fragmentos de sonda resultantes se disocian en solución y la señal del colorante reportero liberado está libre del efecto de enfriamiento del segundo fluoróforo. Se libera una molécula de colorante reportero por cada nueva molécula sintetizada, y la detección del colorante reportero no apagado proporciona la base para la interpretación cuantitativa de los datos.

La TaqMan® RT-PCR puede realizarse utilizando equipos disponibles en el mercado, como, por ejemplo, ABI PRISM 7700™ Sequence Detection System™ (Perkin-Elmer-Applied Biosystems, Foster City, CA, Estados Unidos) o Lightcycler (Roche Molecular Biochemicals, Mannheim, Alemania). En una realización preferente, el procedimiento de la nucleasa 5' se ejecuta en un dispositivo de PCR cuantitativa en tiempo real tal como el ABI PRISM 7700™ Sequence Detection System™. El sistema consta de un termociclador, láser, dispositivo de carga acoplada (CCD), cámara y computadora. El sistema amplifica muestras en un formato de 96 pocillos en un termociclador. Durante la amplificación, se detecta una señal fluorescente inducida por láser en el CCD. El sistema incluye software para ejecutar el instrumento y para analizar los datos.

Los datos del ensayo de la 5'-nucleasa se expresan inicialmente como C_T , o el ciclo umbral. Como se discutió 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 cuando la señal fluorescente se registra por primera vez como estadísticamente significativa es el ciclo umbral (C_T).

5 Para minimizar los errores y el efecto de la variación de muestra a muestra, la RT-PCR generalmente se realiza utilizando una norma interna. La norma interna ideal se expresa a un nivel constante entre diferentes tejidos y no se ve afectada por el tratamiento experimental. Los ARN que se utilizan con mayor frecuencia para normalizar los patrones de expresión génica son los ARNm para los genes de normalización gliceraldehído-3-fosfato-deshidrogenasa (GAPD) y β -actina (ACTB).

10 Una variación más reciente de la técnica de RT-PCR es la PCR cuantitativa en tiempo real, que mide la acumulación de productos de PCR a través de una sonda fluorogénica de doble marca (es decir, la sonda TaqMan®). La PCR en tiempo real es compatible tanto con la PCR competitiva cuantitativa, donde el competidor interno para cada secuencia objetivo se usa para la normalización, como con la PCR comparativa cuantitativa que utiliza un gen de normalización contenido en la muestra, o un gen de normalización para RT-PCR. Para más detalles véase, por ejemplo Held y otros, Genome Research 6:986-994 (1996).

15 Las etapas de un protocolo representativo para perfilar la expresión génica utilizando tejidos fijos embebidos en parafina como fuente de ARN, incluido el aislamiento de ARNm, la purificación, la extensión del cebador y la amplificación, se dan en varios artículos publicados en revistas {por ejemplo: TE Godfrey y otros. J. Molec. Diagnósticos 2: 84-91 [2000]; K. Specht y otros, Am. J. Pathol. 158: 419-29 [2001]}. Brevemente, un procedimiento representativo comienza con el corte de secciones de tejido tumoral embebido en parafina de aproximadamente 10 μ m de grosor. Luego se extrae el ARN y se eliminan las proteínas y el ADN. Después del análisis de la concentración de ARN, se pueden incluir las etapas de reparación y/o amplificación de ARN, si es necesario, y el ARN se transcribe de manera inversa mediante el uso de promotores específicos del gen seguido de RT-PCR.

25 Micromatrices

La expresión génica diferencial también se puede identificar o confirmar mediante la técnica de micromatrices. Por lo tanto, el perfil de expresión de los genes asociados con el cáncer de mama se puede medir en tejido tumoral fresco o embebido en parafina, mediante el uso de la tecnología de micromatrices. En este procedimiento, las secuencias de polinucleótidos de interés (incluidos los ADNc y los oligonucleótidos) se colocan en placas, o se agrupan, en un sustrato de microchip. Las secuencias ordenadas se hibridan luego con sondas de ADN específicas de células o tejidos de interés. Al igual que en el procedimiento de la RT-PCR, típicamente la fuente de ARNm es el ARN total aislado de tumores de seres humanos o líneas celulares tumorales, y los tejidos normales o líneas celulares correspondientes. Por lo tanto, el ARN puede aislarse de una variedad de tumores primarios o líneas celulares tumorales. Si la fuente de ARNm es un tumor primario, se puede extraer ARNm, por ejemplo, de muestras de tejido congeladas o archivadas embebidas en parafina y fijadas (por ejemplo, fijadas en formalina), que se preparan y se conservan de forma rutinaria en la práctica clínica diaria.

40 En una realización específica de la técnica de micromatrices, los insertos amplificados por PCR de clones de ADNc se aplican a un sustrato en una matriz densa. Preferiblemente, se aplican al menos 10.000 secuencias de nucleótidos al sustrato. Los genes de la micromatriz, inmovilizados en el microchip a 10.000 elementos cada uno, son adecuados para la hibridación en condiciones rigurosas. Las sondas de ADNc marcadas con fluorescencia pueden generarse mediante la incorporación de nucleótidos fluorescentes mediante transcripción inversa de ARN extraído de tejidos de interés. Las sondas marcadas de ADNc, aplicadas al chip, hibridan con especificidad en cada punto de ADN en el arreglo. Después de un lavado riguroso para eliminar las sondas unidas de forma no específica, el chip se escanea mediante microscopía láser confocal u otro procedimiento de detección, como una cámara CCD. La cuantificación de la hibridación de cada elemento ordenado permite la evaluación de la abundancia de ARNm correspondiente. Con fluorescencia de doble color, las sondas de ADNc marcadas por separado generadas a partir de dos fuentes de ARN se hibridan por pares con el arreglo. La abundancia relativa de los transcritos de las dos fuentes correspondientes a cada gen especificado se determina así simultáneamente. La escala miniaturizada de la hibridación permite una evaluación conveniente y rápida del patrón de expresión para un gran número de genes. Se ha demostrado que dichos procedimientos tienen la sensibilidad necesaria para detectar los transcritos raros, que se expresan en unas pocas copias por célula, y para detectar reproduciblemente aproximadamente al menos dos veces las diferencias en los niveles de expresión (Scheny y otros, Proc. Natl. Acad. Sci. USA 93(2):106-149 (1996)). El análisis de micromatrices puede realizarse mediante equipos disponibles en el mercado, siguiendo los protocolos del fabricante, como el uso de la tecnología Affymetrix GenChip o la tecnología de micromatrices de Incyte.

60 El desarrollo de procedimientos de micromatrices para el análisis a gran escala de la expresión génica permite buscar sistemáticamente marcadores moleculares de clasificación del cáncer y de predicción de los resultados en una variedad de tipos de tumores.

Análisis en Serie de la Expresión Génica.

65 El análisis en serie de la expresión génica (SAGE) es un procedimiento que permite el análisis simultáneo y cuantitativo de una gran cantidad de transcritos de genes, sin la necesidad de proporcionar una sonda de hibridación individual para

cada transcrito. Primero, se genera una etiqueta de secuencia corta (aproximadamente de 10-14 pb) que contiene información suficiente para identificar de forma única un transcrito, siempre que la etiqueta se obtenga desde una posición única dentro de cada transcrito. Luego, muchos transcritos se unen para formar moléculas seriales largas, que pueden secuenciarse, revelando la identidad de las múltiples etiquetas simultáneamente. El patrón de expresión de cualquier población de transcritos puede evaluarse cuantitativamente determinando la abundancia de etiquetas individuales e identificando el gen correspondiente a cada etiqueta. Para más detalles véase, por ejemplo Velculescu y otros, *Science* 270:484-487 (1995); y Velculescu y otros, *Cell* 88:243-51 (1997).

