

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 829 912**

51 Int. Cl.:

C12Q 1/6886 (2008.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **10.09.2014** **PCT/US2014/055041**

87 Fecha y número de publicación internacional: **19.03.2015** **WO15038682**

96 Fecha de presentación y número de la solicitud europea: **10.09.2014** **E 14843312 (1)**

97 Fecha y número de publicación de la concesión europea: **09.09.2020** **EP 3044335**

54 Título: **Predecir la reaparición de cáncer de mama**

30 Prioridad:

11.09.2013 US 201361876757 P

45 Fecha de publicación y mención en BOPI de la
traducción de la patente:
02.06.2021

73 Titular/es:

BIO THERANOSTICS, INC. (100.0%)
9640 Towne Centre Drive, Suite 200
San Diego, CA 92121, US

72 Inventor/es:

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

74 Agente/Representante:

PONS ARIÑO, Ángel

ES 2 829 912 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

Predecir la reaparición de cáncer de mama

5 Antecedentes de la invención

(1) Campo de la invención

La divulgación se refiere a la identificación y el uso de perfiles o patrones de expresión génica, con relevancia clínica para la reaparición de cáncer de mama. En particular, la divulgación proporciona ensayos para determinar la probabilidad de reaparición de cáncer después del tratamiento inicial con una terapia contra el cáncer de mama, tales como tamoxifeno adyuvante o un inhibidor de la aromatasa.

(2) Descripción de la técnica relacionada

El cáncer de mama receptor de estrógeno positivo es una enfermedad con un riesgo prolongado de reaparición. Después de 5 años de tamoxifeno adyuvante, los pacientes tienen un riesgo sostenido de reaparición de la enfermedad y muerte durante al menos 15 años después del diagnóstico. El seguimiento a largo plazo de ensayos iniciales fundamentales de inhibidores de la aromatasa adyuvantes, incluido el ensayo Arimidex, Tamoxifeno, Solo o en Combinación (ATAC, por sus siglas en inglés) y el estudio del Breast International Group (BIG) 1-98 (Cuzick et al., 2010), muestran una tasa continua de reaparición de aproximadamente 2 % por año después de la terapia inicial, más de la mitad de todas las reapariciones se producen después de 5 años de terapia endocrina adyuvante. Estos hallazgos enfatizan la necesidad de una terapia adyuvante prolongada y un biomarcador que pueda guiar el proceso de toma de decisiones sobre el tratamiento.

Las firmas de expresión multigénicas estudiadas en la última década para la evaluación del riesgo de reaparición en el cáncer de mama receptor de estrógeno positivo se basan principalmente en la medición cuantitativa de la expresión génica relacionada con la proliferación. Estas firmas multigénicas, incluida la puntuación de reaparición de 21 genes (Oncotype DX; Genomic Health, Redwood City, CA, EE.UU.), son fuertes predictores de reaparición distante, pero su capacidad pronóstica disminuye al evaluar el riesgo más allá de los 5 años desde el diagnóstico (Sgroi et al., 2012). Por el contrario, los predictores de reaparición tardía no están bien caracterizados y pueden estar asociados diferentes mecanismos con reapariciones tempranas y tardías. Se necesitan biomarcadores para identificar a los pacientes tratados de forma adecuada con solo 5 años de terapia endocrina y, a la inversa, aquellos con mayor riesgo de reaparición tardía que podrían justificar una terapia endocrina adyuvante u otra terapia prolongadas.

El trabajo anterior desarrolló y validó el ensayo del índice de cáncer de mama (BCI, por sus siglas en inglés) que consta de dos conjuntos de biomarcadores de expresión génica desarrollados de forma independiente: índice de grado molecular (MGI, por sus siglas en inglés) y *HOXB13/IL17BR*. El MGI, un predictor de cinco genes que recapitula el grado y la proliferación del tumor, es altamente pronóstico en pacientes con cáncer de mama receptor de estrógeno positivo. *HOXB13/IL17BR*, que se desarrolló independientemente del grado o la proliferación del tumor, es pronóstico de reapariciones distantes tempranas y tardías, y es predictivo del beneficio prolongado del inhibidor de la aromatasa adyuvante en pacientes con cáncer de mama receptor de estrógeno positivo en estadio temprano. Tanto el BCI como los ensayos de puntuación de reaparición de 21 genes miden la expresión génica mediante PCR cuantitativa en tiempo real, aunque difieren en los genes que detectan. IHC4 es otro modelo de pronóstico que mide la expresión de proteínas de cuatro de los biomarcadores inmunohistoquímicos más informativos: receptores de estrógenos, receptores de progesterona, HER2 y Ki-67 (Cusick et al., 2011), ninguno de los cuales está codificado por genes en el ensayo BCI. El BCI no se ha evaluado en pacientes con cáncer de mama receptor de estrógeno negativo o triple negativo. Véanse las Patentes de EE.UU. 7.930.105 y 7.504.214, Publicaciones de Patente de EE.UU. 2011/0136680 y 2013/0281502, y publicación de patente PCT WO/2012/079059.

El documento WO 2012/079059 divulga tratamientos posteriores para el pronóstico del cáncer de mama. Laurel A Habel et al. (Breast Cancer Res. 14 de marzo de 2013;15(2):R24) divulga *HOXB13:IL17BR* y el índice de grado molecular y el riesgo de muerte por cáncer de mama entre pacientes con enfermedad invasiva ganglios linfáticos negativa. El documento WO 2009/108215 divulga la clasificación de tumores y el pronóstico del cáncer. Ma XJ et al. (J Clin Oncol. 1 de Oct de 2006; 24(28):4611-9) divulga que el índice de expresión de *HOXB13:IL17BR* es un factor pronóstico en el cáncer de mama en estadio temprano.

Por tanto, está claro que existe una necesidad de biomarcadores para mejorar el riesgo-beneficio de la terapia endocrina adyuvante prolongada para la reaparición tardía en pacientes con cáncer de mama receptor de estrógeno positivo. La presente invención aborda esta necesidad.

