

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 667 547**

51 Int. Cl.:

A61K 48/00 (2006.01)

A61K 31/7105 (2006.01)

A61K 31/7088 (2006.01)

A61P 35/00 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **09.12.2011 PCT/US2011/064290**

87 Fecha y número de publicación internacional: **14.06.2012 WO12079059**

96 Fecha de presentación y número de la solicitud europea: **09.12.2011 E 11847205 (9)**

97 Fecha y número de publicación de la concesión europea: **21.02.2018 EP 2648762**

54 Título: **Pronóstico de cáncer de mama después del tratamiento**

30 Prioridad:

09.12.2010 US 421627 P

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
11.05.2018

73 Titular/es:

BIOETHERANOSTICS, INC. (50.0%)
9640 Towne Centre Drive Suite 200
San Diego, CA 92121, US y
THE GENERAL HOSPITAL CORPORATION
(50.0%)

72 Inventor/es:

SGROI, DENNIS;
ERLANDER, MARK G.;
ZHANG, YI y
SCHNABEL, CATHERINE A

74 Agente/Representante:

PONS ARIÑO, Ángel

ES 2 667 547 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

Pronóstico de cáncer de mama después del tratamiento

5 **Campo de la divulgación**

La divulgación se refiere a la identificación y uso de perfiles, o patrones, de expresión génica con relevancia clínica para el cáncer de mama. En particular, la divulgación se basa en parte en la identificación de genes que se expresan en correlación con la probabilidad de recidiva de cáncer después de un tratamiento inicial con un inhibidor de aromatasa u otra terapia endocrina. Los niveles de expresión génica forman un índice molecular que es capaz de predecir un desenlace clínico, así como el pronóstico, para un paciente después del tratamiento inicial con un inhibidor de aromatasa u otra terapia endocrina.

Los perfiles de expresión génica, representados tanto en formatos de expresión de ácido nucleico, expresión de proteínas, o en otros formatos de expresión, pueden utilizarse para predecir el desenlace clínico posterior al tratamiento de sujetos que padecen cáncer de mama, predecir la recidiva de cáncer y/o predecir la aparición de cáncer metastásico. Los perfiles también pueden utilizarse en el estudio de un pronóstico de sujetos. Cuando se utiliza para un pronóstico, los perfiles se utilizan para determinar el tratamiento del cáncer basándose en la probabilidad de esperanza de vida, recidiva de cáncer y/o metástasis de cáncer.

20 **Antecedentes de la divulgación**

El tratamiento del cáncer de mama ha sido un campo muy estudiado y que ha suscitado mucho interés. Después de un diagnóstico inicial de cáncer de mama mediante el análisis de una muestra de células de cáncer de mama de un sujeto, los métodos de tratamiento a menudo comienzan con la extirpación quirúrgica de las células tumorales. En los casos de cáncer de mama dependiente de hormonas, tales como cáncer de mama positivo a receptores de estrógenos (RE+), la cirugía va seguida de un antagonismo de estrógenos para reducir el crecimiento o recrecimiento tumoral. En muchos casos, el tratamiento con tamoxifeno antiestrogénico se utiliza durante cinco años para reducir el riesgo de recidiva de la enfermedad y por lo tanto la mortalidad mediada por cáncer de mama.

Lamentablemente, los datos del campo indican que más de la mitad de todas las recidivas del cáncer de mama se producen después de cinco años de tratamiento con tamoxifeno adyuvante.

Goss et al. (J. Clin. Oncol., 26(12):1948-1955, 2008) comunican resultados de un ensayo clínico en el que se examina del uso de letrozol iniciado tres meses después de cinco años de tratamiento con tamoxifeno adyuvante en sujetos con cáncer de mama RE+ primario. Los resultados sugieren que el tratamiento con letrozol después de tamoxifeno mejora la supervivencia sin cáncer de mama y la supervivencia sin cáncer de mama distante.

Pero Goss et al. no proporcionan medios para predecir qué sujetos, tratados durante cinco años con tamoxifeno, podrían beneficiarse de un tratamiento posterior con letrozol. Por lo tanto, no había medios para dirigir un tratamiento con letrozol solo a los sujetos para los que se espera un beneficio. De este modo, el tratamiento con letrozol se aplicó a sujetos para los cuales no se esperaba ningún beneficio, lo que dio como resultado un tratamiento excesivo de la población de sujetos sin cáncer de mama tratados durante cinco años con tamoxifeno.

Las citas de los documentos de la presente memoria no deben interpretarse como que reflejan una admisión de que cualquiera es relevante en la técnica anterior. Además, sus citas no son indicativas de una búsqueda de divulgaciones relevantes. Todas las indicaciones con respecto a los datos o contenidos de los documentos se basan en información disponible y no es una admisión de su precisión o corrección.

50 **Breve resumen de la divulgación**

La divulgación se basa en parte en el descubrimiento y la determinación de niveles de expresión génica en células tumorales de cáncer de mama, que están correlacionados con un cambio beneficioso en la quimioterapia contra el cáncer de mama. En algunos casos, el cambio es de una forma de terapia endocrina a otra. Los niveles de expresión se pueden usar para proporcionar información pronóstica, como la recidiva del cáncer, e información predictiva, como la capacidad de respuesta a ciertas terapias. Por lo tanto, la invención proporciona:

- un método para determinar si se espera que un sujeto que se ha sometido a la extirpación de un cáncer de mama RE+ y que se ha tratado con una primera terapia endocrina durante un periodo de tiempo sin recidiva de cáncer, se beneficie de un cambio a una segunda terapia endocrina distinta, que comprende: preparar ADNc a partir de ácidos nucleicos de una muestra de células de cáncer de mama RE+ del sujeto, determinar el nivel de expresión del gen *HoxB13* a partir de dicho ADNc, y clasificar al sujeto como que se espera que se beneficie del tratamiento con una segunda terapia endocrina distinta después de suspender la primera terapia endocrina, en el que dicha clasificación está basada en el nivel de expresión elevado de *HoxB13*, en el que (i) la primera terapia endocrina es un tratamiento con un modulador selectivo de receptores de estrógeno (SERM) o un regulador negativo selectivo de receptores de estrógeno (SERD) y la segunda terapia endocrina es un tratamiento con un

inhibidor de aromatasas (IA), o (ii) la primera terapia endocrina es un tratamiento con un IA y la segunda terapia endocrina es un tratamiento con un SERM o un SERD; y

- una segunda terapia endocrina para su uso en un método de tratamiento de cáncer de mama RE+ en un sujeto que se ha sometido a extirpación de cáncer de mama RE+ y que se ha tratado con una primera terapia endocrina durante un periodo de tiempo sin recidiva de cáncer, en el dicho método comprende clasificar al sujeto como que se espera que se beneficie de una segunda terapia endocrina de acuerdo con el método descrito anteriormente, y en el que:

(i) la primera terapia endocrina es un tratamiento con un modulador selectivo de receptores de estrógeno (SERM) o un regulador negativo selectivo de receptores de estrógeno (SERD) y la segunda terapia endocrina es un tratamiento con un inhibidor de aromatasas (IA); o

(ii) la primera terapia endocrina es un tratamiento con un IA y la segunda terapia endocrina es un tratamiento con un SERM o un SERD.

En el presente documento se desvelan métodos para identificar, o clasificar, una población de sujetos inicialmente tratados con una terapia con antiestrógenos o antiaromatasa, en al menos dos subpoblaciones. Se esperaría que una primera subpoblación se beneficiase de un cambio en la terapia, tal como un cambio a otra terapia distinta con antiestrógenos o antiaromatasa. No se esperaría que una segunda subpoblación se beneficiase. En algunos casos, la terapia inicial es con tamoxifeno, tal como una terapia con tamoxifeno adyuvante durante un periodo de aproximadamente cinco años o menor. Opcionalmente, el cambio es a letrozol, o a otra terapia con antiaromatasa. La divulgación incluye medios para una población de sujetos tratados de esta manera, y con tratamiento durante la enfermedad sin cáncer de mama, para clasificarse en la primera y/o la segunda subpoblación.

Los métodos de la invención se basan en el nivel de expresión de HoxB13 en células de cáncer de mama RE+ de un sujeto. En algunas realizaciones, puede utilizarse una proporción de dos genes de expresión de HoxB13 con respecto a la expresión de IL17BR (o proporción HoxB13: IL17BR) (véase Ma et al., J. Clin. Oncol., 24:4611-9 (2006)). En realizaciones alternativas, puede utilizarse una proporción de dos genes de HoxB13 con respecto a la expresión de CHDH.

La proporción HoxB13:IL17BR (H: I) se descubrió basándose en un estudio de nuevos biomarcadores predictivos de un desenlace clínico más allá de los factores de pronóstico convencionales. Los pacientes que desarrollaron recidivas de cáncer se emparejaron con los que no con respecto a un estadio o grado tumoral. Se descubrió que la simple proporción H: I era adecuada para predecir la recidiva de cáncer en pacientes con cáncer de mama positivos a receptores de estrógeno (RE+) que recibían terapia con tamoxifeno adyuvante. Estudios posteriores (Goetz et al., Clin Cancer Res. 12:2080-7 (2006); Jerevall et al., Breast Cancer Res. Treat 107(2):225-34 (2007); Jansen et al., J. Clin.Oncol. 25:662-8 (2007)) han demostrado también en ensayos clínicos tanto retrospectivos como aleatorizados, que la proporción es tanto de pronóstico, tal como que es un indicador de agresividad tumoral, como predictiva del beneficio de tamoxifeno.

En realizaciones adicionales, la invención incluye uno o más genes adicionales en combinación con la expresión de HoxB13. La combinación puede ser con cualquiera de uno, dos, tres, cuatro o los cinco genes que se desvelan adicionalmente a continuación.

Los genes adicionales de la divulgación codifican Bub1B (gemación no inhibida por benzimidazoles 1 beta) o proteína quinasa 6 activada por p21 (PAK6); CENPA (proteína A del centrómero, isoforma a); NEK2 (quinasa 2 relacionada con NIMA o quinasa 2 relacionada con "un gen que nunca entra en mitosis"); RACGAP1 (proteína 1 activadora de Rac GTPasa) y RRM2 (ribonucleótido reductasa M2). En el presente documento el uso de solo estos cinco genes se denomina índice de grado molecular (IGM). Los aspectos de la divulgación incluyen composiciones y métodos como se describen para el uso de la expresión de HoxB13, con o sin expresión de IL17BR en combinación con uno o más niveles de expresión de uno o más de los cinco genes anteriores para estudiar, proporcionar información de pronóstico y/o proporcionar predicciones de respuesta clínica.

Por tanto, la divulgación se basa en parte en el descubrimiento de que el nivel o los niveles de expresión génicos son útiles para proporcionar determinaciones de pronóstico (tal como la probabilidad de recidiva de cáncer en forma de recidiva de cáncer de mama, bien local o distalmente o en forma de metástasis) y determinaciones predictivas (tales como respuesta a un ciclo de tratamiento) para un sujeto. El uso de los siete genes desvelados se denomina Índice de Cáncer de Mama (ICM).