Análisis de la expresión génica por secuenciación paralela masiva de la firma (MPSS)

Este procedimiento, descrito por Brenner y otros, *Nature Biotechnology* 18:630-634 (2000), es un enfoque de secuenciación que combina secuenciación de firmas no basadas en gel con clonación *in vitro* de millones de moldes en microperlas de 5 µm de diámetro separadas. Primero, una biblioteca de microesferas de moldes de ADN está construida *in vitro* por clonación. Esto es seguido por el conjunto de un arreglo plano de microperlas que contienen moldes en una celda de flujo a una alta densidad (típicamente mayor que 3×10^6 microperlas/cm²). Los extremos libres de los moldes clonados en cada microperla se analizan simultáneamente, mediante el uso de un procedimiento de secuenciación de firma basado en fluorescencia que no requiere separación de los fragmentos de ADN. Se ha demostrado que este procedimiento proporciona simultáneamente y con precisión, en una sola operación, cientos de miles de secuencias de firma de genes de una biblioteca de ADNc de levadura.

Descripción general del aislamiento, purificación y amplificación de ARNm

Las etapas de un protocolo representativo para perfilar la expresión génica utilizando tejidos fijados embebidos en parafina como fuente de ARN, incluido el aislamiento de ARNm, la purificación, la extensión del cebador y la amplificación, se proporcionan en varios artículos publicados en revistas (por ejemplo: T.E. Godfrey y otros. *J. Molec. Diagnósticos* 2: 84-91 [2000]; K. Specht y *al.*, *Am. J. Pathol.* 158: 419-29 [2001]). Brevemente, un procedimiento representativo comienza con el corte de secciones de tejido tumoral embebido en parafina de aproximadamente 10 µm de grosor. Luego se extrae el ARN y se eliminan las proteínas y el ADN. Después del análisis de la concentración de ARN, se pueden incluir las etapas de reparación y/o amplificación de ARN, si es necesario, y el ARN se transcribe de manera inversa mediante el uso de promotores específicos del gen seguido de RT-PCR. Finalmente, los datos se analizan para identificar la(s) mejor(es) opción(es) de tratamiento disponibles para el paciente en función del patrón característico de expresión génica identificado en la muestra de tumor examinada, dependiendo de la probabilidad pronosticada de recurrencia del cáncer.

Conjunto de genes de cáncer de mama, subsecuencias genéticas analizadas y aplicación clínica de datos de expresión génica

Un aspecto importante de la presente invención es usar la expresión medida de ciertos genes en el tejido de cáncer de mama para proporcionar información de pronóstico o predictiva. Para este propósito, es necesario corregir (normalizar) tanto las diferencias en la cantidad de ARN analizado como la variabilidad en la calidad del ARN utilizado. Por lo tanto, el ensayo generalmente mide e incorpora la expresión de ciertos genes de normalización, incluidos genes de normalización bien conocidos, como ACTB, GAPD, GUSB, RPL0 y TFRC, como se muestra en el siguiente ejemplo. Alternativamente, la normalización puede basarse en la señal media o mediana (C_T) de todos los genes analizados o un gran subconjunto de los mismos (enfoque de normalización global). A continuación, a menos que se indique lo contrario, la expresión génica significa expresión normalizada.

Diseño de cebadores de PCR basados en intrones y sondas

De acuerdo con un aspecto de la presente invención, los cebadores y las sondas de PCR están diseñados en base a secuencias de intrones presentes en el gen a amplificar. En consecuencia, el primer paso en el diseño del cebador/sonda es la delineación de secuencias de intrones dentro de los genes. Esto puede hacerse mediante un programa disponible públicamente, como el programa ADN BLAT desarrollado por Kent, WJ, *Genome Res* 12(4): 656-64 (2002), o mediante el programa BLAST, incluidas sus variaciones. Las etapas posteriores siguen procedimientos bien establecidos de cebador de PCR y diseño de sonda.

Para evitar señales no específicas, es importante enmascarar secuencias repetitivas dentro de los intrones al diseñar los cebadores y las sondas. Esto se puede lograr fácilmente utilizando el programa Repeat Masker, disponible en línea a través del Baylor College of Medicine, que analiza las secuencias de ADN contra una biblioteca de elementos repetitivos y devuelve una secuencia de consulta en la que los elementos repetitivos están enmascarados. Las secuencias de intrón enmascaradas se pueden usar para diseñar secuencias de cebadores y sondas utilizando cualquier paquete de diseño de cebadores / sondas disponible comercialmente o de cualquier otra manera, como Primer Express (Applied Biosystems); MGB ensayo por diseño (Applied Biosystems); Primer3 (Steve Rozen y Helen J. Skaletsky (2000) Primer3 en la WWW para usuarios generales y para programadores de biólogos. En: Krawetz S, Misener S (eds) *Procedimientos y protocolos bioinformáticos: Methods in Molecular Biology*. Humana Press, Totowa, NJ, páginas 365-386).

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

Para obtener más pautas para el cebador de PCR y el diseño de la sonda, véase, por ejemplo Dieffenbach, CW y otros, "General Concepts for PCR Primer Design" en: PCR Primer, A Laboratory Manual, Cold Spring Harbor Laboratory Press, Nueva York, 1995, páginas 133-155 ; Innis y Gelfand, "Optimization of PCRs" en: PCR Protocols, A Guide to Methods and Applications, CRC Press, Londres, 1994, páginas 5-11; y Plasterer, T.N. Primerselect: Primer and probe design. Methods Mol. Biol. 70:520-527 (1997), cuyas divulgaciones completas incorporadas por referencia expresamente en este documento.

B.3 Algoritmos y Procedimientos Estadísticos

La presente invención aprovecha ciertos algoritmos y procedimientos estadísticos, que se describen en la solicitud pendiente No de serie 10/883,303.

Cuando se usa la RT-PCR cuantitativa (qRT-PCR) para medir los niveles de ARNm, las cantidades de ARNm se expresan en unidades de C_T (ciclo umbral) (Held y otros, Genome Research 6:986-994 (1996)). La suma promedio de los C_T de los ARNm de referencia se establece en algún número, por ejemplo, cero, y cada C_T medido de los ARNm de prueba se da en relación con este punto. Por ejemplo, si, para una muestra de tumor de cierto paciente, se encuentra que el promedio de los C_T de los 5 genes de referencia es 31 y el C_T del gen de prueba X es 35, el valor informado para el gen X es -4 (es decir, 31-35).

Como primer paso después de 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 lo tanto, los genes que se sabe que están asociados con la proliferación celular constituirán el "Grupo de Proliferación" (eje o subconjunto). Los genes que se sabe que están asociados con la invasión por el cáncer de tejido adyacente constituirán el "Grupo de Invasión" (eje o subconjunto). Los genes asociados con las vías de señalización del receptor del factor de crecimiento clave constituirán el "Grupo de Factor de Crecimiento" (eje o subconjunto), también denominado grupo GRB7. Los genes que se sabe que participan en la activación o señalización a través del receptor de estrógenos (ESR1) constituirán el "Grupo de receptores de estrógenos (ESR1)" (eje o subconjunto), y así sucesivamente. Esta lista de subconjuntos, por supuesto, no es limitativa. Los subconjuntos creados dependerán del cáncer en particular, es decir, cáncer de mama, próstata, páncreas, pulmón, etc. En general, los genes cuya expresión se sabe que se correlacionan entre sí, o que se sabe que están involucrados en la misma vía se agrupan en el mismo subconjunto.