Breve resumen de la invención

La presente invención se refiere a un método para evaluar el cáncer de mama de un sujeto, comprendiendo el método medir los niveles de expresión del ARNm de *HoxB13*, *IL17BR*, *Bub1B*, *CENPA*, *NEK2*, *RACGAP1* y *RRM2* en una muestra de células de cáncer de mama ER+ del sujeto; calcular una relación de niveles de expresión de

HoxB13/IL17BR (H:I); sumar los niveles de expresión de Bub1B, CENPA, NEK2, RACGAP1 y RRM2 utilizando coeficientes para obtener un índice MGI; combinar linealmente H:I y MGI como variables continuas para formar un valor de Índice de Cáncer de Mama (BCI); y clasificar la muestra, basándose en el valor de BCI, como se indica:

- 5 (i) un riesgo bajo o un riesgo alto de reaparición de cáncer en el sujeto, sin categoría de riesgo intermedio, en donde un valor de BCI mayor se correlaciona con un mayor riesgo de reaparición de cáncer y un valor de BCI menor se correlaciona con un menor riesgo de reaparición de cáncer;
- 10 (ii) necesidad o falta de necesidad de tratamiento, en donde un valor de BCI menor se correlaciona con no necesitar tratamiento con un inhibidor de la aromatasa, terapia dirigida o terapia endocrina después de un tratamiento inicial con un inhibidor de la aromatasa, terapia dirigida o terapia endocrina durante cinco años o menos.

15 La divulgación se basa en parte en el descubrimiento de que (a) un esquema de dos categorías (riesgo alto, riesgo bajo) se puede utilizar de manera eficaz en el análisis de BCI para evitar la incertidumbre de la clasificación de riesgo intermedio de la técnica anterior; (b) un modelo de BCI lineal (BCI-L) tiene una capacidad de pronóstico superior para el riesgo de reaparición que un modelo cúbico (BCI-C); y (c) los descubrimientos anteriores permiten una aplicación más simple y precisa de BCI para proporcionar información de pronóstico, tal como la reaparición de cáncer e información predictiva, tal como la capacidad de respuesta a determinadas terapias, que se puede utilizar para la
20 selección de opciones terapéuticas.

En el presente documento se proporciona un método para determinar el riesgo de reaparición de cáncer en un sujeto afectado con cáncer de mama. El método comprende (a) determinar los niveles de expresión del ARNm de una pluralidad de genes en una muestra de células de cáncer de mama ER+ del sujeto; y (b) clasificar si el sujeto tiene un
25 riesgo bajo o alto de reaparición de cáncer basándose en el análisis de los niveles de expresión del ARNm de la pluralidad de genes en el momento del diagnóstico de la enfermedad del cáncer de mama. En este método, el análisis de la pluralidad de genes proporciona un riesgo de reaparición de cáncer después de recibir aproximadamente cinco años de terapia adyuvante que es menos de aproximadamente el 5 % en el grupo de riesgo bajo en comparación con conjuntos de datos retrospectivos de pacientes con cáncer de mama ER+ con más de cinco años de resultado, o
30 muestras representativas de los mismos. En muchas realizaciones, el sujeto ha recibido aproximadamente cinco años de terapia adyuvante y está sin enfermedad después de esa terapia.

Por tanto, en algunas realizaciones, la presente divulgación está dirigida a un método para determinar el riesgo de reaparición de cáncer en un sujeto afectado con cáncer de mama. El método comprende
35 determinar los niveles de expresión del ARNm de *HoxB13*, *IL17BR*, *Bub1B*, *CENPA*, *NEK2*, *RACGAP1* y *RRM2* en una muestra de células de cáncer de mama ER+ del sujeto; sumar los niveles de expresión para formar un valor del Índice de Cáncer de Mama (BCI) donde un valor de BCI más alto se correlaciona con un riesgo más alto de reaparición de cáncer y un valor de BCI más bajo se correlaciona con un riesgo más bajo de reaparición de cáncer; y
40 clasificar la muestra, basándose en el valor de BCI, como indicativa de un riesgo bajo o alto de reaparición de cáncer en el sujeto, sin categoría de riesgo intermedio.

En otras realizaciones, la divulgación está dirigida a un método para determinar la capacidad de respuesta al tratamiento de un sujeto afectado con cáncer de mama. El método comprende
45 determinar los niveles de expresión del ARNm de *HoxB13*, *IL17BR*, *Bub1B*, *CENPA*, *NEK2*, *RACGAP1* y *RRM2* en una muestra de células de cáncer de mama ER+ del sujeto; sumar los niveles de expresión para formar un valor de Índice de Cáncer de Mama (BCI), donde un valor de BCI más bajo se correlaciona con la capacidad de respuesta al tratamiento adicional con un inhibidor de la aromatasa, terapia dirigida o terapia endocrina después de un tratamiento inicial con un inhibidor de la aromatasa, terapia dirigida o terapia
50 endocrina durante cinco años o menos; y clasificar la muestra, basándose en el BCI, como que indica dicha capacidad de respuesta o falta de ella.

Adicionalmente, la divulgación está dirigida a un método para tratar a un sujeto afectado con cáncer de mama. El método comprende
55 determinar los niveles de expresión del ARNm de *HoxB13*, *IL17BR*, *Bub1B*, *CENPA*, *NEK2*, *RACGAP1* y *RRM2* en una muestra de células de cáncer de mama ER+ del sujeto; sumar los niveles de expresión para formar un valor de Índice de Cáncer de Mama (BCI), donde un valor de BCI más bajo se correlaciona con (a) la capacidad de respuesta al tratamiento adicional con un inhibidor de la aromatasa, terapia dirigida o terapia endocrina después de un tratamiento inicial con un inhibidor de la aromatasa, terapia dirigida o terapia endocrina durante cinco años o menos y (b) un menor riesgo de reaparición distante; y
60 tratar al sujeto de manera consistente con la determinación del valor de BCI para el sujeto.

Breve descripción de los dibujos

- 65 La FIG. 1 es un diagrama CONSORT de ATAC para el consort descrito en el presente documento.
La FIG. 2 son gráficos que muestran el rendimiento de grupos de riesgo especificados previamente basándose en

BCI-C y BCI-L para las reapariciones distantes generales en 10 años en todos los pacientes ER+ NO. Panel A - BCI-C; Panel B - BCI-L.

La FIG. 3 es un gráfico que muestra el riesgo general de reaparición distante en 10 años en función del índice lineal BCI continuo en pacientes ER+ NO.

La FIG. 4A y 4B son gráficos que muestran el rendimiento de grupos de riesgo especificados previamente basándose en los modelos BCI-C y BCI-L para la reaparición distante general en 10 años en pacientes ER+ NO HER2 negativo. Panel A) BCI-C; 4B) BCI-L.

La FIG. 5A y 5B son gráficos que muestran el rendimiento de los grupos de riesgo con BCI especificados previamente para las reapariciones distantes tempranas y tardías en pacientes con ER+ NO. 5A) reaparición distante temprana en 0-5 años; 5B) reaparición distante tardía en 5-10 años. La población PI se refiere a los grupos de riesgo bajo e intermedio especificados previamente, mientras que P2 se refiere al grupo de riesgo alto de reaparición temprana. P3 se refiere al grupo de riesgo bajo especificado previamente, mientras que P4 se refiere a los grupos de riesgo intermedio y alto de reaparición tardía.