Cuando los niveles de expresión del ICM se analizaron utilizando la reacción en cadena de la polimerasa con transcripción inversa en tiempo real (RT-PCR), se descubrió que la combinación proporcionaba una estadificación superior de riesgo de recidiva en sujetos tratados durante cinco años con terapia de tamoxifeno. Esto refleja un descubrimiento inesperado porque identifica por primera vez un indicador para realizar un cambio beneficioso de las terapias del cáncer de mama.

En aspectos adicionales, la expresión de HoxB13 y/o el ICM, pueden usarse para predecir la recidiva tardía de cáncer en un paciente con cáncer de mama. Como ejemplos no limitantes de recidiva tardía se incluye cinco años después

de tratamiento con tamoxifeno, pero también se incluye 4 años después, 3 años después o 2 años después, o menos tiempo de tratamiento con tamoxifeno. De manera similar, la expresión de Hob13 y/o el ICM, pueden usarse para predecir la respuesta a letrozol o a otra terapia con antiestrógenos o antiaromatasa después de los periodos de tiempo indicados anteriormente para inhibir la recidiva tardía.

5 Las realizaciones de la divulgación incluyen un método de ensayo con valor pronóstico y valor predictivo para estadificar sujetos con cáncer de mama RE+ original y posterior tratamiento sin cáncer de mama. Como un pronóstico, la estadificación puede basarse en niveles de expresión diferenciales que se correlacionan con, y por lo tanto indican, la necesidad de un cambio en las terapias del cáncer de mama como un ejemplo no limitante. Como un ejemplo no limitante, la estadificación (basada en niveles de expresión) puede utilizarse para predecir la sensibilidad endocrina (tal como sensibilidad a letrozol como un ejemplo no limitante) y/o la predicción del beneficio de los inhibidores antiestrógeno y/o antiaromatasa. La detección de la expresión génica puede ser, por supuesto, en cualquier muestra adecuada que contenga células, como se describe en el presente documento. Como ejemplos no limitantes de células para su uso en la divulgación se incluyen las células recién aisladas del sujeto, las congeladas después del aislamiento y las que se fijan y/o incluyen, tales como las que se fijan en formalina y se incluyen en parafina (FFPE, por las siglas *formalin-fixed, paraffin-embedded*). En la mayoría de las realizaciones, las células son células de mama, tales como células de cáncer de mama.

20 En algunas realizaciones, ventajosamente, se utiliza un método basado en los niveles de expresión en una muestra de un sujeto que contiene células de cáncer de mama, tal como una muestra de carcinoma ductal *in situ* (DCIS). Como un ejemplo no limitante, la célula puede proceder de una muestra histológica preoperatoria utilizada para diagnosticar el cáncer en el sujeto. Para dicho sujeto, el tratamiento de referencia es la cirugía, prefiriéndose la cirugía conservadora de mama sobre una mastectomía radical para extirpar el DCIS. Esto a menudo viene seguido de una radioterapia posoperatoria, opcionalmente con terapia endocrina, tal como tratamiento con tamoxifeno, un modulador selectivo de receptores de estrógeno (SERM), un regulador negativo selectivo de receptores de estrógeno (SERD), o un inhibidor de aromatasa (IA), tal como letrozol. En otros casos posoperatorios, la terapia endocrina se administra sin radiación, y opcionalmente, con quimioterapia.

30 La presente divulgación se refiere a la identificación de un sujeto que se espera que se beneficie de un cambio en la terapia endocrina, tal como de un tipo de terapia endocrina a otro, después de una supervivencia sin cáncer de mama durante el ciclo de la terapia endocrina inicial. En realizaciones adicionales, el cambio puede realizarse después de un ciclo inicial de terapia endocrina durante 5 años, 4 años, 3 años o 2 años. La divulgación también incluye detectar la expresión génica, en la que una expresión alta de HoxB13 es un indicador de una probabilidad aumentada de recidiva de cáncer en el sujeto después de una terapia endocrina inicial, tal como terapia adyuvante con tamoxifeno. Por lo tanto, los métodos pueden incluir la identificación del sujeto como probable, o poco probable, de experimentar recidiva local de cáncer, e incluir además un cambio en las modalidades de tratamiento en el sujeto para abordar un desenlace esperado. Como un ejemplo no limitante, la determinación de una probabilidad de recidiva en ausencia de un tratamiento posterior inicial, prolongado, la terapia puede utilizarse para confirmar la idoneidad de, o seleccionar, una terapia prolongada con un cambio en la modalidad antiestrógeno y/o antiaromatasa utilizada.

45 En algunos casos, los métodos desvelados pueden utilizarse para seleccionar o eliminar terapias en mujeres premenopáusicas o postmenopáusicas, que se han sometido a un tratamiento con terapia endocrina y permanecen sin cáncer durante ese tiempo. Las mujeres premenopáusicas incluyen las que tienen menos de aproximadamente 35 años. El método puede incluir analizar una muestra de un sujeto que contenga células de cáncer de mama para determinar la expresión de los genes desvelados. Como un ejemplo no limitante, la célula puede ser una de una muestra histológica preoperatoria utilizada para el diagnóstico del cáncer en el sujeto.

50 Como ejemplos no limitantes de terapia endocrina se incluyen el tratamiento con un SERM, tal como tamoxifeno, o un SERD, o inhibidor de aromatasa (IA). Como ejemplos no limitantes de un IA se incluyen inhibidores no esteroideos tales como letrozol y anastrozol e inhibidores esteroideos irreversibles tal como exemestano.

Descripción detallada de modos de realización práctica de la divulgación

Definiciones de términos utilizados en el presente documento:

60 Un "patrón" o "perfil" o "firma" de expresión génica se refiere a la expresión relativa de uno o más genes entre dos o más resultados clínicos, resultados de cáncer, recidiva de cáncer y/o resultados de supervivencia que se correlacionan con ser capaz de diferenciar entre dichos resultados. En algunos casos, el resultado es el cáncer de mama.

65 Un "gen" es un polinucleótido que codifica un producto distinto, ya sea ARN o proteico en naturaleza. Se aprecia que más de un polinucleótido puede ser capaz de codificar un producto distinto. El término incluye alelos y polimorfismos de un gen que codifica el mismo producto, o un análogo de mismo funcionalmente asociado (incluyendo ganancia, pérdida o modulación de función) del mismo, basándose en la localización cromosómica y en la capacidad de recombinar durante la mitosis normal.

Los términos “correlacionarse” o “correlación” o equivalentes de los mismos, se refieren a una asociación entre la expresión de uno o más genes y un estado fisiológico de una célula con respecto a la exclusión de uno o más estados distintos identificados utilizando los métodos descritos en el presente documento. Un gen puede expresarse a un nivel más alto o más bajo y aun correlacionarse con uno o más estados o resultados de cáncer.

5 Un “polinucleótido” es una forma polimérica de nucleótidos de cualquier longitud, o bien ribonucleótidos o desoxirribonucleótidos. El término se refiere solo a la estructura primaria de la molécula. Por tanto, este término incluye ADN y ARN mono y bicatenario. Esto también incluye tipos de modificaciones conocidas que incluyen marcadores conocidos en la técnica, metilación, “caperuzas”, sustitución de uno o más de los nucleótidos de origen natural con un análogo, y modificaciones internucleotídicas tales como enlaces no cargados (por ejemplo, fosforotioatos, fosforditioatos, etc.), así como formas no modificadas del polinucleótido.

15 El término “amplificar” se utiliza en sentido amplio que significa crear un producto de amplificación que puede realizarse enzimáticamente con ADN o ARN polimerasas. “Amplificación”, como se usa en el presente documento, se refiere generalmente al proceso de producir múltiples copias de una secuencia deseada, particularmente la de una muestra. “Múltiples copias” significa al menos dos copias. Una “copia” no significa necesariamente una secuencia perfecta complementaria o idéntica a la secuencia molde.

20 Por correspondiente se entiende que una molécula de ácido nucleico comparte una cantidad sustancial de identidad de secuencia con otra molécula de ácido nucleico. Una cantidad sustancial significa al menos 95%, normalmente al menos 98% y más normalmente al menos 99%, y la identidad de secuencia se determina utilizando el algoritmo BLAST, como se describe en Altschul et al., J. Mol. Biol. 215:403-410 (1990) (utilizando la configuración por defecto publicada, es decir los parámetros $w=4$, $t=17$). En la técnica se conocen generalmente métodos para amplificar ARNm, e incluyen PCR con transcripción inversa (RT-PCR) y los descritos en la solicitud de patente de Estados Unidos 10/062.857 (presentada el 25 de octubre de 2001), así como en las solicitudes de patente provisionales de Estados Unidos 60/298.847 (presentada el 15 de junio de 2001) y 60/257.801 (presentada el 22 de diciembre de 200). Otro método que puede utilizarse es la PCR cuantitativa (o PCRc). Como alternativa, el ARN puede marcarse directamente como ADNc correspondiente mediante métodos conocidos en la técnica.

30 Una “micromatriz” es una matriz lineal o bidimensional de regiones preferentemente distintas, teniendo cada una de ellas una zona definida, formada en la superficie de un soporte sólido tal como, pero sin limitación, una membrana de vidrio, de plástico o sintética. La densidad de las distintas regiones en una micromatriz se determina mediante los números totales de polinucleótidos inmovilizados a detectar sobre la superficie de un solo soporte en fase sólida, preferentemente al menos aproximadamente 50/cm², más preferente al menos aproximadamente 100/cm², incluso más preferentemente al menos aproximadamente 500/cm², aunque preferentemente por debajo de aproximadamente 1.000/cm². Preferentemente, las matrices contienen menos de aproximadamente 500, aproximadamente 1000, aproximadamente 1500, aproximadamente 2000, aproximadamente 2500 o aproximadamente 3000 polinucleótidos inmovilizados en total. Como se usa en el presente documento, una micromatriz de ADN es una matriz de oligonucleótidos o polinucleótidos colocados en una microplaca o en otras superficies utilizadas para hibridar con polinucleótidos amplificados o clonados de una muestra. Dado que la se conoce la posición de cada grupo de cebadores particular en la matriz, las identidades de una muestra de polinucleótidos pueden determinarse basándose en su unión en una posición particular en la micromatriz.