En el siguiente paso, el nivel de cada ARNm medido en el tumor en un subconjunto se multiplica por un coeficiente que refleja su contribución relativa dentro del conjunto, al riesgo de recurrencia del cáncer para obtener un producto, y este producto se agrega a los otros productos calculados de manera similar mediante el uso de niveles de ARNm en el subconjunto y sus coeficientes, para obtener un término, por ejemplo, un término de proliferación, un término de invasión, un término de factor de crecimiento, etc. Por ejemplo, en el caso del cáncer de mama invasivo negativo para ganglios linfáticos, el término de factor de crecimiento (Grupo GRB7) es $(0,45 \text{ a } 1,35) \times \text{GRB7} + (0,05 \text{ a } 0,15) \times \text{ERBB2}$, como, por ejemplo, $0,9 \times \text{GRB7} + 0,1 \times \text{ERBB2}$ (véase el ejemplo a continuación).

La contribución de cada término a la puntuación de recurrencia general se pondera mediante el uso de un coeficiente adicional. Por ejemplo, en el caso del cáncer de mama invasivo con ganglios linfáticos negativos, el coeficiente del término del Grupo GRB7 puede estar entre 0,23 y 0,70.

Además, para algunos términos, como el factor de crecimiento y los términos de proliferación, se realiza un paso más. Si la relación entre el término y el riesgo de recurrencia no es lineal, se utiliza una transformación funcional no lineal del término, como un umbral.

La suma de los términos obtenidos proporciona la puntuación de recurrencia (RSu), que predice la probabilidad de recurrencia del cáncer en el curso normal de la enfermedad.

La escala RS que se genera por el algoritmo de la presente invención se puede ajustar de varias maneras. Por lo tanto, el rango podría seleccionarse de manera que la escala corra de 0 a 10, de 0 a 50 o de 0 a 100, por ejemplo.

Por ejemplo, en el enfoque de escala particular descrito en el Ejemplo a continuación, la puntuación de recurrencia escalada se calcula en una escala de 0 a 100. Por conveniencia, se agrega 10 a cada valor de C_T medido, y un RS sin escala se calcula como se describe anteriormente. Las ecuaciones para calcular el RS y el SRS se proporcionan en el siguiente ejemplo.

Al calcular la puntuación de recurrencia, o cualquier variable que se utilice para calcular la puntuación de recurrencia, cualquier gen puede ser sustituido por otro gen que se coexpresa con el primer gen en el cáncer particular probado con un coeficiente de Pearson $\geq 0,5$. De manera similar, cualquier gen individual, o gen dentro de un grupo de genes (subconjunto) incluidos en los procedimientos predictivos y de pronóstico de la presente invención puede ser sustituido por otro gen que coexpresa con el primer gen en el cáncer particular probado con un coeficiente de Pearson $\geq 0,5$.

B.4 Quimioterapia contra el cáncer

Los agentes quimioterapéuticos que se utilizan en el tratamiento del cáncer se pueden dividir en varios grupos, según su mecanismo de acción. Algunos agentes quimioterapéuticos dañan directamente el ADN y el ARN. Al interrumpir la replicación del ADN, tales quimioterapéuticos detienen completamente la replicación o dan como resultado la producción de ADN o ARN sin sentido. Esta categoría incluye, por ejemplo, cisplatino (Platinol®), daunorrubicina (Cerubidine®), doxorubicina (Adriamycin®) y etopósido (VePesid®). Otro grupo de agentes quimioterapéuticos contra el cáncer interfiere con la formación de nucleótidos o desoxirribonucleótidos, de manera que se bloquea la síntesis de ARN y la replicación celular. Los ejemplos de medicamentos en esta clase incluyen metotrexato (Abitrexate®), mercaptopurina (Purinethol®), fluorouracilo (Acrucil®) e hidroxiurea (Hydrea®). Una tercera clase de agentes quimioterapéuticos efectúa la síntesis o descomposición de los husos mitóticos y como resultado, interrumpe la división celular. Los ejemplos de medicamentos en esta clase incluyen Vinblastina (Velban®), Vincristina (Oncovin®) y taxanos, como Pacitaxel (Taxol®) y Tocetaxel (Taxotere®). Actualmente, el Tocetaxel está aprobado en los Estados Unidos para tratar a pacientes con cáncer de mama localmente avanzado o metastásico después del fracaso de la quimioterapia previa y pacientes con cáncer de pulmón de células no pequeñas localmente avanzado o metastásico después del fracaso de la quimioterapia previa basada en platino. La predicción de la respuesta del paciente a todos estos y otros agentes quimioterapéuticos está específicamente dentro del ámbito de la presente invención.

En una realización específica, la quimioterapia incluye el tratamiento con un derivado de taxano. Los taxanos incluyen, sin limitación, paclitaxel (Taxol®) y docetaxel (Taxotere®), que se usan ampliamente en el tratamiento del cáncer. Como se discutió anteriormente, los taxanos afectan las estructuras celulares llamadas microtúbulos, que juegan un papel importante en las funciones celulares. En el crecimiento celular normal, los microtúbulos se forman cuando una célula comienza a dividirse. Una vez que la célula deja de dividirse, los microtúbulos se descomponen o destruyen. Los taxanos impiden que los microtúbulos se rompan, lo que bloquea la proliferación celular.

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

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

El tratamiento con cualquier combinación de estos y otros fármacos quimioterapéuticos se contempla específicamente.

La mayoría de los pacientes reciben quimioterapia inmediatamente después de la extirpación quirúrgica del tumor. Este enfoque se conoce comúnmente como terapia adyuvante. Sin embargo, la quimioterapia se puede administrar también antes de la cirugía, como el llamado tratamiento neoadyuvante. Aunque el uso de quimioterapia neoadyuvante se origina en el tratamiento del cáncer de mama avanzado e inoperable, también ha ganado aceptación en el tratamiento de otros tipos de cáncer. La eficacia de la quimioterapia neoadyuvante se ha probado en varios ensayos clínicos. En el ensayo multicéntrico National Surgical Adjuvant Breast and Bowel Project B-18 (NSAB B-18) (Fisher y otros, J. Clin. Oncología 15:2002-2004 (1997); Fisher y otros, J. Clin. Oncología 16: 2672-2685 (1998)) la terapia neoadyuvante se realizó con una combinación de adriamicina y ciclofosfamida ("régimen AC"). En otro ensayo clínico, la terapia neoadyuvante se administró mediante el uso de una combinación de 5-fluorouracilo, epirubicina y ciclofosfamida ("régimen FEC") (van Der Hage y otros, J. Clin. Oncol. 19:4224-4237 (2001)). Los ensayos clínicos más recientes también han utilizado regímenes de tratamiento neoadyuvante que contienen taxanos. Véase, por ejemplo Holmes y otros, J. Natl. Cancer Inst. 83:1797-1805 (1991) y Moliterni y otros, Seminars in Oncology, 24:S17-10-S-17-14 (1999). Para obtener más información sobre la quimioterapia neoadyuvante para el cáncer de mama, véase, Cleator y otros, Endocrine-Related Cancer 9:183-195 (2002).