La FIG. 6A y 6B son gráficos que muestran el riesgo de reaparición distante temprana y tardía en función del índice BCI continuo en pacientes ER+ NO. 6A) riesgo de reaparición distante temprana en 0-5 años; 6B) riesgo de reaparición distante tardía en 5-10 años. Las líneas verticales definen los bordes entre los grupos de riesgo con BCI bajo, intermedio (Inter) y alto especificados previamente. La FIG. 7 son gráficos que muestran el rendimiento de los grupos de riesgo con BCI especificado previamente para las reapariciones distantes tempranas y tardías en pacientes ER+ NO HER2 negativo. Panel A: reaparición distante temprana (0-5 años); B) riesgo de reaparición distante tardía (5-10 años).

La FIG. 8 son gráficos que muestran el riesgo de reaparición distante temprana (0-5 años) y tardía (5-10 años) en función del índice BCI en pacientes ER+ NO HER2 negativo. Panel A: riesgo de reaparición distante temprana; Panel B: riesgo de reaparición distante tardía.

La FIG. 9 son gráficos que muestran el rendimiento de los grupos de riesgo de BCI y RS especificados previamente y los grupos de riesgo categórico determinados post-hoc de IHC4 para las reapariciones distantes generales en 10 años en pacientes ER+ NO, ambos brazos combinados y el brazo de anastrozol (ANA) y tamoxifeno (TAM) por separado. Panel A - BCI en ambos brazos combinados; Panel B - RS en ambos brazos combinados; Panel C - IHC4 en ambos brazos combinados; Panel D - BCI en el brazo de anastrozol solo; Panel E - RS en el brazo de anastrozol solo; Panel F - IHC4 en el brazo de anastrozol solo; Panel G - BCI en el brazo de tamoxifeno solo; Panel H - RS en el brazo de tamoxifeno solo; Panel I - IHC4 en el brazo de tamoxifeno solo.

La FIG. 10 es un gráfico que muestra el rendimiento de los grupos de riesgo especificados previamente para las reapariciones distantes generales en 10 años en pacientes ER+ ganglios positivos.

Descripción detallada de la invención

Como se usa en el presente documento, las formas en singular "un", "uno/una" y "el/la" pretenden incluir las formas en plural también, a menos que el contexto indique claramente lo contrario. Adicionalmente, el uso de "o" pretende incluir "y/o", a menos que el contexto indique claramente lo contrario.

Definiciones

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, reaparición de cáncer y/o resultados de supervivencia que se correlacionan con la capacidad de distinguir entre dichos resultados. En algunos casos, el resultado es el de cáncer de mama.

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

Los términos "correlacionar" 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 exclusión de uno o más estados distintos identificados mediante el uso de los métodos como se describe en el presente documento. Un gen puede expresarse en un nivel más alto o más bajo y aún estar correlacionado con uno o más estados o resultados de cáncer.

Un "polinucleótido" es una forma polimérica de nucleótidos de cualquier longitud, ya sean ribonucleótidos o desoxirribonucleótidos. Este término se refiere solo a la estructura primaria de la molécula. Por tanto, este término incluye ADN y ARN bicatenarios y monocatenarios. También incluye tipos conocidos de modificaciones 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 (p. ej., fosforotioatos, fosforditioatos, etc.), así como formas no modificadas del polinucleótido.

El término "amplificar" se usa en sentido amplio para significar que la creación de un producto de amplificación se puede preparar enzimáticamente con ADN o ARN polimerasas, por ejemplo, usando la reacción en cadena de la

polimerasa (PCR), como se sabe en la técnica. "Amplificación", como se usa en el presente documento, generalmente se refiere al proceso de producir múltiples copias de una secuencia deseada, particularmente las de una muestra. "Múltiples copias" significa al menos 2 copias. Una "copia" no significa necesariamente una complementariedad o identidad de secuencia perfecta respecto de la secuencia plantilla.

5 Por correspondiente se entiende que una molécula de ácido nucleico comparte una cantidad sustancial de identidad de secuencia con otra molécula de ácido nucleico. Cantidad sustancial significa al menos el 95 %, habitualmente al menos el 98 % y más habitualmente al menos el 99 %, y la identidad de la secuencia se determina utilizando el algoritmo BLAST, como se describe en Altschul et al., J. Mol. Biol. 215:403-410 (1990) (utilizando la configuración

10 predeterminada publicada, es decir, los parámetros $w = 4$, $t = 17$). Los métodos para amplificar el ARNm se conocen generalmente en la técnica e incluyen la PCR con transcripción inversa (RT-PCR) y los descritos en la Patente de EE.UU. 6.794.141. Otro método que puede usarse es la PCR cuantitativa (o Q-PCR). Como alternativa, el ARN puede marcarse directamente como el correspondiente ADNc mediante métodos conocidos en la técnica.

15 Una "micromatriz" es una matriz lineal o bidimensional de regiones preferentemente discretas, teniendo cada una un área, formada sobre la superficie de un soporte sólido tal como, pero sin limitación, vidrio, plástico, o membrana sintética. La densidad de las regiones discretas de una micromatriz se determina por el número total de polinucleótidos inmovilizados a detectar sobre la superficie de un único soporte en fase sólida, preferentemente al menos

20 aproximadamente 50/cm², más preferentemente al menos aproximadamente 100/cm², incluso más preferentemente al menos aproximadamente 500/cm², pero 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

25 polinucleótidos situada sobre un chip u otras superficies utilizada para hibridarse con polinucleótidos amplificados o clonados de una muestra. Como la posición de cada grupo particular de cebadores en la matriz es conocida, las identidades de los polinucleótidos de una muestra se pueden determinar basándose en su unión a una posición particular en la micromatriz.

Debido a que la divulgación se basa en la identificación de genes que están sobreexpresados o subexpresados, una

30 realización de la divulgación implica determinar la expresión mediante hibridación de ARNm, o una versión amplificada o clonada del mismo, de una célula de muestra a un polinucleótido que es exclusivo de 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

35 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 previa se refiere a un aumento o disminución de 1 del valor numérico establecido. Incluso más preferidos son los polinucleótidos de al menos o aproximadamente 50, al menos o aproximadamente 100, al menos o aproximadamente 150, al menos o

40 aproximadamente 200, al menos o aproximadamente 250, al menos o aproximadamente 300, al menos o aproximadamente 350, o al menos o 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 del 10 % del valor numérico establecido. Dichos polinucleótidos también pueden denominarse sondas de polinucleótidos que son capaces de hibridar con secuencias de los genes, o porciones

45 exclusivas de los mismos, descritas en el presente documento. Preferentemente, las secuencias son las del ARNm codificado por los genes, el ADNc correspondiente a dichos ARNm, y/o versiones amplificadas de dichas secuencias. En realizaciones preferidas de la divulgación, las sondas de polinucleótidos están inmovilizadas en una matriz, otros dispositivos, o en puntos individuales que localizan las sondas.