45 Dado que la divulgación se basa en la identificación de genes que están sobre- o infra- expresados, una realización de la divulgación implica la determinación de la expresión por hibridación de ARNm, o una versión amplificada o clonada del mismo, de una célula de muestra con un polinucleótido que es exclusivo en una secuencia génica particular. Los polinucleótidos preferidos de este tipo contienen al menos aproximadamente 20, al menos aproximadamente 22, al menos aproximadamente 24, al menos aproximadamente 26, al menos aproximadamente 28, al menos aproximadamente 30 o al menos aproximadamente 32 pares de bases consecutivos de una secuencia génica que no se encuentra en otras secuencias génicas. El término “aproximadamente”, como se usa en la oración anterior, se refiere a un aumento o disminución de 1 desde el valor numérico indicado. Son incluso más preferidos los polinucleótidos de al menos o de aproximadamente 50, de al menos o de aproximadamente 100, de al menos o de aproximadamente 150, de al menos o de aproximadamente 200, de al menos o de aproximadamente 250, de al menos o de aproximadamente 300, de al menos o de aproximadamente 350 o de al menos o de aproximadamente 400 pares de bases de una secuencia génica que no se encuentra en otras secuencias génicas. El término “aproximadamente”, como se usa en la oración anterior, se refiere a un aumento o disminución de 10 % del valor numérico indicado. Dichos polinucleótidos también pueden denominarse sondas polinucleotídicas que son capaces de hibridar con secuencias de los genes, o con partes exclusivas de las mismas, descritas en el presente documento. Preferentemente, las secuencias son las secuencias de ARNm codificadas por los genes, el ADNc correspondiente a dichos ARNm y/o versiones amplificadas de dichas secuencias. En realizaciones preferidas de la divulgación, las sondas polinucleotídicas se inmovilizan en una matriz, en otros dispositivos, o en manchas individuales que localizan las sondas.

65 En otra realización de la divulgación, toda o parte de una secuencia desvelada puede amplificarse y detectarse por métodos tales como la reacción en cadena de la polimerasa (PCR) y sus variaciones, tales como, pero sin limitación, PCR cuantitativa (PCRc), PCR con transcripción inversa (RT-PCR) y PCR en tiempo real, opcionalmente RT-PCR

en tiempo real. Dichos métodos utilizarán uno o dos cebadores que son complementarios a partes de una secuencia desvelada, utilizándose los cebadores para cebar la síntesis de ácidos nucleicos. Los ácidos nucleicos recién sintetizados se marcan opcionalmente y pueden detectarse directamente o por hibridación con un polinucleótido de la divulgación. Los ácidos nucleicos recién sintetizados pueden ponerse en contacto con polinucleótidos (que contienen secuencias) de la divulgación, en condiciones que permitan su hibridación. Como alternativa, y en otra realización de la divulgación, la expresión génica puede determinarse mediante análisis de proteínas expresadas en una muestra de células de interés utilizando uno o más anticuerpos específicos para uno o más epítopos de productos génicos (proteínas) individuales en dicha muestra de células. Preferentemente, dichos anticuerpos se marcan para facilitar su detección después de unirse al producto génico.

El término "marcador" se refiere a una composición capaz de producir una señal detectable, indicativa de la presencia de la molécula marcada. Como marcadores adecuados se incluyen radioisótopos, cromóforos nucleotídicos, enzimas, sustratos, moléculas fluorescentes, restos quimioluminiscentes, partículas magnéticas, restos bioluminiscentes y similares. Como tal, un marcador es cualquier composición detectable mediante espectroscopia, fotoquímica, bioquímica, inmunoquímica, medios eléctricos, ópticos o químicos.

El término "soporte" se refiere a soportes convencionales, tales como perlas, partículas, tiras reactivas, fibras, filtros, membranas y soportes de silano o silicato tales como portaobjetos.

Como se usa en el presente documento, una "muestra de tejido de cáncer" o "muestra de célula de cáncer" se refiere a una muestra aislada de tejido que contiene células de un individuo que padece el cáncer correspondiente. La muestra puede ser de un material extraído mediante un procedimiento quirúrgico, tal como una biopsia. Dichas muestras son aislados primarios (a diferencia de las células cultivadas) y pueden extraerse mediante cualquier medio adecuado reconocido en la técnica. En algunas regiones la "muestra" puede extraerse mediante un método no invasivo, incluyendo, pero sin limitación, abrasión, aspiración con aguja fina.

Una "muestra de tejido mamario" o "muestra de células de mama" se refiere a una muestra de tejido o líquido mamario aislada de un individuo que se sospecha que padece cáncer de mama, o que está en riesgo de desarrollarlo. Dichas muestras son aislados primarios (a diferencia de las células cultivadas) y pueden extraerse mediante cualquier medio no invasivo, incluyendo, pero sin limitación, lavado ductal, aspiración con aguja fina, biopsia con aguja, con los dispositivos y métodos descritos en la patente de Estados Unidos nº. 6.328.709, o con cualquier medio adecuado reconocido en la técnica. Como alternativa, la "muestra" puede extraerse mediante un método invasivo, incluyendo, pero sin limitación, biopsia quirúrgica.

La "expresión" y "expresión génica" incluyen la transcripción y/o traducción de material de ácido nucleico. Por supuesto los términos también pueden limitarse, si se indicase, en referencia solo a la transcripción de ácidos nucleicos.

Como se usa en el presente documento, el término "comprendiendo", y sus homólogos, se utilizan en su sentido inclusivo; es decir, equivalente al término "incluyendo" y sus homólogos correspondientes.

Las condiciones que "permiten" que se produzca un suceso, o condiciones que son "adecuadas" para que se produzca un suceso, tal como hibridación, extensión de cadena, y similares, o condiciones "adecuadas", son condiciones que no impiden que dichos sucesos se produzcan. Por tanto, estas condiciones permiten, potencian, facilitan y/o son conductoras del suceso. Dichas condiciones, conocidas en la técnica y descritas en el presente documento, dependen, por ejemplo, de la naturaleza de la secuencia de nucleótidos, de las condiciones de temperatura y del tampón. Estas condiciones también dependen de si se desea que se produzca el suceso, tal como hibridación, escisión, extensión de cadena o transcripción.

Una "mutación" de secuencia, como se usa en el presente documento, se refiere a cualquier alteración de secuencia en la secuencia de un gen de interés desvelado en el presente documento, en comparación con una secuencia referencia. Una mutación de secuencia incluye cambios de un solo nucleótido, o alteraciones de más de un nucleótido en una secuencia, debido a mecanismos tales como sustitución, delección o inserción. El polimorfismo de un solo nucleótido (SNP, *Single Nucleotide Polymorphism*) es también una mutación de secuencia, como se usa en el presente documento. Dado que la presente divulgación se basa en el nivel relativo de expresión génica, en la realización práctica de la divulgación también pueden analizarse mutaciones en regiones no codificantes de genes, como se desvela en el presente documento.

"Detección" incluye cualquier medio de detección, incluyendo la detección directa e indirecta de la expresión génica y cambios en su interior. Por ejemplo, pueden observarse directa o indirectamente productos "menos detectables", y el término indica cualquier reducción (incluyendo la ausencia de señal detectable). De manera similar, un producto "más detectable" significa cualquier aumento, observado tanto directa como indirectamente.

Los aumentos y disminuciones en la expresión de las secuencias desveladas se definen en los siguientes términos basándose en el porcentaje o factores de cambio sobre la expresión en células normales. Los aumentos pueden ser de 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 120, 140, 160, 180 o 200 % con respecto a los niveles de expresión en

células normales. Como alternativa, los factores de aumento pueden ser de 1, 1,5, 2, 2,5, 3, 3,5, 4, 4,5, 5, 5,5, 6, 6,5, 7, 7,5, 8, 8,5, 9, 9,5 o 10 veces los niveles de sobreexpresión en células normales. Las disminuciones pueden ser de 10, 20, 30, 40, 50, 55, 60, 65, 70, 75, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98, 99 o 10 0% con respecto a los niveles de expresión en células normales.

5 A menos que se defina de otra manera, todos los términos técnicos y científicos utilizados en el presente documento tienen el mismo significado que el que normalmente entiende un experto habitual en la técnica a la que pertenece esta divulgación.

10 Generalidades

Los patrones de expresión génica desvelados en el presente documento son factores predictivos para determinar el beneficio terapéutico en un cambio en una terapia endocrina. En algunos casos, como un ejemplo no limitante, la predicción es en pacientes con cáncer de mama sin afectación ganglionar, tales como pacientes RE+ sin afectación ganglionar.

15 Para determinar los niveles de expresión de genes en la realización práctica de la presente divulgación, puede utilizarse cualquier método conocido en la técnica. En algunas realizaciones, se usa la expresión basada en la detección de ARN que se hibrida con los genes identificados y desvelados en el presente documento. Esto se realiza fácilmente mediante cualquier método de detección o de amplificación+detección de ARN conocido o reconocido como equivalente en la técnica, tal como, pero sin limitación, PCR con transcripción inversa, los métodos desvelados en la solicitud de patente de Estados Unidos nº. de serie 10/062,857 (presentada el 25 de octubre de 2001) así como en la solicitud de patente provisional de Estados Unidos 60/298.847 (presentada el 15 de junio de 2001) y 60/257.801 (presentada el 2 de diciembre del 2000), y métodos para detectar la presencia, o ausencia, de secuencias estabilizadoras o desestabilizadoras de ARN.

30 Como alternativa, puede utilizarse la expresión basada en la detección del estado del ADN. La detección del ADN de un gen identificado como metilado o deleciónado, puede utilizarse para genes que tienen expresión disminuida. Esto puede realizarse fácilmente mediante métodos basados en PCR conocidos en la técnica, incluyendo, pero sin limitación, PCRc. A la inversa, la detección del ADN de un gen identificado como amplificado puede utilizarse para genes que tienen expresión aumentada en correlación con un resultado de cáncer de mama particular. Esto puede realizarse fácilmente mediante métodos basados en PCR, hibridación *in situ* con fluorescencia (FISH, acrónimo de *fluorescence in situ hybridization*) e hibridación cromogénica *in situ* (CISH, acrónimo de *chromosome in situ hybridization*), conocidos en la técnica.

35 También puede utilizarse una expresión basada en la detección de una presencia, un aumento o una disminución en los niveles o actividad de proteínas. La detección puede realizarse mediante cualquier método basado en inmunohistoquímica (IHQ), en sangre (especialmente para proteínas secretadas), en anticuerpos (incluyendo autoanticuerpos contra la proteína), en células exfoliadas (del cáncer), en espectroscopia de masas, y en la obtención de imágenes (incluyendo el uso de ligandos marcados) conocido en la técnica y reconocidos como apropiado para la detección de la proteína. Los métodos basados en anticuerpos y en la obtención de imágenes son adicionalmente útiles para la localización de tumores después de la determinación del cáncer utilizando células obtenidas mediante un procedimiento no invasivo (tal como lavado ductal o aspiración con aguja fina), en el que se desconoce la fuente de las células cancerosas. Para localizar uno o más carcinomas en un paciente, puede utilizarse un anticuerpo o ligando marcado.