B.5 Kits de la invención

Los materiales para usar en los procedimientos de la presente invención son adecuados para la preparación de kits producidos de acuerdo con procedimientos bien conocidos. Por lo tanto, la divulgación proporciona kits que comprenden agentes, que pueden incluir sondas y/o cebadores específicos o selectivos de genes, para cuantificar la expresión de los genes divulgados para predecir el resultado pronóstico o la respuesta al tratamiento. Tales kits pueden contener opcionalmente reactivos para la extracción de ARN de muestras tumorales, en particular muestras de tejido fijadas en parafina y/o reactivos para la amplificación de ARN. Además, los kits pueden comprender opcionalmente el(los) reactivo(s) con una descripción o etiqueta de identificación o instrucciones relacionadas con su uso en los procedimientos de la presente invención. Los kits pueden comprender contenedores (que incluyen placas de microtitulación adecuadas para su uso en una implementación automatizada del procedimiento), cada uno con uno o

más de los diversos reactivos (típicamente en forma concentrada) que se utilizan en los procedimientos, que incluyen, por ejemplo, micromatrices prefabricados, tampones, los nucleótidos trifosfatos apropiados (p. ej., dATP, dCTP, dGTP y dTTP; o rATP, rCTP, rGTP y UTP), transcriptasa inversa, ADN polimerasa, ARN polimerasa y una o más sondas y cebadores unidos a un promotor reactivo con la ARN polimerasa).

Los procedimientos proporcionados por la presente invención también pueden automatizarse parcial o de manera total. Todos los aspectos de la presente invención también se pueden practicar de manera que un número limitado de genes adicionales que se coexpresan con los genes divulgados, por ejemplo, como lo demuestran los altos coeficientes de correlación de Pearson, se incluyan en pruebas de pronóstico o predictivas además de los genes divulgados.

Una vez descrita la invención, la misma se entenderá más fácilmente mediante la referencia al siguiente ejemplo, que se proporciona a modo de ilustración, y no pretende limitar la invención de ninguna manera.

Ejemplo

Un estudio de quimioterapia neoadyuvante en cáncer de mama invasivo: Perfil de expresión génica de tejido de biopsia central embebido en parafina

Este estudio se realizó para identificar genes o grupos de genes que predican la sensibilidad o resistencia del paciente a la quimioterapia. El estudio utilizó tejido y datos del Estudio B-20 de NSABP: "A Clinical Trial to Determine the Worth of Chemotherapy and Tamoxifen over Tamoxifen Alone in the Management of Patients with Primary Invasive Breast Cancer, Negative Axillary Nodes and Estrogen-Receptor-Positive Tumors." Fisher y otros, J Natl Cancer Inst 89 (22):1673-1682 (1997).

Diseño del estudio

Criterios de inclusión del paciente: Inscrito en NSABP Estudio B-20. Criterios de exclusión del paciente: No hay bloqueo tumoral disponible desde el diagnóstico inicial en el archivo NSABP; sin tumor o muy poco tumor en el bloque según lo evaluado por el examen del portaobjetos de H&E por el patólogo; ARN insuficiente (<275 ng) para análisis de RT-PCR; CT promedio no normalizado para los 5 genes de referencia <35; Clínicamente no elegible o sin seguimiento.

Ensayo de laboratorio

Se analizaron muestras fijas de tejido tumoral de mama embebido en parafina de hasta 600 pacientes que fueron tratadas al inicio del estudio con TAM más quimioterapia en el estudio B-20. El ARN previamente extraído de tejido tumoral de mama embebido en parafina fija de hasta 252 pacientes que fueron tratados al ingreso al estudio con TAM solo, en el estudio B-20, se reanalizó. La expresión de 16 genes relacionados con el cáncer y 5 genes de referencia se evaluó cuantitativamente para cada paciente mediante el uso de TaqMan® RT-PCR, que se realizó por triplicado con entrada de ARN a 2 ng por reacción.

El algoritmo de expresión génica que se definió prospectivamente antes del análisis por RT-PCR del tejido tumoral en este estudio se usó para calcular una Puntuación de Recurrencia para cada paciente.

Revisión y preparación de patología

Grupo 1: Se excluyeron del estudio los casos sin tumor o con muy poco tumor (<5 % del área ocupada por células cancerosas invasivas en comparación con el área ocupada por otros elementos epiteliales, como epitelio normal, cambio fibroquístico o DCIS/LCIS).

Grupo 2: Casos con regiones en el portaobjetos que tienen elementos prominentes no tumorales (como músculo liso, hemorragia, fibrosis, hiperplásico, epitelio y/o mama normal; pero no DCIS, LCIS o necrosis) donde los elementos no tumorales estaban lo suficientemente localizados y eran lo suficientemente abundante (> 50 % del tejido total en el portaobjetos) para ser susceptible de macro-disección. La macro-disección se realizó en estos casos.

Grupo 3: Todos los demás casos se analizaron sin disección.

Supervivencia del paciente

Para el análisis primario, la supervivencia sin recurrencia distante (DRFS) se basó en el tiempo (en años) desde la cirugía hasta la primera recurrencia distante. La enfermedad contralateral, otros segundos cánceres primarios y las muertes previas a la recurrencia a distancia se consideraron eventos censuradores.

Expresión génica

Los niveles de expresión de 21 genes que se utilizaron en el cálculo de la puntuación de recurrencia se informaron como valores del ensayo GHI. La Tabla 1 proporciona las identidades de 16 genes de prueba y 5 de referencia. Los valores de expresión génica se normalizaron en relación con la media de los 5 genes de referencia. Se sabe que los genes de referencia son relativamente invariables en el cáncer de mama, así como en diversas condiciones de muestra y procedimiento, lo que los hace útiles para normalizar los efectos extraños. Las medidas de expresión normalizadas de referencia generalmente varían de 0 a 15, donde un aumento de una unidad generalmente refleja un aumento de 2 veces en la cantidad de ARN. Los 21 genes pre-especificados para el análisis se enumeran en la Tabla 1.

Tabla 1

Panel de expresión génica

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

Análisis estadístico

La Puntuación de Recurrencia contiene factores pronósticos y predictivos. Con el fin de identificar los genes predictivos del tratamiento en el cáncer de mama, el objetivo principal fue explorar la relación entre la expresión génica y el DRFS en las pacientes tratadas. Para dichos análisis, se utilizaron datos de pacientes tratados y no tratados para discriminar los genes predictivos del tratamiento de los genes puramente pronósticos. Para identificar los genes predictivos del tratamiento de quimioterapia, tanto los pacientes tratados con TAM solamente como los pacientes tratados con TAM y quimioterapia se incluyeron en el Estudio B-20 de NSABP.

Se utilizaron modelos de riesgos proporcionales de Cox para examinar la interacción entre el efecto del tratamiento y la expresión génica de Cox, J Royal Stat Soc Serie B 34(2):187-220(1972); Therneau y Gramsch, Modeling Survival Data: Extending the Cox Model, Springer, Nueva York, NY (2000) ISBN 0-387-98784-3. Existe una interacción entre el tratamiento y la expresión génica si el efecto del tratamiento depende del nivel de expresión génica; es decir, si la expresión génica es un factor predictivo del tratamiento (Fisher, Statistical Methods and Scientific Inference, Oliver y Boyd, Edimburgo (1974); Savage The foundations of Statistics, John Wiley, Nueva York (1964). La prueba de razón de verosimilitud se usó para identificar genes predictivos de tratamiento estadísticamente significativos al comparar el modelo reducido que excluye la expresión génica por interacción del tratamiento versus el modelo completo competidor que incluye la expresión génica por interacción del tratamiento.