En otra realización de la divulgación, la totalidad o parte de una secuencia divulgada se puede amplificar y detectar por métodos tales como la reacción en cadena de la polimerasa (PCR) y variaciones de la misma, tales como, pero

50 sin limitación, PCR cuantitativa (Q-PCR), PCR con transcripción inversa (RT-PCR) y PCR en tiempo real, opcionalmente RT-PCR en tiempo real. Dichos métodos utilizarían uno o dos cebadores que son complementarios a porciones de una secuencia divulgada, donde los cebadores se utilizan para cebar la síntesis de ácidos nucleicos. Los ácidos nucleicos nuevamente sintetizados opcionalmente se marcan, y se pueden detectar directamente o mediante hibridación con un polinucleótido de la divulgación. Los ácidos nucleicos nuevamente sintetizados se pueden poner

55 en contacto con polinucleótidos (que contienen secuencias) de la divulgación en condiciones que permiten su hibridación.

Como alternativa, y en otra realización de la divulgación, la expresión génica se puede determinar mediante el análisis de la proteína expresada en una muestra celular de interés mediante el uso de uno o más anticuerpos específicos

60 para uno o más epítopos de productos génicos individuales (proteínas) en dicha muestra celular. Dichos anticuerpos se marcan preferentemente para permitir su fácil 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. Los marcadores adecuados incluyen radioisótopos, cromóforos de nucleótidos, enzimas,

65 sustratos, moléculas fluorescentes, restos quimioluminiscentes, partículas magnéticas, restos bioluminiscentes y similares. Como tal, una etiqueta es cualquier composición detectable por medios espectroscópicos, fotoquímicos,

bioquímicos, inmunoquímicos, 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 tal como portaobjetos de vidrio.

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

Una "muestra de tejido mamario" o "muestra de células mamarias" se refiere a una muestra de tejido mamario o líquido aislados de un individuo sospechoso de estar afectado o en riesgo de desarrollar, cáncer de mama. Estas muestras son aislados primarios (a diferencia de las células cultivadas) y pueden recogerse por cualquier medio no invasivo, incluyendo, pero sin limitación, lavado ductal, aspiración con aguja fina, biopsia con aguja, los dispositivos y métodos descritos en la Patente de EE.UU. N° 6.328.709, o cualquier otro medio adecuado reconocido en la técnica. Como alternativa, la "muestra" puede recogerse por un método invasivo, incluyendo, pero sin limitación, biopsia quirúrgica.

"Expresión" y "expresión génica" incluyen la transcripción y/o traducción de material de ácido nucleico. Por supuesto, el plazo también puede ser limitado, si así se indica, como refiriéndose únicamente a la transcripción de ácidos nucleicos.

Como se usa en el presente documento, la expresión "que comprende" y sus afines se utilizan en su sentido inclusivo; es decir, equivalente al término "incluido" y sus correspondientes afines.

Condiciones que "permiten" que se produzca un evento o condiciones que son "adecuados" para que se produzca un evento, tal como la hibridación, la extensión de la hebra, y similares, o las condiciones "adecuadas" son condiciones que no evitan que se produzcan tales eventos. Por tanto, estas condiciones permiten, potencian, facilitan y/o favorecen el evento. Tales condiciones, conocidas en la técnica y descritas en el presente documento, dependen de, por ejemplo, la naturaleza de la secuencia de nucleótidos, temperatura y condiciones del tampón. Estas condiciones también dependen de qué evento se desee, tal como la hibridación, escisión, extensión de la hebra o transcripción.

"Mutación" de secuencia, como se usa en el presente documento, se refiere a cualquier alteración de secuencia en la secuencia de un gen divulgado en el presente documento de interés en comparación con una secuencia de 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 la sustitución, eliminación o inserción. El polimorfismo de un solo nucleótido (SNP) también es una mutación de secuencia como se usa en el presente documento. Debido a que la presente divulgación se basa en el nivel relativo de expresión génica, las mutaciones en regiones no codificantes de genes como se divulga en el presente documento también pueden ensayarse en la práctica de la divulgación.

"Detección" incluye cualquier medio de detección, incluida la detección directa e indirecta de la expresión génica y los cambios en las mismas. Por ejemplo, pueden observarse productos "detectablemente menos" directa o indirectamente, y el término indica cualquier reducción (incluida la ausencia de señal detectable). De forma similar, producto "detectablemente más" significa cualquier aumento, ya sea observado directa o indirectamente.

Las diferencias en la expresión de las secuencias divulgadas entre dos afecciones que se evalúan (p. ej., riesgo alto o bajo de reaparición) se definen en los siguientes términos en base a cambios porcentuales o de factor en la expresión entre las dos afecciones. Las diferencias entre las dos afecciones pueden ser de 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 120, 140, 160, 180 o 200 %.

Como alternativa, los aumentos o disminuciones de una afección a la otra afección 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.

A menos que se defina de otro modo, todos los términos técnicos y científicos usados en el presente documento tienen el mismo significado que el entendido habitualmente por un experto en la materia a la que pertenece esta divulgación.

La presente invención se basa en parte en varios descubrimientos analizados con más detalle a continuación. El primer descubrimiento es que un esquema de dos categorías (riesgo alto, riesgo bajo) se puede utilizar de manera eficaz en el análisis de BCI para evitar la incertidumbre de la clasificación de riesgo intermedio de la técnica anterior. Véase, p. ej., el Ejemplo 2 a continuación. El segundo descubrimiento es que un modelo lineal BCI (BCI-L) tiene una capacidad de pronóstico superior para el riesgo de reaparición que un modelo cúbico (BCI-C). Estos descubrimientos permiten una aplicación más simple y precisa de BCI para proporcionar información de pronóstico, tal como la reaparición de cáncer e información predictiva, tal como la capacidad de respuesta a determinadas terapias, que se puede utilizar para la selección de opciones terapéuticas. Como tal, BCI podría identificar un grupo de riesgo bajo que tiene un riesgo de reaparición de menos del 5 % que abarca más del 60 % de los pacientes con cáncer de mama (Ejemplo 2). Esto

hace que más del 60% de los pacientes que se someten a la terapia inicial renuncien a la terapia prolongada con poco riesgo de reaparición.