50 Una realización que para determinar la expresión utiliza un análisis basado en ácidos nucleicos, es mediante la inmovilización en un soporte sólido, de una o más secuencias de los genes identificados en el presente documento, que incluye, pero sin limitación, un sustrato sólido como una matriz o perlas o tecnología basada en perlas como se conoce la técnica. Como alternativa, también pueden utilizarse análisis de expresión basados en soluciones conocidos en la técnica.

55 El gen (o genes) inmovilizado puede estar en forma de polinucleótidos que son exclusivos o, de otra manera, específicos para el gen (o genes), de tal forma que el polinucleótido tendría la capacidad de hibridar con un ADN o ARN correspondiente al gen (o genes). Estos polinucleótidos pueden ser la longitud completa del gen (o genes) o ser secuencias cortas de los genes (hasta un nucleótido más cortas que la secuencia de longitud completa conocida en la técnica, por deleción a partir del extremo 5' o 3' de la secuencia) que opcionalmente están mínimamente interrumpidas (tal como mediante pares de bases no complementarias insertadas o mal emparejadas) de tal manera que la hibridación con un ADN o ARN correspondiente al gen (o genes) no esté afectada. En algunos casos, los polinucleótidos utilizados son del extremo 3' del gen, tal como dentro de aproximadamente 350, aproximadamente 300, aproximadamente 250, aproximadamente 150, aproximadamente 100 aproximadamente o 50 nucleótidos dese la señal de poliadelinación o sitio de poliadelinación de un gen o secuencia expresada. Los polinucleótidos que contienen mutaciones con respecto a las secuencias de los genes divulgados también pueden utilizarse siempre y cuando la presencia de las mutaciones todavía permita la hibridación para producir una señal detectable.

65 El gen (o genes) inmovilizado puede utilizarse para determinar el estado de las muestras de ácido nucleico

preparadas a partir del cáncer de muestra, o mama, o célula (o células) para los que el resultado del sujeto de la muestra (por ejemplo, el paciente del que se obtuvo la muestra) no se conoce o para la confirmación de un resultado que está asignado al sujeto de la muestra. Sin limitar la divulgación, tal célula puede ser de un paciente con cáncer de mama RE+. El polinucleótido (o polinucleótidos) inmovilizado necesita solo ser suficiente para hibridar de forma
5 específica con las moléculas de ácido nucleico correspondientes obtenidas de la muestra en condiciones adecuadas.

Como apreciarán los expertos en la materia, algunas de las correspondientes secuencias indicadas anteriormente incluyen tramos de 3' poli A (o poli T en la cadena complementaria) que no contribuyen a la singularidad de las
10 secuencias divulgadas. La divulgación puede, por lo tanto, llevarse a la práctica con secuencias que carecen de los tramos 3' poli A (o poli T). La singularidad de las secuencias divulgadas se refiere a las partes o totalidades de las secuencias que se encuentran solo en los ácidos nucleicos del gen divulgado, que incluyen secuencias exclusivas encontradas en la porción 3' no traducida de los genes. Las secuencias exclusivas preferentes para la práctica de la divulgación son las que contribuyen a las secuencias consenso para cada uno de los tres conjuntos, de forma que
15 las secuencias exclusivas serán útiles en la detección de la expresión en una diversidad de individuos en lugar de ser específicas para un polimorfismo presente en algunos individuos. Como alternativa, pueden utilizarse secuencias exclusivas para un individuo o una subpoblación. Las secuencias exclusivas preferidas son, preferentemente, de las longitudes de los polinucleótidos de la divulgación, como se comenta en el presente documento.

Para determinar los niveles de expresión (aumentados o disminuidos) de las secuencias descritas anteriormente en la práctica de la divulgación, puede utilizarse cualquier método conocido en la técnica. En una realización de la divulgación, se utiliza la expresión basada en la detección de ARN que hibrida con los polinucleótidos que contienen las secuencias descritas anteriormente. Esto se realiza fácilmente mediante cualquier método de detección o de
20 amplificación+detección de ARN, conocido o reconocido como equivalente en la técnica, tal como, pero sin limitación, PCR con transcripción inversa (opcionalmente, PCR en tiempo real), los métodos divulgados en la solicitud de patente en Estados Unidos N.º de serie 10/062.587 titulada "Amplificación de Ácidos Nucleicos" presentada el 25 de octubre de 2001, así como la solicitud de patente provisional de Estados Unidos 60/198.847 (presentada el 15 de junio de 2001) y 60/257.801 (presentada el 22 de diciembre de 2000), los métodos divulgados en la patente de Estados Unidos N.º 6.291.170 y PCR cuantitativa. También pueden utilizarse métodos para
25 identificar la estabilidad aumentada del ARN (que da como resultado una observación de una expresión aumentada) o la estabilidad disminuida del ARN (que da como resultado una observación de una expresión disminuida). Estos métodos incluyen la detección de secuencias que aumentan o disminuyen la estabilidad de los ARNm que contienen las secuencias de los genes. Estos métodos también incluyen la detección de la degradación aumentada de ARNm.

En algunas realizaciones de la divulgación, se utilizan polinucleótidos que tienen secuencias presentes en las regiones 3' no traducidas y/o no codificantes de las anteriores secuencias divulgadas para detectar los niveles de expresión de las secuencias génicas en células cancerosas o de mama. Tales polinucleótidos pueden contener, opcionalmente, secuencias encontradas en las partes 3' de las regiones codificantes de las anteriores secuencias divulgadas. Los polinucleótidos que contienen una combinación de secuencias de las regiones codificantes y 3' no
30 codificantes tienen preferentemente las secuencias dispuestas de forma contigua, pero sin secuencias heterólogas intervinientes.

Como alternativa, la divulgación puede llevarse a la práctica con polinucleótidos que tengan secuencias presentes en las regiones 5' no traducida y/o no codificante de las secuencias génicas en células cancerosas o de mama, para
35 detectar su nivel de expresión. Tales polinucleótidos pueden contener, opcionalmente, secuencias encontradas en las partes 5' de las regiones codificantes. Los polinucleótidos que contienen una combinación de secuencias de las regiones codificantes y 5' no codificantes tienen, preferentemente, las secuencias dispuestas de forma contigua, pero sin secuencias heterólogas intervinientes. La divulgación también puede llevarse a la práctica con secuencias presentes en las regiones codificantes de las secuencias génicas divulgadas.

Los polinucleótidos no limitativos contienen secuencias de las regiones 3' o 5' no traducidas y/o no codificantes de al menos aproximadamente 20, al menos aproximadamente 24, al menos aproximadamente 26, al menos aproximadamente 28, al menos aproximadamente 30, al menos aproximadamente 32, al menos aproximadamente 34, al menos aproximadamente 36, al menos aproximadamente 38, al menos aproximadamente 40, al menos aproximadamente 42, al menos aproximadamente 44 o al menos aproximadamente 46 nucleótidos consecutivos. El término "aproximadamente", utilizado en la oración anterior se refiere a un aumento o disminución de 1 a partir del valor numérico establecido. Son incluso más preferentes los polinucleótidos que contienen secuencias de al menos o aproximadamente 50, al menos o aproximadamente 100, al menos o aproximadamente 150, al menos o aproximadamente 200, al menos o aproximadamente 250, al menos o aproximadamente 300, al menos o aproximadamente 350 o al menos o aproximadamente 400 nucleótidos consecutivos. El término "aproximadamente",
40 utilizado en la oración anterior se refiere a un aumento o disminución del 10% a partir del valor numérico establecido.

Las secuencias del extremo 3' o 5' de las anteriores regiones codificantes descritas como se encuentran en los polinucleótidos de la divulgación, tienen las mismas longitudes que las descritas anteriormente, excepto que estarían naturalmente limitadas por la longitud de la región codificante. El extremo 3' de una región codificante puede incluir secuencias hasta la mitad 3' de la región codificante. Por el contrario, el extremo 5' de una región codificante puede
65

incluir secuencias hasta la mitad 5' de la región codificante. Por supuesto, las anteriores secuencias descritas, o las regiones codificantes y los polinucleótidos que contienen partes de las mismas, pueden utilizarse en su totalidad.

5 Los polinucleótidos que combinan las secuencias de una región 3' no traducida y/o no codificante y el extremo 3' asociado de la región codificante, pueden ser al menos o aproximadamente de 100, al menos o aproximadamente de 150, al menos o aproximadamente de 200, al menos o aproximadamente de 250, al menos o aproximadamente de 300, al menos o aproximadamente de 350 o al menos o aproximadamente de 400 nucleótidos consecutivos. Preferentemente, los polinucleótidos utilizados son del extremo 3' del gen, tal como dentro de aproximadamente 350, aproximadamente 300, aproximadamente 250, aproximadamente 200, aproximadamente 150, 10 aproximadamente 100 o aproximadamente 50 nucleótidos desde la señal de poliadelinación o sitio de poliadelinación de un gen o secuencia expresada. Los polinucleótidos que contienen mutaciones con respecto a las secuencias de los genes divulgados también pueden utilizarse siempre y cuando la presencia de las mutaciones todavía permita la hibridación para producir una señal detectable.

15 En otra realización de la divulgación, pueden utilizarse polinucleótidos que contienen deleciones de nucleótidos del extremo 5' y/o 3' de las secuencias anteriores divulgadas. Las deleciones son preferentemente de 1-5, 5-10, 10-15, 15-20, 20-25, 25-30, 30-35, 35-40, 40-45, 45-50, 50-60, 60-70, 70-80, 80-90, 90-100, 100-125, 125-150, 150-175 o 175-200 nucleótidos del extremo 3' y/o 5', aunque el grado de la deleción estaría limitado naturalmente por la longitud de las secuencias divulgadas y la necesidad de ser capaz de utilizar los polinucleótidos para la detección de los niveles de expresión.

20 Otros polinucleótidos de la divulgación del extremo 3' de las anteriores secuencias divulgadas incluyen los cebadores y las sondas opcionales para la PCR cuantitativa. En algunas realizaciones, los cebadores y sondas son los que amplifican una región de menos de aproximadamente 350, menos de aproximadamente 300, menos de aproximadamente 250, menos de aproximadamente 200, menos de aproximadamente 150, menos de aproximadamente 100 o menos de aproximadamente 50 nucleótidos de la señal de poliadelinación o sitio de poliadelinación de un gen o secuencia expresada.

25 En otras realizaciones más de la divulgación, pueden utilizarse polinucleótidos que contienen partes de las anteriores secuencias divulgadas que incluyen el extremo 3'. Tales polinucleótidos contendrían al menos o aproximadamente 50, al menos o aproximadamente 100, al menos o aproximadamente 150, al menos o aproximadamente 200, al menos o aproximadamente 250, al menos o aproximadamente 300, al menos o aproximadamente 350 o al menos o aproximadamente 400 nucleótidos consecutivos del extremo 3' de las secuencias divulgadas.