Puntuación de Recurrencia

La puntuación de recurrencia (RS) en una escala de 0 a 100 se deriva de las medidas de expresión de referencia normalizadas de la siguiente manera:

$$RS_u = 0,47 \times \text{Puntuación Umbral del Grupo GRB7} - 0,34 \times \text{Puntuación del Grupo ESR1} + 1,04 \times \text{Puntuación Umbral del Grupo de Proliferación} + 0,10 \times \text{Puntuación del Grupo de Invasión} + 0,05 \times CD68 - 0,08 \times GSTM1 - 0,07 \times BAG1$$

5

10

15 donde:

$$\text{Puntuación del Grupo GRB7} = 0,9 \times \text{GRB7} + 0,1 \times \text{ERBB2}$$

20

$$\text{Puntuación Umbral del Grupo GRB7} = \begin{cases} 8 & \text{Si Puntuación del Grupo GRB7} < 8 \\ \text{Puntuación del Grupo GRB7} & \text{De otra manera} \end{cases}$$

25

$$\text{Puntuación del Grupo ESR1} = (x \text{ Esrt1} + x \text{ PGR} + \text{BCL2} + \text{SCUBE2})/4$$

30

$$\text{Puntuación del Grupo de Proliferación} = (\text{BIRC5} + \text{MK167} + \text{MYBL2} + \text{CCNB1} + \text{STK6})/5;$$

35

$$\text{Puntuación Umbral del Grupo de Proliferación} = \begin{cases} 6,5 & \text{Si Puntuación del Grupo de Prolif.} < 6,5 \\ \text{Puntuación del Grupo de Proliferación} & \text{De otra manera} \end{cases}$$

40

$$\text{Puntuación del Grupo de Invasión} = (\text{CTSL2} + \text{MMP11})/2$$

45

El RS_u (La Puntuación de Recurrencia sin escala) se vuelve a escalar para que esté entre 0 y 100:

50

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

55 Clasificación en tres grupos

El RS se utilizó para determinar un grupo de riesgo de recurrencia para cada paciente. Los puntos de corte entre los grupos de recurrencia de riesgo bajo, intermedio y alto se definirán de la siguiente manera:

60

Grupo de riesgo	Puntuación de Recurrencia
Bajo riesgo de recurrencia	Menos de 18
Riesgo intermedio de recurrencia.	Mayor o igual que 18 y menor que 31
Alto riesgo de recurrencia	Mayor o igual que 31

65

Resultados

La Tabla 2 muestra que seis de las variables probadas interactuaron con la respuesta de quimioterapia beneficiosa, medida por DRFS a 10 años, con significación estadística ($P < 0,1$), llamada, RS, Puntuación Umbral del Grupo de Proliferación (ProlThres), MYBL2, Puntuación del Grupo de Invasión, SCUBE2 y Puntuación del Grupo ESR1. El análisis de interacción para RS se realizó en la mitad inferior del rango total de 100 puntos, como lo indica el término RS/50 en la Tabla 2.

Tabla 2

Análisis de interacción					
Variable	Estimado	Valor-p	H.R.	IC del 95 % para H.R.	
IntRS/50	-1,151	0,038	0,316	0,107	0,936
IntProlThres	-1,12114	0,038	0,325	0,112	0,943
IntMYBL2	-0,4043	0,049	0,667	0,445	0,999
IntInvasionGoup	-0,64788	0,055	0,523	0,269	1,016
IntSCUBE2	0,221844	0,062	1,248	0,988	1,577
IntESR1Group	0,279682	0,093	1,322	0,953	1,834

Como se muestra en la Tabla 2, una mayor expresión de los siguientes genes y conjuntos de genes se correlaciona con una mayor probabilidad de supervivencia a 10 años sin recurrencia distante: RS; MYBL2; Puntuación Umbral del Grupo de Proliferación; Puntuación del Grupo de Invasión. El aumento de la expresión de los siguientes genes se correlaciona con una menor probabilidad de una respuesta beneficiosa al tratamiento: SCUBE2; Puntuación del Grupo ESR1. Es de destacar que los componentes claves individuales del algoritmo RS, llamado ProlifAxisthresh, InvasionGroup y Grupo ESR1, influyen independientemente en la respuesta a la quimioterapia en una dirección de acuerdo con el aumento de RS correspondiente a una mayor probabilidad de beneficio de la quimioterapia.

La Figura 1 muestra la relación entre la categoría del grupo de riesgo de RS (riesgo bajo, intermedio y alto) y el porcentaje de beneficio de la quimioterapia en la población de NSABP B-20 a los 10 años. El beneficio promedio entre los pacientes de alto riesgo (definido por $RS > 31$) fue de aproximadamente 28 %, con límites de confianza del 95 % que abarcan 12-42 %. Es decir, en este grupo, la quimioterapia promedio disminuyó el riesgo absoluto de recurrencia a los 10 años en un 28 %. Esto es notable porque los pacientes de alto riesgo sin quimioterapia en promedio tienen un riesgo absoluto de recurrencia de poco más del 30 %, lo que indica que la quimioterapia puede reducir la tasa relativa de recurrencia en alrededor del 90 % en este grupo de pacientes. En el caso de los pacientes de riesgo intermedio (definido por RS entre 18 y 31) el beneficio promedio fue casi cero, con límites de confianza del 95 % que abarcan de -10 a + 10 %. En el caso de los pacientes de bajo riesgo (definido por $RS < 18$) el beneficio promedio fue casi cero, con límites de confianza del 95 % que abarcan de -4 a + 4 %.

Estos resultados tienen utilidad para guiar la decisión sobre si tratar con quimioterapia a una paciente con cáncer de mama temprano, positivo para ESR1. La validación del algoritmo del Puntuación de Recurrencia en el brazo de tratamiento de NSABP B14 TAM demostró que los pacientes en el grupo de alto riesgo tienen un riesgo > 30 % de recurrencia del cáncer de mama a los 10 años. Los datos presentados aquí indican que esta población de alto riesgo tiene un beneficio muy sustancial del tratamiento con quimioterapia si eligen tomarlo, reduciendo potencialmente la recurrencia a la de los pacientes de bajo riesgo. Por otro lado, la población de bajo riesgo tratada con TAM, que tiene un riesgo de recurrencia sin quimioterapia del ~7 %, puede esperar que la quimioterapia produzca una reducción relativamente pequeña del riesgo.

Debido a que la RS es una variable continua, la RS numérica precisa para un paciente dado puede usarse para indicar la probabilidad individual de beneficio de la quimioterapia del paciente. Esto se muestra en la Figura 2.