Estas ventajas se proporcionan adjuntas como métodos para determinar el pronóstico y/o la reaparición de cáncer mediante el ensayo de los patrones de expresión divulgados en el presente documento. Entonces, cuando la interpretación subjetiva se haya utilizado previamente para determinar el pronóstico y/o el tratamiento de los pacientes con cáncer, esta divulgación proporciona patrones objetivos de expresión génica, que se pueden utilizar solos o en combinación con criterios subjetivos para proporcionar una evaluación más precisa de los resultados del paciente, incluida la supervivencia y la reaparición de cáncer.

Por tanto, en algunas realizaciones, se proporciona un método para determinar el riesgo de reaparición de cáncer en un sujeto afectado con cáncer de mama. El método comprende (a) determinar los niveles de expresión del ARNm de una pluralidad de genes en una muestra de células de cáncer de mama ER+ del sujeto; y (b) clasificar si el sujeto tiene un riesgo bajo o alto de reaparición de cáncer basándose en el análisis de los niveles de expresión del ARNm de la pluralidad de genes en el momento del diagnóstico de la enfermedad del cáncer de mama. En este método, el análisis de la pluralidad de genes proporciona un riesgo de reaparición de cáncer después de recibir aproximadamente cinco años de terapia adyuvante que es menos de aproximadamente el 5 % en el grupo de riesgo bajo en comparación con conjuntos de datos retrospectivos de pacientes con cáncer de mama ER+ con más de cinco años de resultado, o muestras representativas de los mismos. En muchas realizaciones, el sujeto ha recibido aproximadamente cinco años de terapia adyuvante y está sin enfermedad después de esa terapia.

Estos métodos se pueden practicar utilizando solo una clasificación de riesgo bajo y alto. En realizaciones alternativas, también se clasifica una categoría de riesgo intermedio.

En algunas realizaciones, el grupo de riesgo bajo comprende más del 50 % del conjunto de datos. En otras realizaciones, el grupo de riesgo bajo comprende más del 55 % del conjunto de datos. En realizaciones adicionales, el grupo de riesgo bajo comprende más del 60 % del conjunto de datos.

Estos métodos pueden evaluar el riesgo de reaparición en cualquier período de tiempo, p. ej., 5 años o menos, mayor de 5 años, a los 10 años, mayor de 10 años, etc. Adicionalmente, estos métodos pueden evaluar el riesgo de reaparición distante o reaparición local, o ambas.

Puede usarse cualquier pluralidad de genes para estos métodos, siempre que, cuando se analizan sus niveles de expresión, son capaces de identificar un grupo de riesgo bajo que tiene menos del 5 % de riesgo de reaparición. En algunas realizaciones, uno de la pluralidad de genes es *HoxB13*. En otras realizaciones, uno de la pluralidad de genes es *IL17BR*. En realizaciones adicionales, se determina la proporción de niveles de expresión de *HoxB13/IL17BR* (H:I).

En realizaciones adicionales, la pluralidad de genes comprende 1, 2, 3, 4 o los 5 genes de MGI, *Bub1B*, *CENPA*, *NEK2*, *RACGAP1*, y *RRM2*. En algunas realizaciones, se determinan los niveles de expresión de *Bub1B*, *CENPA*, *NEK2*, *RACGAP1*, y *RRM2*. En algunas de esas realizaciones, también se determina la relación de niveles de expresión de *HoxB13/IL17BR* (H:I). En varios aspectos de esas realizaciones, los niveles de expresión de *Bub1B*, *CENPA*, *NEK2*, *RACGAP1*, y *RRM2* se suman y se aplica un coeficiente para obtener un índice MGI, y H:I y MGI se combinan como variables continuas en un valor de BCI.

Cuando se usa BCI, el BCI puede calcularse de cualquier forma conocida en la técnica. En algunas realizaciones, el BCI se calcula evaluando el riesgo individual de reaparición de cáncer como parte de una variable BCI continua, donde el riesgo de reaparición aumenta en una relación lineal con la variable BCI.

La presente divulgación también está dirigida a un método para determinar el riesgo de reaparición de cáncer en un sujeto afectado con cáncer de mama. El método comprende determinar los niveles de expresión del ARNm de *HoxB13*, *IL17BR*, *Bub1B*, *CENPA*, *NEK2*, *RACGAP1* y *RRM2* en una muestra de células de cáncer de mama ER+ del sujeto; sumar los niveles de expresión para formar un valor del Índice de Cáncer de Mama (BCI) donde un valor de BCI más alto se correlaciona con un riesgo más alto de reaparición de cáncer y un valor de BCI más bajo se correlaciona con un riesgo más bajo de reaparición de cáncer; y clasificar la muestra, basándose en el valor de BCI, como indicativa de un riesgo bajo o alto de reaparición de cáncer en el sujeto, sin categoría de riesgo intermedio.

Algunos de los métodos de la divulgación se basan en los niveles de expresión de determinados genes, incluyendo el nivel de expresión de *HoxB13*, en células de cáncer de mama de un sujeto como componente del BCI. En algunas realizaciones, se usa una proporción de dos genes de la expresión de *HoxB13* a la expresión de *IL17BR* (o relación *HoxB13: IL17BR*) (Publicaciones de Solicitud de Patente de EE.UU. 2005/0239079, 2005/0239083 y 2006/0154267). En realizaciones alternativas de un índice de cáncer de mama, se puede usar una relación de dos genes de la expresión de *HoxB13* a la expresión de *CHDH*.

La relación *HoxB13: IL17BR* (H:I) se descubrió basándose un estudio de nuevos biomarcadores predictivos del

resultado clínico más allá de los factores pronósticos convencionales. Los pacientes que desarrollaron reapariciones de cáncer se hicieron coincidir con los que no lo hicieron con respecto al estadio y grado del tumor. Se descubrió que la relación H:I simple es adecuada para predecir la reaparición de cáncer en pacientes con cáncer de mama receptor de estrógeno positivo (ER+) que reciben terapia adyuvante con tamoxifeno. Estudios posteriores (Ma et al., 2006; Goetz et al., 2006; Jerevall et al., 2007; Jansen et al., 2007) han demostrado además que la relación es tanto pronóstica, al ser un indicador de la agresividad del tumor como predictiva del beneficio del tamoxifeno (es decir, respuesta/resistencia al tamoxifeno) en ensayos clínicos tanto retrospectivos como aleatorizados.