30 La divulgación también incluye polinucleótidos utilizados para detectar la expresión génica en células de mama. Los polinucleótidos pueden comprender un polinucleótido más corto que consista en secuencias encontradas en los genes anteriores en combinación con secuencias heterólogas no encontradas de forma natural en combinación con las secuencias. Los ejemplos no limitativos incluyen secuencias cortas de vectores de clonación o presentes en fragmentos de restricción utilizados para preparar sondas marcadas o cebadores como se describe en el presente documento.

HoxB13 y H/I

35 Los métodos de la divulgación basados en los niveles de expresión de HoxB 13 en células de cáncer de mama de un sujeto, pueden utilizarse como un índice de predicción del beneficio en el cambio de una terapia endocrina después de un ciclo inicial de terapia endocrina. En algunas realizaciones, puede utilizarse una proporción de dos genes de expresión de HoxB13 con respecto a la expresión de IL17BR (o proporción de HoxB13:IL17BR) en la manera comunicada por Ma et al. (J. Clin. Oncol., 24:4611-9 (2006)). En realizaciones alternativas, puede utilizarse una proporción de dos genes de expresión de HoxB13 con respecto a la expresión de CHDH.

40 En los casos en los que se utilice solo la expresión de HoxB13 o la proporción de HoxB13:IL17BR (H:I), puede utilizarse un valor límite para definir las células de cáncer de mama que tienen un valor "alto" y uno "bajo" correspondiente a la expresión. En algunas realizaciones, puede utilizarse un límite para definir células de cáncer de mama que tienen un valor "H/I alto" y uno "H/I bajo". Como un ejemplo no limitativo, puede utilizarse el valor de 0,06 de la manera comunicada por Ma et al. En otras realizaciones, el límite puede ser la expresión promedio de HoxB 13 en células de cáncer de mama de sujetos que lo padecen. En posibles realizaciones adicionales, el límite puede ser el valor promedio de H/I en células de cáncer de mama de sujetos que lo padecen, determinado por la expresión de Hob13 promedio/la expresión de IL17BR promedio.

IGM

Los genes divulgados a continuación tienen las siguientes funciones y máximos de expresión indicados en el ciclo celular:

65

Gen	Máximo expresión	de	Funciones en el ciclo celular
BUB1B	G2/M		punto de control de ensamblaje del uso mitótico
CENPA	G2/M		ensamblaje del centrómero
NEK2	G2/M		duplicación del centrómero
RACGAP1	No determinado		inicio de la citocinesis
RRM2	S		replicación de ADN

5 Las secuencias de estos genes se han comunicado y caracterizado anteriormente en la técnica. Por ejemplo, el 6 de septiembre de 2007, el gen BUB1B humano (también conocido como quinasa 6 activada por la proteína p21 o PAK 6) se identificó mediante Unigene Hs.631699 y se caracterizó mediante 273 secuencias correspondientes. El 6 de marzo de 2010, la misma información del gen se identificó mediante UniGene Hs.513645 y se caracterizó como que correspondía al cromosoma 15 en la posición 15q14 y como se confirma mediante 23 secuencias de ARNm y 549 secuencias EST (acrónimo de *expressed sequence tag*, etiqueta de secuencia expresada).

10 También el 6 de septiembre de 2007, el gen CENPA humano se identificó mediante Hs.1594 (con 129 secuencias correspondientes). El 6 de marzo de 2010, se caracterizó la misma información del gen como que correspondía al cromosoma 2 en la posición 2p24-p21 y como se confirma mediante 10 secuencias de ARNm y 119 secuencias EST.

15 También el 6 de septiembre de 2007, el gen NEK2 humano se identificó mediante Hs.153704 (con 221 secuencias correspondientes). El 6 de marzo de 2010, se caracterizó la misma información del gen como que correspondía al cromosoma 1 en la posición 1q32.2-q41 y como se confirma mediante 17 secuencias ARNm y 205 secuencias EST.

20 También el 6 de septiembre de 2007, el gen RACGAP1 humano se identificó mediante Hs.696319 (con 349 secuencias correspondientes). El 6 de marzo de 2010, se identificó la misma información del gen mediante UniGene Hs.505469 y se caracterizó como que correspondía al cromosoma 12 en la posición 12q13.12 y como se confirma mediante 15 secuencias de ARNm y 398 secuencias EST.

25 También el 6 de septiembre de 2007, el gen RRM2 humano se identificó mediante Hs.226390 con 1348 secuencias correspondientes). El 6 de marzo de 2010, se caracterizó la misma información del gen como que correspondía al cromosoma 2 en la posición 2p25-p24 y como se confirma mediante 25 secuencias de ARNm y 1328 secuencias EST.

30 Las secuencias de ARNm y EST que se corresponden con cada uno de los anteriores identificadores Unigene se incorporan en el presente documento como referencia en su totalidad como si se expusieran totalmente y el experto en la materia puede utilizarlas en la realización práctica de la divulgación como considere apropiado. Las secuencias de ARNm representativas de cada uno de los genes BUB1B, CENPA, NEK2, RACGAP1 y RRM2, se han divulgado en la solicitud de patente de Estados Unidos 12/718.973, publicada como documento US 2011-0136680 A1 el 9 de junio de 2001. Las secuencias divulgadas no son limitativas para la realización práctica de la invención divulgada y se proporcionan como una prueba del conocimiento sustancial en el campo con respecto a las secuencias que son los genes divulgados. Adicionalmente, el experto tiene toda la capacidad de alinear cualquiera de dos o más de las secuencias expresadas conocidas para cada uno de estos genes, para identificar una zona de identidad o de cambios conservados como una región que identifica de forma exclusiva cada uno de estos genes en comparación con otros genes. Además, el experto tiene toda la capacidad de alinear cualquiera de dos o más de las secuencias expresadas conocidas para cada uno de estos genes para identificar una zona exclusiva para una o más de las secuencias expresadas como una región que identifica de forma exclusiva una secuencia expresada conocida con respecto a al menos otra secuencia expresada. Como ejemplo no limitativo, una región exclusiva puede ser una variante de la secuencia expresada para uno de los genes conocidos, de forma que la región puede utilizarse para identificar la expresión de la variante.

45 Las secuencias de los mismos genes también se han identificado y caracterizado en otras especies animales. Por lo tanto, el experto en la materia sabe perfectamente cómo identificar los genes divulgados con respecto a otros genes animales. Opcionalmente, el experto también puede comparar las secuencias conocidas de los genes divulgados de diferentes fuentes animales para identificar regiones y secuencias conservadas exclusivas para estos genes con respecto a otros genes.

Métodos

55 Como se describe en el presente documento, la divulgación incluye la identidad de genes, cuya expresión puede utilizarse para proporcionar información pronóstica relacionada con el cáncer. En particular, los niveles de expresión de estos genes pueden utilizarse en relación al cáncer de mama. En algunos métodos, el perfil de expresión génica se correlaciona con (y, de este modo, tiene la capacidad de discriminar entre) pacientes que se espera que se beneficien de un cambio en la terapia endocrina después de un tratamiento inicial con terapia endocrina durante un periodo de tiempo. En otras realizaciones, la divulgación incluye un método para comparar la expresión génica en

una muestra de células cancerosas de un paciente con respecto al perfil de expresión génica para determinar la probabilidad clínica o el desenlace del tratamiento para el paciente, o el resultado biológico natural, en ausencia de un cambio.

5 Estas realizaciones de la divulgación pueden utilizarse provechosamente para cumplir una necesidad diagnóstica no
satisfecha importante para poder predecir si un paciente probablemente se beneficiará de un cambio en el tipo de
tratamiento. Por ejemplo, un valor alto de la proporción H:I está muy asociado a una respuesta a un cambio de la
10 terapia de primera línea con tamoxifeno durante hasta 5 años a una terapia con letrozol. El cambio puede producirse
en cualquier momento después de la terapia de primera línea, tal como inmediatamente después, al cabo de tres
meses después de la finalización de la terapia de primera línea, al cabo de seis meses después de la finalización de
la terapia de primera línea, al cabo de nueve meses después de la finalización de la terapia de primera línea, al cabo
de 12 meses después de la finalización de la terapia de primera línea, al cabo de 18 meses después de la
finalización de la terapia de primera línea o al cabo de 24 meses (o más) después de la finalización de la terapia de
primera línea.

15 De esta forma, la divulgación incluye un método para identificar a un paciente, de una población de pacientes con
células de cáncer de mama ER+, tratado con una primera terapia endocrina y sin cáncer durante un periodo de
tiempo, como que pertenece a una subpoblación de pacientes con un mejor pronóstico si se trata con una terapia
endocrina alternativa. En algunos casos, el cáncer de mama en el sujeto es cáncer de mama sin afectación
20 ganglionar. La divulgación proporciona un medio no subjetivo para la identificación de pacientes en la subpoblación.

La divulgación también incluye un método de determinación de un pronóstico y/o del resultado de supervivencia
evaluando la expresión de los patrones divulgados en el presente documento. De esta forma, cuando anteriormente
puede haberse utilizado una interpretación subjetiva para determinar el pronóstico y/o el tratamiento de los pacientes
25 de cáncer, esta divulgación proporciona patrones de expresión génica objetivos, que pueden utilizarse solos o en
combinación con criterios subjetivos para proporcionar una evaluación más precisa de los resultados de los
pacientes, incluyendo la supervivencia y la recidiva del cáncer.

30 En algunas realizaciones, la divulgación proporciona un método para determinar el tratamiento terapéutico de un
paciente con cáncer determinando el pronóstico para dicho paciente, evaluando los niveles de expresión de una
muestra de células cancerosas de dicho paciente, descritos en el presente documento, y seleccionando un
tratamiento para un paciente con tal expresión génica. La evaluación puede incluir medir o detectar o determinar el
nivel de expresión de los genes de cualquier modo adecuado descrito en el presente documento o conocido por el
35 experto. En muchos casos, el cáncer es cáncer de mama, y el sujeto es un paciente humano. De forma adicional, las
células cancerosas pueden ser las de un tumor y/o proceder de un sujeto sin afectación ganglionar (cáncer sin
afectación de ganglios linfáticos) o con afectación ganglionar (cáncer con afectación de ganglios linfáticos).

El nivel de expresión requerido puede ser el que se identifica mediante los métodos descritos en el presente
documento para los genes utilizados. De forma adicional, la evaluación puede incluir preparar ARN de la muestra,
40 opcionalmente para su uso en una PCR (reacción en cadena de la polimerasa) u otro método analítico, como se
describe en el presente documento. El método de la PCR es opcionalmente RT-PCR (PCR con transcripción
inversa) o PCR cuantitativa, tal como RT-PCR en tiempo real. Como alternativa, la evaluación puede realizarse
utilizando una matriz, tal como una micromatriz como se conoce en el campo pertinente. Opcionalmente, la muestra
de células cancerosas se disecciona de tejido extirpado u obtenido de dicho sujeto. Como se describe en el presente
45 documento, puede utilizarse una diversidad de tipos de muestra, incluyendo, como un ejemplo no limitativo, una
muestra fijada en formalina e incluida en parafina (FFPE). Como se describe en el presente documento, el método
puede incluir evaluar o determinar la proporción H:I (proporción de los niveles de expresión de HoxB13 e IL17BR) en
la muestra, como se divulga en el presente documento.