Tabla 3

	Reactivo	Gen	Acceso	Oligo	Secuencia	Longitud
5	Sentido derecho	ACTB	NM_001101	S0034/B-acli.f2	CAGCAGATGTGGATCAGCAAG	21
	Sentido contrario	ACTB	NM_001101	S0036/B-acli.r2	GCATTTGCGGTGGACGAT	18
	Sonda	ACTB	NM_001101	S4730/B-acli.p2	AGGAGTATGACGAGTCCGGCCCC	23
	Sentido derecho	BAG1	NM_004323	S1386/BAG1.f2	CGTTGTCAGCACTTGAATACAA	23
10	Sentido contrario	BAG1	NM_004323	S1387/BAG1.r2	GTTCAACCTCTTCTGTGGACTGT	24
	Sonda	BAG1	NM_004323	S4731/BAG1.p2	CCCAATTAACATGACCCGGCAACCAT	26
	Sentido derecho	BCL2	NM_000633	S0043/Bcl2.f2	CAGATGGACCTAGTACCCACTGAGA	25
	Sentido contrario	BCL2	NM_000633	S0045/Bcl2.r2	CCTATGATTTAAGGGCATTITTTCC	24
15	Sonda	BCL2	NM_000633	S4732/Bcl2.p2	TTCCACGCCGAAGGACAGCGAT	22
	Sentido derecho	CCNB1	NM_031966	S1720/CCNB1.f2	TTCAGGTTGTTGCAGGAGAC	20
	Sentido contrario	CCNB1	NM_031966	S1721/CCNB1.r2	CATCTTCTTGGCACACAAT	20
	Sonda	CCNB1	NM_031966	S4733/CCNB1.p2	TGTCCTCATTATTGATCGGTTTCATGCA	27
20	Sentido derecho	CD68	NM_001251	S0067/CD68.f2	TGGTCCCAGCCCTGTGT	18
	Sentido contrario	CD68	NM_001251	S0069/CD68.r2	CTCCTCCACCCTGGGTTGT	19
	Sonda	CD68	NM_001251	S4734/CD68.p2	CTCCAAGCCCAGATTTCAGATTTCGAGTCA	28
	Sentido derecho	SCUBE2	NM_020974	S1494/SCUBE2.f2	TGACAATCAGCACACCTGCAT	21
25	Sentido contrario	SCUBE2	NM_020974	S1495/SCUBE2.r2	TGTGACTACAGCCGTGATCCTTA	23
	Sonda	SCUBE2	NM_020974	S4735/SCUBE2.p2	CAGGCCCTCTTCCGAGCGGT	20
	Sentido derecho	CTSL2	NM_001333	S4354/CTSL2.f1	TGTCCTACTGAGCGAGCAGAA	21
	Sentido contrario	CTSL2	NM_001333	S4355/CTSL2.r1	ACCATTGCAGCCCTGATTG	19
30	Sonda	CTSL2	NM_001333	S4356/CTSL2.p1	CTTGAGGACGCGAACAGTCCACCA	24
	Sentido derecho	ESR1	NM_000125	S0115/EstR1.f1	CGTGGTGCCCCTCTATGAC	19
	Sentido contrario	ESR1	NM_000125	S0117/EstR1.r1	GGCTAGTGGGCGCATGTAG	19
	Sonda	ESR1	NM_000125	S4737/EstR1.p1	CTGGAGATGCTGGACGCC	19
35	Sentido derecho	GAPD	NM_002046	S0374/GAPD.f1	ATTCCACCCATGGCAAATTC	20
	Sentido contrario	GAPD	NM_002046	S0375/GAPD.r1	GATGGGATTTCCATTGATGACA	22
	Sonda	GAPD	NM_002046	S4738/GAPD.p1	CCGTTCTCAGCCTTGACGGTGC	22
	Sentido derecho	GRB7	NM_005310	S0130/GRB7.f2	CCATCTGCATCCATCTTGTT	20
40	Sentido contrario	GRB7	NM_005310	S0132/GRB7.r2	GGCCACCAGGGTATTATCTG	20
	Sonda	GRB7	NM_005310	S4726/GRB7.p2	CTCCCCACCCTTGAGAAGTGCCT	23
	Sentido derecho	GSTM1	NM_000561	S2026/GSTM1.r1	GGCCCAGCTTGAATTTTTCA	20
	Sentido contrario	GSTM1	NM_000561	S2027/GSTM1.f1	AAGCTATGAGGAAAAGAAGTACACGAT	27
	Sonda	GSTM1	NM_000561	S4739/GSTM1.p1	TCAGCCACTGGCTTCTGTCATAATCAGGAG	30
45	Sentido derecho	GUSB	NM_000181	S0139/GUS.f1	CCCCTCAGTAGCCAAGTCA	20
	Sentido contrario	GUSB	NM_000181	S0141/GUS.r1	CACGCAGGTGGTATCAGTCT	20
	Sonda	GUSB	NM_000181	S4740/GUS.p1	TCAAGTAAACGGGCTGTTTTCAAACA	27
	Sentido derecho	ERBB2	NM_004448	S0142/HER2.f3	CGGTGTGAGAAGTGCAGCAA	20
50	Sentido contrario	ERBB2	NM_004448	S0144/HER2.r3	CCTCTCGCAAAGTGCTCCAT	19
	Sonda	ERBB2	NM_004448	S4729/HER2.p3	CCAGACCATAGCACACTCGGGCAC	24
	Sentido derecho	MKI67	NM_002417	S0436/MKI67.f2	CGGACTTTGGGTGCGACTT	19
	Sentido contrario	MKI67	NM_002417	S0437/MKI67.r2	TTACAACCTTCCACTGGGACGAT	24
55	Sonda	MKI67	NM_002417	S4741/MKI67.p2	CCACTTGTCGAACCACCGCTCGT	23
	Sentido derecho	MYBL2	NM_002466	S3270/MYBL2.f1	GCCGAGATCGCCAAGATG	18
	Sentido contrario	MYBL2	NM_002466	S3271/MYBL2.r1	CTTTTGATGGTAGAGTCCAGTGATTC	27
	Sonda	MYBL2	NM_002466	S4742/MYBL2.p1	CAGCATTGTCTGTCTCCCTGGCA	24
	Sentido derecho	PGR	NM_000926	S1336/PR.f6	GCATCAGGCTGTCATTATGG	20
	Sentido contrario	PGR	NM_000926	S1337/PR.r6	AGTAGTTGTGCTGCCCTTCC	20
	Sonda	PGR	NM_000926	S4743/PR.p6	TGTCCTTACCTGTGGGAGCTGTAAGGTC	28
	Sentido derecho	RPLP0	NM_001002	S0258/RPLP0.f2	CCATTCTATCATCAACGGGTACAA	24
	Sentido contrario	RPLP0	NM_001002	S0258/RPLP0.r2	TCAGCAAGTGGGAAGGTGTAATC	23

ES 2 778 851 T3

	Sonda	RPLP0	NM_001002	S4744/RPLP0.p2	TCTCCACAGACAAGGCCAGGACTCG	25
	Sentido derecho	STK6	NM_003600	S0794/STK6.f2	CATCTTCCAGGAGGACCACT	20
	Sentido contrario	STK6	NM_003600	S0795/STK6.r2	TCCGACCTTCAATCATTCA	20
5	Sonda	STK6	NM_003600	S4745/STK6.p2	CTCTGTGGCACCCCTGGACTACCTG	24
	Sentido derecho	MMP11	NM_005940	S2067/MMP11.f3	CCTGGAGGCTGCAACATACC	20
	Sentido contrario	MMP11	NM_005940	S2068/MMP11.r3	TACAATGGCTTTGGAGGATAGCA	23
	Sonda	MMP11	NM_005940	S4746/MMP11.p3	ATCCTCCTGAAGCCCTTTTCGCAGC	25
10	Sentido derecho	BIRC5	NM_001168	S0259/BIRC5.f2	TGTTTTGATTCCCGGGCTTA	20
	Sentido contrario	BIRC5	NM_001168	S0261/BIRC5.r2	CAAAGCTGTCAGCTCTAGCAAAAG	24
	Sonda	BIRC5	NM_001168	S4747/BIRC5.p2	TGCCTTCTTCCCTCCCTCACTTCTCACCT	28
	Sentido derecho	TFRC	NM_003234	S1352/TFRC.f3	GCCAACTGCTTTCATTTGTG	20
15	Sentido contrario	TFRC	NM_003234	S1353/TFRC.r3	ACTCAGGCCCATTTCTTTA	20
	Sonda	TFRC	NM_003234	S4748/TFRC.p3	AGGGATCTGAACCAATACAGAGCAGACA	28