El BCI incluye la expresión de uno o más genes adicionales en combinación con la expresión de *HoxB13/IL17BR*. La combinación puede ser con cualquiera de uno, dos, tres, cuatro o los cinco genes divulgados adicionalmente como sigue. Estos genes adicionales de la divulgación codifican *Bub1B* ("gemación no inhibida por benzimidazoles I beta) o cinasa 6 activada por proteína p21 (PAK6); *CENPA* (proteína A del centrómero, isoforma a); *NEK2* (cinasa 2 relacionada con NIMA o cinasa 2 relacionada con "nunca en el gen a de la mitosis"); *RACGAP1* (Proteína activadora 1 de Rac GTPasa); y *RRM2* (ribonucleótido reductasa M2). El uso de estos cinco genes solos se denomina en el presente documento Índice de Grado Molecular (MGI). Se analizan los métodos para calcular el MGI, p. ej., en el Ejemplo 2 a continuación y en la Publicación de Patente de EE.UU. 2011/0136680. En algunas realizaciones, el MGI se calcula sumando los niveles de expresión de *Bub1B*, *CENPA*, *NEK2*, *RACGAP1*, y *RRM2* utilizando coeficientes para el nivel de expresión de cada gen. Los coeficientes se pueden determinar mediante cualquier método conocido en la técnica. En diversas realizaciones, los coeficientes se determinan a partir del análisis de componentes principales.

Los aspectos de la divulgación incluyen composiciones y métodos descritos para el uso de la expresión de *HoxB13*, con la expresión de *IL17BR*, en combinación con el o los niveles de expresión de uno o más de los cinco genes anteriores para estudiar, para proporcionar información de pronóstico y/o proporcionar predicciones de la capacidad de respuesta clínica.

Los 5 genes de MGI tienen papeles en el ciclo celular e indicaron la expresión máxima como sigue:

Gen	Pico de expresión	Papel en el Ciclo Celular
<i>Bub1B</i>	G2/M	punto de control del ensamblaje del husillo mitótico
<i>CENPA</i>	G2/M	ensamblaje del centrómero
<i>NEK2</i>	G2/M	duplicación del centrómero
<i>RACGAP1</i>	No Determinado	Inicio de la citocinesis
<i>RRM2</i>	S	Replicación de ADN

Véase la solicitud de patente PCT WO/2012/079059 para obtener detalles sobre la identidad de estos genes.

Por tanto, la divulgación se basa en parte en el descubrimiento de que los niveles de expresión génica son útiles para proporcionar determinaciones pronósticas y predictivas para un sujeto. El uso de los siete genes divulgados se denomina Índice de Cáncer de Mama (BCI).

Como se demuestra en los Ejemplos, BCI proporciona una estratificación superior del riesgo de reaparición en pacientes con cáncer de mama al asignar sujetos con riesgo intermedio a riesgo bajo o riesgo alto, durante un período inicial de hasta cinco años de terapia endocrina, terapia dirigida o tratamiento con un inhibidor de la aromatasa. Por tanto, el BCI es ventajoso sobre otras firmas de expresión génica contemporáneas porque la identificación de dos en lugar de tres grupos de riesgo distintos en cada período de tiempo (agrupando el riesgo intermedio y bajo juntos para determinar la reaparición temprana y el riesgo intermedio y alto juntos para la reaparición tardía) permite la eliminación de la categoría de riesgo intermedio que puede representar hasta el 40 % de los pacientes con cáncer de mama receptor de estrógeno positivo. En algunos casos, El BCI se aplica en el contexto de la reaparición tardía de la enfermedad porque permite un medio de identificar a los pacientes que pueden evitar la terapia endocrina adyuvante prolongada y sus efectos secundarios adversos.

Los factores clínico-patológicos como el estado ganglionar y el tamaño del tumor se asocian con un mayor riesgo de reaparición tardía; sin embargo, los resultados divulgados en el presente documento representan un refinamiento, permitiendo una evaluación individualizada del riesgo de reaparición tardía y proporcionando una mejora estadísticamente significativa en el rendimiento pronóstico más allá de los factores clínico patológicos. Dicho de otra manera, los métodos de la divulgación se pueden practicar sin el uso de, u opcionalmente junto con, el uso de factores clínico patológicos tales como el estado ganglionar y el tamaño del tumor. Otros ensayos basados en la expresión génica (EndoPredict y PAM 50) tienen capacidad de pronóstico de reaparición tardía más allá de los factores clínico patológicos. Estos estudios validan aún más el uso clínico de ensayos de base molecular para la evaluación del riesgo de reaparición tardía de la enfermedad.

Para determinar los niveles de expresión de genes en la práctica de la presente divulgación, se puede utilizar cualquier

método conocido en la técnica. En algunas realizaciones, se usa la expresión basada en la detección de ARNm que hibrida con los genes identificados y divulgados en el presente documento. Esto se lleva a cabo fácilmente según cualquier método de detección o amplificación+detección de ARNm conocido o reconocido como equivalente en la técnica tales como, pero sin limitación, PCR con transcripción inversa, los métodos descritos en la Patente de EE.UU. 6.794.141, y los métodos para detectar la presencia o ausencia, de secuencias estabilizantes o desestabilizantes del ARN. En diversas realizaciones, el ARNm se convierte en ADNc.

La capacidad para discriminar se confiere mediante la identificación de la expresión de los genes individuales como relevante y no por la forma del ensayo utilizado para determinar el nivel real de expresión. Un ensayo puede utilizar cualquier característica de identificación de un gen individual identificado como se divulga en el presente documento, siempre que el ensayo refleje, cuantitativa o cualitativamente, 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 de identificación incluyen, pero sin limitación, secuencias de los ácidos nucleicos exclusivas utilizadas para codificar (ADN) o expresar (ARN), dicho gen o epítomos específicos o actividades de, una proteína codificada por dicho gen. Todo lo que se necesita es la identidad del gen o los genes necesarios para discriminar entre los resultados de cáncer y una muestra apropiada que contiene células para su uso en un ensayo de expresión. De forma similar, la naturaleza de la muestra que contiene células no es limitante, ya que en los métodos divulgados puede usarse tejido fresco, tejido recién congelado y tejido fijado, tal como tejidos embebidos en parafina fijados con formalina (FFPE).

También se puede utilizar la expresión basada en la detección de presencia, aumento o la disminución en los niveles o actividad de las proteínas. La detección puede realizarse mediante cualquier método basado en (IHC), basado en sangre (especialmente para proteínas secretadas), basado en anticuerpos (incluidos los autoanticuerpos contra la proteína), basado en células exfoliadas (del cáncer), basado en espectroscopía de masas y basado en imagen (incluido el uso de ligando marcado) conocido en la técnica y reconocido como apropiado para la detección de la proteína. Los métodos basados en anticuerpos e imágenes también son útiles para la localización de tumores después de la determinación del cáncer mediante el uso de células obtenidas mediante un procedimiento no invasivo (como el lavado ductal o la aspiración con aguja fina), donde se desconoce la fuente de las células cancerosas. Puede usarse un anticuerpo o ligando marcado para localizar el o los carcinomas dentro de un paciente.