50 Como ejemplo no limitativo, los cinco genes del IGM pueden evaluarse y utilizarse para detectar niveles de
expresión que corresponden a un valor que es de "riesgo alto" (que está por encima del límite) para el IGM, o para
detectar niveles de expresión que corresponden a un valor que es de "riesgo bajo" (que está en límite, o por debajo
del límite) para el IGM, como se divulga en el presente documento. En algunos casos, el umbral del límite del índice
de grado molecular, IGM, puede ser 0 (cero), tal como cuando las mediciones de los niveles de expresión se
55 normalizan a 0 (cero) con una desviación típica de 1. En realizaciones alternativas, el límite puede ser o ser
aproximadamente 0,05, ser o ser aproximadamente 0,10, ser o ser aproximadamente 0,15, ser o ser
aproximadamente 0,20, ser o ser aproximadamente 0,25, ser o ser aproximadamente -0,05, ser o ser
aproximadamente -0,10, ser o ser aproximadamente -0,15, ser o ser aproximadamente -0,20, ser o ser
aproximadamente -0,25, ser o ser aproximadamente -0,30, ser o ser aproximadamente -0,35, ser o ser
60 aproximadamente -0,40, ser o ser aproximadamente -0,45, ser o ser aproximadamente -0,50, ser o ser
aproximadamente -0,55, ser o ser aproximadamente -0,60, ser o ser aproximadamente -0,65, ser o ser
aproximadamente -0,70, ser o ser aproximadamente -0,75, ser o ser aproximadamente - 0,80, ser o ser
aproximadamente -0,85, ser o ser aproximadamente -0,90, ser o ser aproximadamente -0,95, ser o ser
aproximadamente -1,0, ser o ser aproximadamente -1,1, ser o ser aproximadamente -1,2, ser o ser
65 aproximadamente -1,3, ser o ser aproximadamente -1,4, ser o ser aproximadamente -1,5, ser o ser
aproximadamente -1,6, ser o ser aproximadamente -1,7, ser o ser aproximadamente -1,8, ser o ser

aproximadamente -1,9, ser o ser aproximadamente -2,0 o menor. Con respecto a la proporción H:I, su determinación puede realizarse como se describe en Ma *et al.*, Cancer Cell, 5: 607-16 (2004) y en Ma *et al.* (2006) como se hace referencia en el presente documento. Por ejemplo, puede utilizarse un valor de 0,06 para determinar si una muestra tiene una proporción H:I de "riesgo alto" ($> 0,06$) o de "riesgo bajo" ($\leq 0,06$).

5 De esta forma, utilizando un umbral, o límite, de 0 (cero) como un ejemplo no limitativo para el IGM, con los cinco genes, los métodos divulgados proporcionan dos posibles resultados de ensayo para una muestra dada: un "índice de grado molecular, IGM, de riesgo alto" corresponde a un valor por encima de 0 (cero) y un "IGM de riesgo bajo" corresponde a un valor ≤ 0 . Un "IGM de riesgo alto" es indicativo de un cáncer de "riesgo alto", que incluye cáncer de mama que es análogo al de un tumor de Grado III, como se define en los métodos y patrones conocidos en el campo. Un "IGM de riesgo bajo" es indicativo de un cáncer de "riesgo bajo", que incluye cáncer de mama, que es análogo al de un tumor de Grado I, como se define mediante en métodos y patrones conocidos en el campo.

15 En una realización de la divulgación, se proporciona un método para determinar el riesgo o la probabilidad de recidiva de cáncer de un sujeto después de un tratamiento para el cáncer de mama, tal como la extirpación del cáncer mediante cirugía. El método puede comprender i) preparar ADNc a partir de ácidos nucleicos en una muestra de células de cáncer de mama RE+ (positivo a receptores de estrógeno) extirpada de un sujeto que padece cáncer de mama; ii) determinar los niveles de expresión de los siete genes del Índice del Cáncer de Mama (ICM) divulgado a partir de dicho ADNc para determinar un valor de ICM; iii) identificar que el sujeto ha sido tratado con terapia endocrina durante un periodo de tiempo sin recidiva de cáncer y iv) clasificar el cáncer como que probablemente recurra debido a un valor de ICM de riesgo alto. En algunos casos, el sujeto se ha tratado con terapia endocrina durante aproximadamente 5 años o más, durante aproximadamente 4 años o más, durante aproximadamente 3 años o más, durante aproximadamente 2 años o más o durante aproximadamente 1 año o más.

25 En otra realización de la divulgación, se proporciona un método para determinar la probabilidad de un cambio beneficioso en la terapia endocrina como tratamiento para el cáncer de mama. El método puede comprender i) preparar ADNc a partir de ácidos nucleicos en una muestra de células de cáncer de mama RE+, extirpada de un sujeto que padece de cáncer de mama; ii) determinar el nivel de expresión del gen Hob13 a partir de dicho ADNc; iii) identificar opcionalmente al sujeto como que se ha sometido a extirpación quirúrgica del cáncer de mama; iv) identificar al sujeto como que se ha tratado con una primera terapia endocrina durante un periodo de tiempo sin recidiva del cáncer y v) clasificar al sujeto, como se esperaba, para que se beneficie del tratamiento sin una segunda terapia endocrina distinta después de suspender la primera terapia endocrina, basándose dicha clasificación en un nivel de expresión elevado de HoxB13. En algunos casos, el nivel de expresión elevado de HoxB13 se determina como parte del valor H/I y la clasificación está basada en un valor de H/I alto. En algunos casos, el sujeto se ha tratado con terapia endocrina durante aproximadamente 5 años o más, durante aproximadamente 4 años o más, durante aproximadamente 3 años o más, durante aproximadamente 2 años o más, o durante aproximadamente 1 año o más.

40 En realizaciones adicionales, la divulgación proporciona un método para tratar a un paciente que se ha sometido a una primera terapia endocrina como se describe anteriormente. El método puede comprender la anterior determinación de la probabilidad de un cambio beneficioso en la terapia endocrina como tratamiento para el cáncer de mama seguido del tratamiento del paciente con una segunda terapia endocrina después de finalizar el tratamiento con la primera terapia endocrina.

45 Como ejemplos no limitativos, la primera terapia endocrina puede ser el tratamiento con un SERM (modulador selectivo de receptores de estrógeno) o un SERD (regulador negativo selectivo de receptores de estrógeno) y la segunda terapia endocrina puede ser un tratamiento con un inhibidor de aromatasa. Como alternativa, la primera terapia endocrina puede ser el tratamiento con un inhibidor de aromatasa y la segunda terapia endocrina puede ser el tratamiento con un SERM o un SERD. Las realizaciones incluyen tamoxifeno como la primera terapia endocrina seguido de letrozol como la segunda terapia endocrina, o letrozol como la primera terapia endocrina seguido de tamoxifeno como la segunda terapia endocrina.

55 La divulgación incluye adicionalmente un método de determinación de un factor o indicador de pronóstico de la respuesta clínica en mujeres premenopáusicas y posmenopáusicas. Las mujeres posmenopáusicas pueden definirse como las que tienen ≥ 50 años, mientras que las mujeres premenopáusicas pueden definirse como las que tienen menos de 50 años.

60 La capacidad de discriminar la confiere la identificación de la expresión de genes individuales como importante y no por la forma del ensayo utilizado para determinar el nivel real de expresión. Puede utilizarse un ensayo que identifique cualquier característica de un gen individual identificado como se divulga en el presente documento siempre que el ensayo refleje, de forma cuantitativa o cualitativa, la expresión del gen en el "transcriptoma" (la fracción transcrita de genes en un genoma) o el "proteoma" (la fracción traducida de genes expresados en un genoma). Las características identificativas incluyen, pero sin limitación, secuencias de ácido nucleico exclusivas utilizadas que codifican (ADN), o que expresan (ARN), dicho gen o epítopos específicos para, o actividades de, una proteína que codifica dicho gen. Todo que se necesita es la identidad del gen(es) necesaria para discriminar entre los resultados de cáncer y una muestra que contiene células apropiadas para su uso en un ensayo de expresión.

Asimismo, la naturaleza de la muestra que contiene células no es limitativa, dado que en los métodos divulgados puede utilizarse tejido reciente, tejido reciente congelado y tejido fijado, tal como tejido fijado en formalina e incluido en parafina (FEPE).

- 5 En una realización, la divulgación proporciona la identificación de los patrones de expresión génica analizando la expresión génica global, o casi global, de células individuales o de poblaciones de células homogéneas que se han diseccionado, o aislado o purificado de otra forma, de células contaminantes más allá de lo posible mediante una simple biopsia. Dado que la expresión de numerosos genes fluctúa entre células de diferentes pacientes, así como entre células de la misma muestra del paciente, los niveles de expresión génica pueden determinarse en correspondencia con uno o más genes de "control" o de "normalización", cuya expresión (o expresiones) es relativamente constante en las células de un paciente o entre pacientes.

15 En otro aspecto, la divulgación incluye medios físicos y metodológicos para detectar la expresión de uno o más genes, identificados por los modelos generados por patrones de expresión individuales. Estos medios pueden dirigirse a ensayar uno o más aspectos del molde (o moldes) de ADN subyacentes a la expresión del gen (o genes), del ARN utilizado como un producto intermedio para expresar el gen (o genes), o del producto proteico expresado por el gen (o genes).

20 Una ventaja proporcionada por la divulgación es que las células no cancerosas, contaminantes (tales como linfocitos infiltrantes u otras células del sistema inmunitario) no están presentes para posiblemente afectar a los genes identificados o al análisis posterior de la expresión génica para identificar la recidiva del cáncer y/o el resultado de supervivencia de los pacientes. Tal contaminación está presente cuando se utiliza una biopsia que contiene muchos tipos de células para ensayar perfiles de expresión génica.

25 Aunque la presente divulgación se describe principalmente en el contexto de cáncer humano, tal como cáncer de mama, este puede llevarse a la práctica en el contexto de cáncer de cualquier animal. Los animales preferidos para la aplicación de la presente divulgación son los mamíferos, en particular, los que son importantes para aplicaciones agrícolas (tales como, pero sin limitación, ganado, ovejas, caballos y otros "animales de granja"), modelos animales de cáncer, y animales de compañía (tales como, pero sin limitación, perros y gatos).