Tabla 4

Gen	Número de acceso	Secuencia
ACTB	NM_001101	CAGCAGATGGGATCAGCAAGCAGGAGTATACGAGTCCGGCCCTCCATCGTCCACCGCAATGC
BAG1	NM_004323	CGTTGTACGCACTTGGAAATACAAAGATGGTCCGGGTCATGTTAATGGGMAAAAGAACAGTCCACAGGAAGAGGTGAAC
BCL2	NM_000833	CAGATGACCTAGTACCACCTGAGATTCCACGCGGACGCGATGGGAAAATGCCCTTAATCATAGG
CCNB1	NM_031966	TTCAAGGTGTTGCAGGAGACCATGTACATGACTGCTCCATATTGATC66TTCAJGCGAATAATTGTGTGCCCAAGAAGATG
CD68	NM_001251	TGGTTCACAGCCCTGTGCACCTCCAA9CCAGATTCAGATTCGAGTCATGTACACAAACCAGGGTGGAGGAG
SCUBE2	NM_020974	TGACAAATCAGCACACCTGCAATTCACCGCTCGGMAAGGEGGCCGAGCTGCATGAATAAGGATCACGGCTGTAGTCACA
CTS1L2	NM_004333	TGTCTCACTGAGCGAGCAGAACTGTGCGACTGTTCCGCTCCTCAAGGCAATCAGGGCTGCAATGGT
ESR1	NM_000125	CGTGGTGC00CCTCTATGACCTGCTGCTGGAGATGCTG9ACGCCCACCCGCTACATGCGCCACTAGCC
GAPD	NM_002046	ATTCCACCCATGCGAAATTCCATGECACCGTCAAGGCTGAGAA9GGGAAGCTTGTTCATCAATGGAATCCCATC
GRB7	NM_005310	CCATCTGCATCCATCTTTG66CTCCACCCCTTGAGAA6TGCCTCAGATAATACCCCTGGTGGCC
GSTM1	NM_000561	AAGCTATGAGGAAAAGANGTACACGATGGGGAGCGCTCGTGAATATGACAGAAGCCAGTGGCTGAATGMAAAATTCAAGCTGGCC
GLUS	NM_000181	CCCACCTCAGTAGCCAAAGTCACAATGTTTGGAAAACAGCCCGTCTTACTTGAGCAAGACTGATACACCTGCCGTG
ERBB2	NM_004448	CSGTGTGAGAAGTGCAGCAAGCCCTGTGCCAGTGTGCTATGGTGTGEECATGGACCTTGCAGAGG
MKI67	NM_002417	CEGACTTTGGGTGCCACTTEACGAGCCGTG6TTCGACAAGTGGCC0TTGGGGCCGGATCGTCCAGTGGAAAGTTGTAA
MYBL2	NM_002466	GC0GAGATCGCCAMGAT6TTC0CAGGGGAGCAGACAAATGCTGTGAAGAATCAGTEGAATCTACCATCAAAAG
PGR	NM_000926	GCATCAGGCTGTCAATTATGGTGTCTTACCTGJGGAGCTGTAA9GTCTTTTAAAGGGGCAATGGAAAGGCGCACAACA
RPLP0	NM_001002	CCATTCATCATCAACGGGTACAACAGTCCCTGGCCCTTGTCTGTGGAGACGGATACACCTTCCACTTGGCTGA
STK6	NM_003600	CATCTCCAGGAGGACCACCTCTGTGTGGCACCCTGGACTACCTGCCCCCTGMAAATGATTGAAGGT00GA
MMP11	NM_005940	CCTGGAGGCTGCACATACCCTCATCTGTCAGCCGEGATCCTCTGAAAGCCCTTTTCGACGACTGCTATCCTCCAAGCCCAFTGTA
BIRC5	NM_001168	TGTTTGTATCCCGEGCTTACCAGGTGAGAGTGGGGAGGAGGAGGEGCAGTGTCCCTTTTGTGTAGAGCTGACAGCTTTG
TFR3	NM_008234	GCCAACTGCCTTTCATTTGTGAGGGATCTGAACCAATACAGAGCAGACATAAAGGAAATGGGCCTGAGT

REIVINDICACIONES

1. Un procedimiento para predecir la probabilidad de una respuesta beneficiosa a la quimioterapia de un ser humano diagnosticado con cáncer de mama, que comprende

5 (a) determinar cuantitativamente, en una muestra biológica que comprende células de cáncer de mama obtenidas de dicho sujeto, el valor de una o más de las siguientes variables:

(i) una Puntuación de Recurrencia (RS),
 en el que la RS se calcula calculando primero una Puntuación de Recurrencia sin cerrar (RS_U):

$$RS_U = 0,47 \times \text{Puntuación Umbral del Grupo GRB7} - 0,34 \times \text{Puntuación del Grupo ESR1} + 1,04 \times \text{Puntuación Umbral del Grupo de Proliferación} + 0,10 \times \text{Puntuación del Grupo de Invasión} + 0,05 \times CD68 - 0,08 \times GSTM1 - 0,07 \times BAG1$$

en el que:

$$\text{La Puntuación del Grupo GRB7} = 0,9 \times GRB7 + 0,1 \times ERBB2$$

y la Puntuación Umbral del Grupo GRB7 = 8 si la puntuación del grupo GRB7 es <8 e = la puntuación del grupo GRB7 si la puntuación del grupo GRB7 es ≥8;

$$\text{La Puntuación del Grupo ESR1} = (ESR1 + PGR + BCL2 + SCUBE2)/4;$$

$$\text{La Puntuación del Grupo de Proliferación} = (BIRC5 + MK167 + MYBL2 + CCNB1 + STK6)/5;$$

y la Puntuación Umbral del Grupo de Proliferación = 6,5 si la Puntuación del Grupo de Proliferación es <6,5 e = la Puntuación del Grupo de Proliferación si la Puntuación del Grupo de Proliferación es ≥6,5; y

$$\text{La Puntuación del Grupo de Invasión} = (CTSL2 + MMP11)/2$$

y posteriormente calculando el RS escalado:

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

(ii) Puntuación del Grupo ESR1,
 en el que la puntuación del Grupo ESR1 = (ESR1 + PGR + BCL2 + SCUBE2)/4;

(iii) Puntuación del Grupo de Invasión,
 en el que la puntuación del grupo de invasión = CTSL2 + MMP11)/2;

(iv) Puntuación Umbral del Grupo de Proliferación,
 en el que la Puntuación del Grupo de Proliferación = (BIRC5 + MK167 + MYBL2 + CCNB1 + STK6)/5

y la Puntuación Umbral del Grupo de Proliferación = 6,5 si la Puntuación del Grupo de Proliferación es <6,5 e = la Puntuación del Grupo de Proliferación si la Puntuación del Grupo de Proliferación es ≥6,5; y

(v) el nivel de expresión del transcrito de ARN de al menos uno de MYBL2 y SCUBE2, y (b) determinar la probabilidad de una respuesta beneficiosa a la quimioterapia del sujeto, en el que

(b1) por cada unidad de un aumento en el valor de uno o más de (i), (iii), (iv), o el nivel de expresión del transcrito de ARN de MYBL2, se identifica a dicho sujeto con una probabilidad proporcionalmente mayor de tener una respuesta beneficiosa a dicha quimioterapia; y

5 (b2) por cada unidad de un aumento en el valor de (ii) o el nivel de expresión del transcrito de ARN de SCUBE2, se identifica a dicho sujeto con una probabilidad proporcionalmente disminuida de tener una respuesta beneficiosa a la quimioterapia;

10 en el que los símbolos de genes en las ecuaciones representan los niveles de expresión de los transcritos de ARN de los respectivos genes medidos por reacción en cadena cuantitativa de la transcriptasa inversa-polimerasa (qRT-PCR) y expresados en unidades de ciclo umbral (C_T) normalizadas en relación con los niveles de expresión de uno o más ARNm de referencia.