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

El gen o genes inmovilizados pueden estar en forma de polinucleótidos que son exclusivos o específicos de otro modo para el gen o genes, de modo que el polinucleótido sería capaz 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 corto que la secuencia de longitud completa conocida en la técnica, p. ej., por eliminación del extremo 5' o 3' de la secuencia) que están opcionalmente mínimamente interrumpidos (como por falta de coincidencias o pares de bases no complementarias insertados) de manera que no se ve afectada la hibridación con un ADN o ARN correspondiente al gen o genes. En algunos casos, los polinucleótidos utilizados son del extremo 3' del gen, como dentro de aproximadamente 350, aproximadamente 300, aproximadamente 250, aproximadamente 200, aproximadamente 150, aproximadamente 100, o aproximadamente 50 nucleótidos a partir de la señal de poliadenilación o sitio de poliadenilación de un gen o secuencia expresada.

El experto en la materia es completamente capaz de alinear dos cualquiera o más de las secuencias expresadas conocidas para cada uno de estos genes para identificar un área de identidad o cambios conservados como una región que identifica de forma exclusiva cada uno de estos genes en comparación con otros genes. Asimismo, el experto es completamente capaz de alinear dos cualquiera o más de las secuencias expresadas conocidas para cada uno de estos genes para identificar un área exclusiva de una o más de las secuencias expresadas como una región que identifica de manera exclusiva una secuencia expresada conocida en relación con al menos otra secuencia expresada. Como un ejemplo no limitante, una región exclusiva puede estar en una variante de la secuencia expresada para uno de los genes conocidos, de modo que la región puede usarse para identificar la expresión de la variante.

Las secuencias de los mismos genes también se han identificado y caracterizado de otras especies animales. Por tanto, el experto en el campo sabe claramente cómo identificar los genes divulgados en relación con otros genes animales. El experto también puede comparar opcionalmente las secuencias conocidas de los genes divulgados de diferentes fuentes animales para identificar regiones conservadas y secuencias exclusivas de estos genes en relación con otros genes.

Los polinucleótidos que contienen mutaciones con respecto a las secuencias de los genes divulgados también pueden usarse siempre que la presencia de las mutaciones aún permita la hibridación para producir una señal detectable. El o los genes inmovilizados pueden usarse para determinar el estado de las muestras de ácido nucleico preparadas a partir de una muestra de células de cáncer o mama, para las que no se conoce el resultado del sujeto de la muestra (por ejemplo, el paciente del que se obtiene la muestra) o para la confirmación de un resultado que ya está asignado

al sujeto de la muestra. Sin limitar la divulgación, dicha célula puede ser de un paciente con cáncer de mama ER+. El polinucleótido o polinucleótidos inmovilizados solo necesitan ser suficientes para hibridar específicamente con las moléculas del ácido nucleico correspondientes derivadas de la muestra en condiciones adecuadas.

- 5 Como apreciarán los expertos en la materia, algunas de las secuencias correspondientes indicadas anteriormente incluyen tramos 3' poliA (o poliT en la hebra complementaria) que no contribuyen a la exclusividad de las secuencias divulgadas. Por tanto, la divulgación se puede poner en práctica con secuencias que carecen de los tramos 3' poliA (o poliT). La exclusividad de las secuencias divulgadas se refiere a las porciones o la totalidad de las secuencias que se encuentran solo en los ácidos nucleicos de los genes divulgados, incluyendo secuencias exclusivas encontradas en la porción 3' no traducida de los genes. Las secuencias exclusivas preferidas para la práctica de la divulgación son aquellas que contribuyen a las secuencias consenso para cada uno de los tres conjuntos, de modo que las secuencias exclusivas serán útiles para detectar la expresión en una variedad de individuos en lugar de ser específicas para un polimorfismo presente en algunos individuos. Como alternativa, pueden usarse secuencias exclusivas para un individuo o una subpoblación. Las secuencias exclusivas preferidas son preferentemente de las longitudes de polinucleótidos de la divulgación como se analiza en el presente documento.

También se pueden usar métodos para identificar una estabilidad aumentada del ARN (que da como resultado una observación de expresión aumentada) o una estabilidad disminuida del ARN (que da como resultado una observación de la 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 un aumento de la degradación del ARNm. En algunas realizaciones de la divulgación, los polinucleótidos que tienen secuencias presentes en las regiones 3' no traducidas y/o no codificantes de las secuencias divulgadas anteriormente se usan para detectar niveles de expresión de las secuencias génicas en células de cáncer o mama. Dichos polinucleótidos pueden contener opcionalmente secuencias encontradas en las porciones 3' de las regiones codificantes de las secuencias divulgadas anteriormente.

Los polinucleótidos que contienen una combinación de secuencias de las regiones 3' codificantes y no codificantes tienen preferentemente las secuencias dispuestas de manera contigua, sin secuencias heterólogas intermedias.

Como alternativa, la divulgación se puede practicar con polinucleótidos que tienen secuencias presentes en las regiones 5' no traducidas y/o no codificantes de las secuencias génicas en células de cáncer, o mama, para detectar sus niveles de expresión. Dichos polinucleótidos pueden contener opcionalmente secuencias encontradas en las porciones 5' de las regiones codificantes. Los polinucleótidos que contienen una combinación de secuencias de las regiones 5' codificantes y no codificantes tienen preferentemente las secuencias dispuestas de manera contigua, sin secuencias heterólogas intermedias. La divulgación también se puede practicar con secuencias presentes en las regiones codificantes de las secuencias génicas divulgadas.

Los polinucleótidos no limitantes contienen secuencias de regiones 3' o 5' no traducidas y/o no codificantes de al menos aproximadamente 20, al menos aproximadamente 22, 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" como se usa en la oración previa se refiere a un aumento o disminución de 1 del valor numérico establecido. Incluso son más preferidos 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" como se usa en la oración anterior se refiere a un aumento o disminución del 10 % del valor numérico establecido.

Las secuencias del extremo 3' o 5' de las regiones codificantes descritas anteriormente que se encuentran en los polinucleótidos de la divulgación tienen las mismas longitudes que las descritas anteriormente, excepto que, naturalmente, estarían 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. En cambio, el extremo 5' de una región codificante puede incluir secuencias en la mitad 5' de la región codificante. Por supuesto, las secuencias descritas anteriormente, o las regiones codificantes y polinucleótidos que contienen porciones de las mismas, se puede utilizar en su totalidad.

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 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. Preferentemente, los polinucleótidos utilizados son del extremo 3' del gen, como dentro de aproximadamente 350, aproximadamente 300, aproximadamente 250, aproximadamente 200, aproximadamente 150, aproximadamente 100, o aproximadamente 50 nucleótidos a partir de la señal de poliadenilación o sitio de poliadenilación de un gen o secuencia expresada. Los polinucleótidos que contienen mutaciones con respecto a las secuencias de los genes divulgados también pueden

usarse siempre que la presencia de las mutaciones aún permita la hibridación para producir una señal detectable.