30 Los métodos proporcionados por la divulgación también pueden automatizarse completa o parcialmente.

Kits

35 Los materiales para su uso en los métodos de la presente divulgación son idealmente adecuados para la preparación de kits producidos según procedimientos bien conocidos. Por tanto, la divulgación proporciona, kits que comprenden agentes para la detección de la expresión de los genes divulgados para clasificar tumores o determinar resultados para el cáncer. Tales kits comprenden, opcionalmente, el agente con una descripción o etiqueta identificativa o instrucciones con respecto a su uso en los métodos de la presente divulgación. Tal kit pueden comprender recipientes, cada uno con uno o más de los diversos reactivos (normalmente en forma concentrada) utilizados en los métodos, incluyendo, por ejemplo, micromatrices prefabricadas, tampones, nucleótidos trifosfato apropiados (por ejemplo, dATP, dCTP, dGTP y dTTP; o rATP, rCTP, rGTP y UTP), transcriptasa inversa, ADN polimerasa, ARN polimerasa, y uno o más complejos de cebadores de la presente divulgación (por ejemplo, cebadores de poli(T) de longitud apropiada o aleatorios unidos a un promotor que reacciona con la ARN polimerasa).

45 Normalmente, también se incluirá un conjunto de instrucciones.

Habiendo proporcionado ahora, en líneas generales, la divulgación, esta se entenderá más fácilmente con referencia a los siguientes ejemplos que se proporcionan como ilustración y, salvo que se especifique, no se pretende que limiten la divulgación.

Ejemplos

Ejemplo I: General

Pacientes y muestras de tumor

55 Se utilizaron muestras de la cohorte NCIC CTG MA.17 (véase Goss *et al.*, J. Clin. Oncol., 26(12): 1948-1955, 2008). Se utilizaron 100 casos con 200 controles. Los 100 casos incluyeron 61 casos de recidiva de cáncer a distancia; 17 casos de recidiva de cáncer local; 5 casos de recidiva de cáncer regional; 16 casos de recidiva contralateral; y 1 caso desconocido. De estos, los casos contralaterales y desconocidos se excluyeron.

65 Los datos del seguimiento clínico estaban disponibles para las muestras utilizadas, las cuales eran bloques de tumor fijados en formalina e incluidos en parafina (FFPE) desde el momento del diagnóstico. Se calcularon las razones de probabilidad con el análisis del ICM, la proporción H:I, HoxB13 y el IGM, como variables continuas y categóricas. El análisis multivariado también incluyó la edad, el grado de tumor y el tratamiento en el análisis. Interacción de tratamiento: la edad y el grado de tumor también se incluyeron en el análisis. Los valores de P se calcularon para el

periodo de interacción.

La Tabla 1 resume las características del estudio de casos y controles (N=249).

5

Tabla 1. Características de los pacientes y tumores

<u>Descripción del Factor</u>		<u>MA-17 global (n=5157)</u>	<u>Estudio de casos y controles (n =249)</u>	<u>Casos (n=83)</u>	<u>Controles (n=166)</u>	<u>Valor de P</u>
Edad	<50		9 (4 %)	4 (5 %)	5 (3 %)	0,64
-	>=50. < 60		83 (33 %)	27 (32 %)	56 (34 %)	
	>=60. <70		82 (33 %)	24 (29 %)	58 (35 %)	
	>=70		75 (30 %)	28 (34 %)	47 (28 %)	
Grado de tumor	1		26 (10 %)	6 (7 %)	20 (12 %)	0,28
	2		166 (67 %)	54 (65 %)	112 (67,5 %)	
-	3		57 (23 %)	23 (27 %)	34 (20,5 %)	
Tipo de tumor	Ductal		218 (88 %)	71 (86 %)	147 (88,6 %)	0,63
-	Lobular		31 (12 %)	12 (14 %)	19 (11,4 %)	
Fase N	N0		94 (38 %)	31 (37 %)	63 (38 %)	0,45
	N1		138 (55 %)	44 (53 %)	94 (57 %)	
	N2.N3.NX		17 (7 %)	8 (10 %)	9 (5 %)	
Fase T	T1		110 (44 %)	37 (45 %)	73 (44 %)	0,58
	T2		111 (45 %)	35 (42 %)	76 (46 %)	
	T3		21 (8 %)	7 (8 %)	14 (8 %)	
	T4. TX		7 (3 %)	4 (5 %)	3 (2 %)	
Tratamiento anterior a la quimioterapia	No		148 (59 %)	49 (59 %)	99 (60 %)	0,96
	Sí		101 (41 %)	34 (41 %)	67 (40 %)	
Radiación anterior	No		150 (60 %)	49 (59 %)	101 (61 %)	0,89
Tratamiento	Sí		99 (40 %)	34 (41 %)	65 (39 %)	
Grupo de tratamiento	Letrozol		122 (49 %)	31 (37 %)	91 (55 %)	0,01
	Placebo		127 (51 %)	52 (63 %)	75 (45 %)	

Ensayos de RT-PCR en tiempo real para la proporción H/I y el IGM

- 10 Las secuencias de cebadores y sondas para HOXB13 e IL17BR, así como los genes de control ESR1, PGR, CHDH, ACTB, HMBS, SDHA y UBC, se utilizaron como se describe anteriormente (Ma *et al.*, citado anteriormente). Las secuencias de cebadores y sondas para los cinco genes de grado molecular (BUB1B, CENPA, NEK2, RACGAP1 y RRM2) así como de ERBB2 (HER2) se prepararon utilizando Primer Express (ABI).
- 15 Para la extracción de ARN se utilizaron cortes de cada una de las muestras FFPE. Se utilizó disección macroscópica para enriquecer el contenido tumoral. La extracción de ARN, la transcripción inversa y la RT-PCR TaqMan, utilizando el instrumento ABI 7900HT (Applied Biosystem, Inc), se realizaron como se describe anteriormente (Ma *et al.*, citado anteriormente). Los números de los umbrales de ciclado (CT, *cycling threshold*) se normalizaron con respecto al CT medio de cuatro genes de referencia (ACTB, HMBS, SDHA y UBC). El uso de estos genes está respaldado por informes previos con respecto a estos genes y a las secuencias representativas de cada uno de estos genes conocidos por el experto en la materia. Para representar los niveles relativos de expresión génica se
- 20

tomaron los CT normalizados.

Cálculo de la proporción H/I y del IGM

5 En general, y con respecto al IGM, se prefiere que los niveles de expresión de los genes divulgados se combinen para formar un solo índice que sirva como un fuerte factor de pronóstico e indicador del desenlace (o desenlaces) clínico. El índice es una suma de los niveles de expresión de los genes utilizados y utiliza coeficientes determinados a partir de un componente principal para combinar casos de más de un gen divulgado en un solo índice. Los coeficientes están determinados por factores tales como la desviación típica de cada uno de los niveles de expresión de los genes a través de un conjunto de datos representativo, y el valor de expresión de cada gen en cada muestra. Se controla la calidad del conjunto de datos representativo basándose en los valores de expresión promedios para el gen (o genes) de referencia como se divulga en el presente documento.

15 Dicho de otra manera, y con respecto al IGM, los niveles de expresión normalizados para los cinco genes a partir de las micromatrices o RT-PCR se normalizaron con respecto a una media de 0 y una desviación típica de 1 entre muestras dentro de cada conjunto de datos y después se combinaron en un solo índice por muestra a través del análisis de componentes principales (ACP), que utiliza el primer componente principal. La normalización de los datos de expresión primarios dentro de cada conjunto de datos fue necesaria para tener en cuenta la diferentes plataformas (micromatrices y RT-PCR) y los tipos de muestra (congeladas y FFPE). Como resultado, y siguiendo los parámetros de cuantificación, se genera una fórmula para la suma de los valores de expresión que define el índice. La precisión de los parámetros de cuantificación pueden después analizarse basándose en las medidas, los errores típicos y las desviaciones típicas (con intervalos de confianza) de los niveles de expresión de los genes en todo el conjunto de datos. Por lo tanto, la generación de la fórmula para el índice es dependiente del conjunto de datos, del gen de referencia y de los genes del IGM.

25 La proporción HOXB13:IL17BR se calculó como la diferencia en los niveles de expresión normalizados entre HOXB13 e IL17BR como se describe anteriormente (Ma *et al.*, citado anteriormente). Las medias y las desviaciones típicas para HOXB13 e IL17BR, utilizadas para la normalización de la cohorte de la Tabla 1, pueden obtenerse de un análisis de 190 cortes de tejido FFPE procedentes una cohorte basada en una población distinta de pacientes con cáncer de mama positivos para receptores de estrógenos sin afectación de ganglios linfáticos.

35 Para el IGM, obviamente se eliminaron los valores de C_T sin procesar anómalos antes de calcular el promedio de los valores sobre los duplicados para cada gen y para cada muestra. El valor de C_T sin procesar promedio para cada gen se normalizó después mediante el valor de C_T promedio de cuatro genes de referencia (ACTB, HMBS, SDHA y UBC). Los niveles de expresión normalizados (ΔC_T) para los cinco genes se combinaron en un solo índice por muestra, que puede compararse con un valor límite predeterminado, tal como 0, en el que un IGM alto está por encima del límite y un IGM bajo está por debajo del límite.

ICM continuo

40 Se construyó un modelo de riesgo continuo combinando, como variables continuas, la proporción H:I y el IGM. La linealidad de estas dos variables se comprobó mediante un ajuste a un modelo de regresión de riesgos instantáneos proporcionales de Cox con ejes cúbicos restringidos, y la proporción H:I demostró no linealidad significativa. Se utilizó una función polinómica de H:I para aproximar los modelos restringidos utilizando el Criterio de Información de Akaike. Después, el indicador resultante del modelo de regresión de Cox final se recuantificó en el intervalo de 0 a 10, lo que se denomina ICM.

50 El ICM se categorizó adicionalmente en tres niveles: riesgo bajo, $ICM < 5$; riesgo intermedio, $5 \leq ICM < 6,4$; riesgo alto, $ICM > 6,4$. Estos límites se escogieron de forma que las proporciones resultantes de los grupos de riesgo bajo, intermedio y alto fueron similares a las formadas por los tres grupos de combinación categóricos de la proporción H:I y el IGM.

Límites y análisis estadístico

55 LÍMITE DE H/I: en este estudio puede utilizarse el límite de 0,06 para la proporción HOXB13:IL17BR, previamente definido para estadificar pacientes tratados con tamoxifeno adyuvante en riesgo bajo y alto de recidiva.

60 LÍMITE DEL IGM: el cálculo y el límite para el IGM se definieron sin utilizar ningún dato de desenlace clínico y, en su lugar, fue un límite natural. El análisis inicial de IGM en la cohorte de Uppsala indicó una buena discriminación de tumores de grado 1 y grado 3 utilizando la media (0) como límite, y la agrupación basada en modelos del IGM también indicó una distribución bimodal con un límite natural de alrededor de 0. Este límite se confirmó adicionalmente con el análisis de la curva de eficacia diagnóstica (ROC, acrónimo de *receiver operating characteristics*).