15 2. El procedimiento de la reivindicación 1, en el que uno o más ARNm de referencia se seleccionan del grupo que consiste en ACTB, GAPD, GUSB, RPLPO y TFRC.

3. El procedimiento de la reivindicación 1 o 2, en el que el cáncer de mama se selecciona del grupo que consiste en cáncer de mama invasivo, cáncer de mama en etapa II, cáncer de mama en etapa III y cáncer de mama positivo a ESR1.

20 4. El procedimiento de cualquier reivindicación anterior, en el que la muestra biológica comprende tejido fijo embebido en parafina o tejido congelado obtenido mediante biopsia con aguja fina, biopsia central, aspiración con aguja fina, lavado bronquial o biopsia transbronquial.

25 5. El procedimiento de cualquier reivindicación anterior, que comprende la determinación de al menos dos de dichas variables.

6. El procedimiento de la reivindicación 5, que comprende la determinación de al menos tres de dichas variables.

30 7. El procedimiento de la reivindicación 6, que comprende la determinación de al menos cuatro de dichas variables.

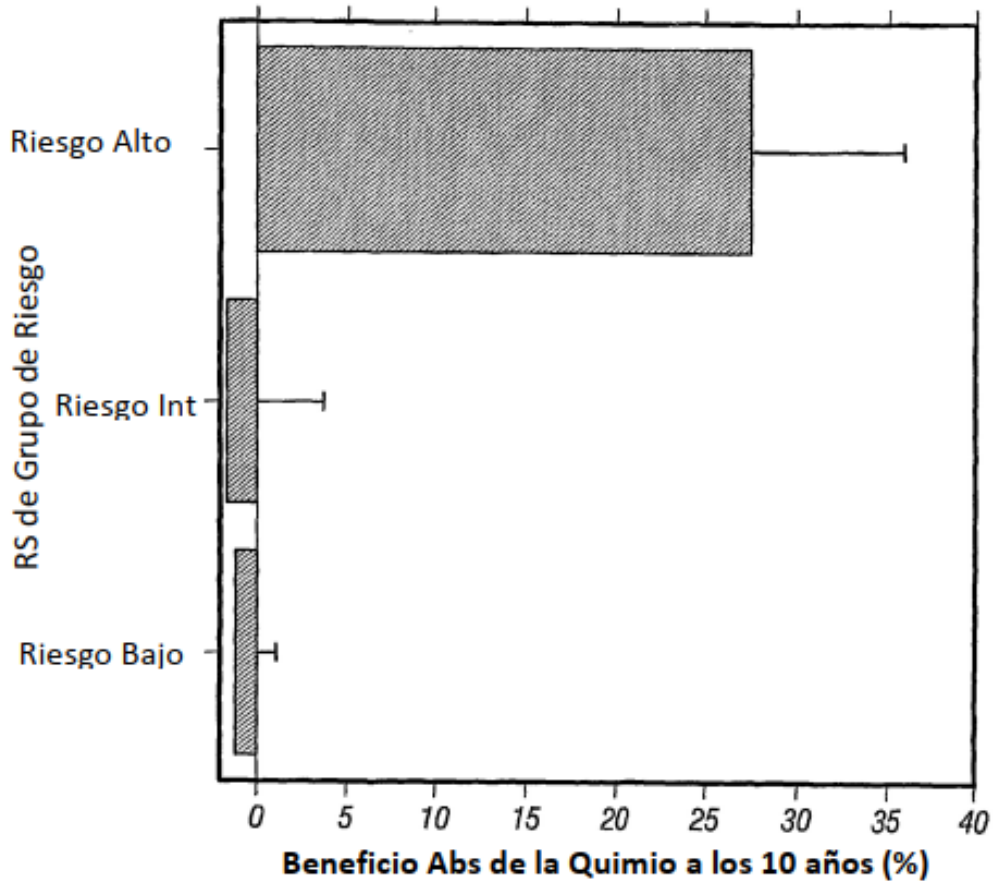


Figura 1

**Estudio NSABP B-20: Beneficio Absoluto de la Quimioterapia como
Función de la Puntuación de Recurrencia (Media del Beneficio \pm e.e.)**

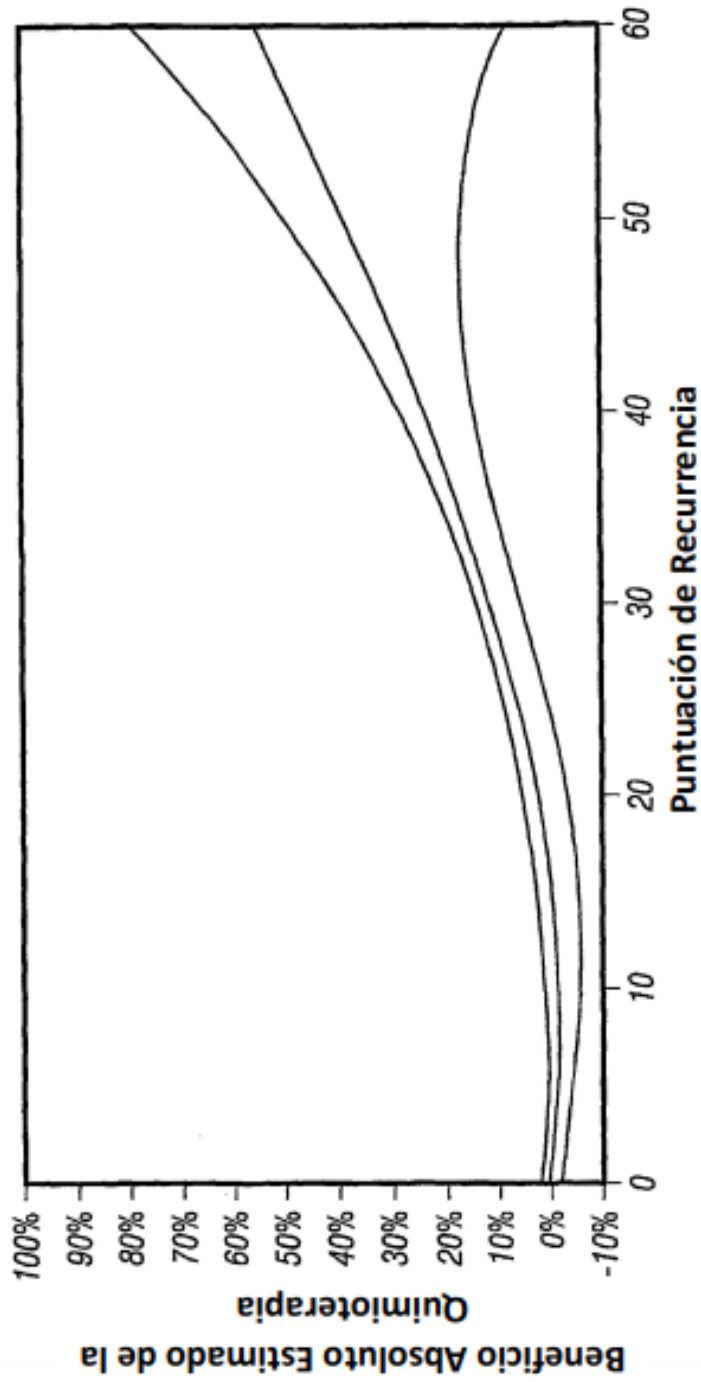


Figura 2