65 ANÁLISIS ESTADÍSTICO: para evaluar la asociación de los índices de expresión génica con el desenlace clínico, se realizó análisis de Kaplan-Meier con una prueba de rango logarítmico y regresión de riesgos instantáneos

proporcionales de Cox. Para evaluar la capacidad pronóstica de los índices de expresión génica después de ajustar factores pronósticos conocidos, se realizaron modelos de regresión de Cox multivariable.

5 La suposición de riesgos instantáneos proporcionales (RIP) se comprobó a través de restos de Schoenfeld a escala; las variables que infringían la suposición de los RIP se ajustaron en el modelo mediante estratificación. Para explicar el diseño de casos y cohortes de la cohorte de la Tabla 1, se utilizó el análisis ponderado de Kaplan-Meier y modelos de regresión de Cox con modificaciones para manejar los diseños de casos y cohortes (véase ^{19,20}, como se implementó en el paquete de supervivencia en R (www.r-project.org)). Para probar la interacción entre el IGM dicotomizado y la proporción H:I en los modelos de regresión de Cox, se utilizó la estadística de Wald en la cohorte de la Tabla 1 y se utilizó la prueba del coeficiente de verosimilitudes en la última cohorte.

Las correlaciones de las variables continuas con los factores categóricos se examinaron utilizando la prueba de Wilcoxon de dos muestras, no paramétrica, o la prueba de Kruskal-Wallis para factores con más de dos niveles.

15 Todos los análisis estadísticos se realizaron en el entorno estadístico R. Todas las pruebas de significación estadística fueron bilaterales y un valor de p <0,05 se consideró significativo.

Ejemplo II: Interpretación del pronóstico

20 La Tabla 2 muestra la distribución de los casos y controles en los grupos de riesgo de ICM continuos.

Tabla 2:

	Casos (n=83)	Controles (n=166)
Grupo de ICM (%)		
Bajo	43,4 %	57,8 %
Intermedio	22,9 %	18,1 %
Alto	33,7 %	24,1 %

La Tabla 3 muestra el análisis univariable en relación con recidiva de cáncer en los sujetos de la cohorte MA. 17

25

Tabla 3. Análisis univariable en relación con recidiva de cáncer		
	Razón de posibilidades (IC 95 %)	Valor de P
Tratamiento (Placebo frente a Letrozol)	2,02 (1,17 - 3,47)	0,01
Grado Tumoral		0,28
II frente a I	1,73 (0,61 - 4,88)	0,30
III frente a I	2,53 (0,81 - 7,90)	0,11
Análisis del ICM		
ICM	2,38 (1,21 - 4,69)	0,01
ICM, Alto frente a Bajo	1,87 (1,00 - 3,50)	0,05
Análisis de componentes del ICM		
HoxB13	1,34 (1,05 - 1,70)	0,02
HoxB13, Alto frente a Bajo	2,17 (1,27 - 3,69)	0,004
H:I	2,52 (1,08 - 5,85)	0,03
H:I, Alto frente bajo	1,68 (1,00 - 2,81)	0,049
ICM	1,83 (0,93 - 3,58)	0,08
ICM, Alto frente a Bajo	1,49 (0,87 - 2,56)	0,15

La Tabla 4 muestra el análisis multivariable en relación con la recidiva de cáncer

Tabla 4. Análisis multivariable en relación con recidiva de cáncer		
Análisis con ICM		
	Razón de posibilidades (IC 95 %)	Valor de P
ICM	2,37 (1,08 - 5,22)	0,03
ICM, Alto frente a Bajo	1,87 (0,88 - 3,95)	0,10
Análisis con componentes del ICM		
	Razón de posibilidades (IC 95 %)	Valor de P
HoxB13	1,35 (1,05 - 1,74)	0,02
HoxB13, Alto frente a Bajo	2,32 (1,32 - 4,10)	0,004
H:I	2,55(1,03-6,32)	0,04
H:I, Alto frente bajo	1,71 (0,98-2,97)	0,06
ICM	1,61 (0,73 - 3,54)	0,24
ICM, Alto frente a Bajo	1,37 (0,73 - 2,55)	0,33

Como se demuestra por lo anterior, el ICM es un índice pronóstico de recidivas tardías de cáncer en pacientes RE+ después de 5 años de tratamiento con tamoxifeno. La expresión de HoxB 13 también es pronóstica de recidivas tardías de cáncer en pacientes RE+ después de 5 años de tratamiento con tamoxifeno.

5 **Ejemplo III: Interacción entre biomarcador y tratamiento**

La Tabla 5 muestra interacciones entre la expresión génica analizada y el tratamiento.

	Valor de P
HoxB13	0,047
H:I	0,97
IGM	0,06
ICM	0,42
ICM Alto frente a Intermedio + Bajo	0,08

10 La Tabla 6 muestra la distribución de la expresión génica de HoxB13 en relación con los tratamientos utilizados. La expresión de HoxB13 en el diagnóstico predice un beneficio para el paciente de la terapia endocrina prolongada con letrozol después 5 años de terapia con tamoxifeno adyuvante.

Tabla 6

	Letrozol		Placebo		Valor de P
	Controles	Casos	Controles	Casos	
HoxB13 Bajo	48	14 (23 %)	46	16 (26 %)	0,83
HoxB13 Alto	43	17 (28 %)	29	36 (55 %)	0,004

15 Bibliografía:

1. Ma *et al.*, Cancer Cell, 5: 607-16 (2004)
 2. Ma *et al.*, J. Clin. Oncol., 24: 4611-9 (2006)
 20 3. Goetz *et al.*, Clin. Cancer Res., 12: 2080-7 (2006)
 4. Jerevall *et al.*, Breast Cancer Res. Treat (2007)
 5. Jansen *et al.*, J. Clin. Oncol. 25: 662-8 (2007)
 6. Cianfrocca *et al.*, Oncologist, 9: 606-16 (2004)
 7. Sotiriou *et al.*, J. Natl. Cancer Inst., 98: 262-72 (2006)
 25 8. van't Veer *et al.*, Nature, 415: 530-6 (2002)
 9. Paik *et al.*, N. Engl. J. Med., 351: 2817-26 (2004)
 10. Desmedt *et al.*, Cell Cycle, 5: 2198-202 (2006)
 11. Loi *et al.* J. Clin. Oncol., 25: 1239-46 (2007)
 12. Sotiriou *et al.*, Nat. Rev. Cancer, 7: 545-53 (2007)
 30 13. Miller *et al.*, Proc. Natl. Acad. Sci. USA, 102: 13550-5 (2005)
 14. Pawitan *et al.*, Breast Cancer Res. 7: R953-64 (2005)
 15. Rundle *et al.*, Cancer Epidemiol Biomarkers Prev., 14: 1899-907 (2005)
 16. Ma *et al.*, Proc. Natl. Acad. Sci. USA, 100: 5974-9 (2003)
 17. Whitfield *et al.*, Mol. Biol. Cell, 13: 1977-2000 (2002)
 35 18. Hirose *et al.*, J. Biol. Chem., 276: 5821-5828 (2001)
 19. Goldhirsch *et al.*, Ann. Oncol., 16: 1569-83 (2005)

REIVINDICACIONES

1. Un método para determinar si se espera que un sujeto, que se ha sometido a la extirpación de un cáncer de mama RE+ y que se ha tratado con una primera terapia endocrina durante un periodo de tiempo sin recidiva de cáncer, se beneficie de un cambio a una segunda terapia endocrina distinta, que comprende:
- 5 preparar ADNc a partir de ácidos nucleicos de una muestra de células de cáncer de mama RE+ del sujeto, determinar el nivel de expresión del gen *HoxB13* a partir de dicho ADNc, y clasificar al sujeto como que se espera que se beneficie del tratamiento con una segunda terapia endocrina distinta después de suspender la primera terapia endocrina, en el que dicha clasificación está basada en el nivel de expresión elevado de *HoxB13*,
- 10 en el que (i) la primera terapia endocrina es un tratamiento con un modulador selectivo de receptores de estrógeno (SERM) o un regulador negativo selectivo de receptores de estrógeno (SERD) y la segunda terapia endocrina es un tratamiento con un inhibidor de aromatasa. Como alternativa, la primera terapia endocrina puede ser el tratamiento con un inhibidor de aromatasa (IA), o (ii) la primera terapia endocrina es un tratamiento con un IA y la segunda terapia endocrina es un tratamiento con un SERM o un SERD.
- 15 2. El método de acuerdo con la reivindicación 1, en el que:
- 20 el nivel de expresión del gen *IL17BR* a partir de dicho ADNc se determina para calcular la proporción de los niveles de expresión de *HOXB13:IL17BR* (proporción H:I), y dicha proporción se utiliza para indicar un nivel de expresión elevado de *HoxB 13*.
- 25 3. El método de la reivindicación 1 o 2, en el que dichos ácidos nucleicos son ARNm procedente de dicha muestra.
4. El método de la reivindicación 3, en el que dicho ARN se utiliza para una PCR (reacción en cadena de la polimerasa).
- 30 5. El método de la reivindicación 1 o 2, en el que dicha determinación comprende utilizar una matriz.
6. El método de la reivindicación 1 o 2, en el que dicha muestra se disecciona de tejido extirpado de dicho sujeto.
7. El método de la reivindicación 4, en el que dicha PCR es una RT-PCR (PCR con transcripción inversa), opcionalmente una RT-PCR en tiempo real.
- 35 8. El método de la reivindicación 1 o 2, en el que dicha muestra es una muestra fijada en formalina e incluida en parafina (FFPE).
- 40 9. El método de la reivindicación 1 o 2, en el que dicho cáncer es carcinoma ductal *in situ* (CDIS) y dicha recidiva de cáncer comprende una recidiva local.
10. El método de cualquier reivindicación precedente, en el que la primera terapia endocrina es un tratamiento con tamoxifeno y la segunda terapia endocrina es un tratamiento con letrozol.
- 45 11. Una segunda terapia endocrina para su uso en un método de tratamiento de cáncer de mama RE+ en un sujeto que se ha sometido extirpación de cáncer de mama RE+ y que se ha tratado con una primera terapia endocrina durante un periodo de tiempo sin recidiva de cáncer, en el dicho método comprende clasificar al sujeto como que se espera que se beneficie de una segunda terapia endocrina de acuerdo con una cualquiera de las reivindicaciones 1 a 9, y en el que:
- 50 (i) la primera terapia endocrina es un tratamiento con un modulador selectivo de receptores de estrógeno (SERM) o un regulador negativo selectivo de receptores de estrógeno (SERD) y la segunda terapia endocrina es un tratamiento con un inhibidor de aromatasa (IA); o
- 55 (ii) la primera terapia endocrina es un tratamiento con un IA y la segunda terapia endocrina es un tratamiento con un SERM o un SERD.
12. La segunda terapia endocrina para su uso en un método de la reivindicación 11, en el que la primera terapia endocrina es tamoxifeno y la segunda terapia endocrina es letrozol.