En otra realización de la divulgación, se pueden usar polinucleótidos que contienen eliminaciones de nucleótidos del extremo 5' y/o 3' de las secuencias divulgadas anteriormente. Las eliminaciones 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 5' y/o 3', aunque la extensión de las eliminaciones estaría naturalmente limitada por la longitud de las secuencias y la necesidad de poder usar los polinucleótidos para la detección de niveles de expresión.

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

En otras realizaciones más de la divulgación, se pueden usar polinucleótidos que contienen porciones de las secuencias divulgadas anteriormente que incluyen el extremo 3'. Dichos 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.

La divulgación también incluye polinucleótidos usados para detectar la expresión génica en células de cáncer de mama. Los polinucleótidos pueden comprender un polinucleótido más corto que consta de secuencias encontradas en los genes anteriores en combinación con secuencias heterólogas que no se encuentran de manera natural en combinación con las secuencias. Ejemplos no limitantes incluyen secuencias cortas de vectores de clonación o presentes en fragmentos de restricción usadas para preparar sondas o cebadores marcados como se describe en el presente documento.

El nivel de expresión necesario puede ser el que se identifica mediante los métodos descritos en el presente documento para los genes usados. Adicionalmente, el ensayo puede incluir preparar ARN a partir de la muestra, opcionalmente para uso en PCR (reacción en cadena de la polimerasa) u otra metodología analítica como se describe en el presente documento. La metodología de PCR es opcionalmente RT-PCR (transcripción inversa-PCR) o PCR cuantitativa, tal como la RT-PCR en tiempo real. Como alternativa, el ensayo puede realizarse mediante el uso de una matriz, tal como una micromatriz, por secuenciación de nueva generación o por cualquier otro método conocido en la técnica. Opcionalmente, la muestra de células cancerosas se disecciona del tejido extraído u obtenido de dicho sujeto. Como se describe en el presente documento, se puede utilizar una variedad de tipos de muestras, incluyendo una muestra fijada en formalina y embebida en parafina (FFPE, por sus siglas en inglés) como ejemplo no limitante. Y como se describe en el presente documento, el método puede incluir analizar o determinar la relación H:I (relación de los niveles de expresión de *HoxB13* y *IL17BR*) en la muestra como se divulga en el presente documento.

A modo de ejemplo no limitante, los cinco genes del MGI pueden analizarse y utilizarse para detectar niveles de expresión que correspondan a un valor de "riesgo alto" (que esté por encima del punto de corte) para MGI, o para detectar niveles de expresión que correspondan a un valor "riesgo bajo" (que es igual o inferior al punto de corte) de MGI, como se divulga en el presente documento. En algunos casos, el umbral de punto de corte de MGI puede ser 0 (cero), como cuando las medidas de los niveles de expresión están normalizadas a 0 (cero) con una desviación estándar de 1. En realizaciones alternativas, el punto de corte puede ser de o aproximadamente 0,05, de o aproximadamente 0,10, de o aproximadamente 0,15, de o aproximadamente 0,20, de o aproximadamente 0,25, de o aproximadamente -0,05, de o aproximadamente -0,10, de o aproximadamente -0,15, de o aproximadamente -0,20, de o aproximadamente -0,25, de o aproximadamente -0,30, de o aproximadamente -0,35, de o aproximadamente -0,40, de o aproximadamente -0,45, de o aproximadamente -0,50, de o aproximadamente -0,55, de o aproximadamente -0,60, de o aproximadamente -0,65, de o aproximadamente -0,70, de o aproximadamente -0,75, de o aproximadamente -0,80, de o aproximadamente -0,85, de o aproximadamente -0,90, de o aproximadamente -0,95, de o aproximadamente -1,0, de o aproximadamente -1,1, de o aproximadamente -1,2, de o aproximadamente -1,3, de o aproximadamente -1,4, de o aproximadamente -1,5, de o aproximadamente -1,6, de o aproximadamente -1,7, de o aproximadamente -1,8, de o aproximadamente -1,9, de o aproximadamente -2,0 o menor. Con respecto a la relación H:I, su determinación puede realizarse como se describe en Ma et al., 2004 y Ma et al., 2006. Por ejemplo, se puede utilizar un valor de 0,06 para determinar si una muestra tiene una relación H:I de "riesgo alto" (> 0,06) o de "riesgo bajo" (0,06).

Entonces, usando un umbral o punto de corte, de 0 (cero) como un ejemplo no limitante para MGI con los cinco genes, los métodos divulgados proporcionan dos posibles resultados de ensayo para una muestra determinada: "MGI de riesgo alto" correspondiente a un valor superior a 0 (cero) y "MGI de riesgo bajo" correspondiente a un valor de 0. Un "MGI de riesgo alto" es indicativo de un cáncer de "riesgo alto", incluido el cáncer de mama que es análogo al de un tumor de Grado III según lo definido por las metodologías y normas conocidas en el campo. Un "MGI de riesgo bajo"

es indicativo de un cáncer de "riesgo bajo", incluyendo el cáncer de mama, que es análogo al de un tumor de Grado I según lo definido por las metodologías y normas conocidas en el campo.

El umbral o los puntos de corte utilizados para determinar el riesgo intermedio de reaparición de cáncer pueden ser los divulgados en el presente documento o dentro de aproximadamente el 2 %, 4 %, 6 %, 8 %, 10 %, 12 %, 14 %, 16 %, 18 %, 20 %, 25 %, 30 %, 35 %, 40 %, 50 %, 60 %, 70 %, 80 %, 90 %, 100 % o más de los mismos.

Como un ejemplo no limitante, la célula cancerosa puede ser una de una muestra histológica preoperatoria o una biopsia, utilizado para diagnosticar el cáncer en el sujeto. Para un sujeto con carcinoma ductal in situ (DCIS, por sus siglas en inglés), el tratamiento de referencia actual es la cirugía, con cirugía conservadora de mama preferida a una mastectomía radical, para quitar el DCIS. Esto suele ir seguido de radioterapia posoperatoria, opcionalmente con terapia endocrina, tal como el tratamiento con tamoxifeno, un modulador selectivo del receptor de estrógeno (SERM, por sus siglas en inglés), un regulador negativo selectivo del receptor de estrógeno (SERD, por sus siglas en inglés), un inhibidor de la aromatasa (AI, por sus siglas en inglés) tal como letrozol, una terapia dirigida tal como la terapia anti-mTOR (p. ej., con Afinitor®) o la terapia anti-HER2 (p. ej., con Herceptin®) y/o quimioterapia, utilizando cualquier compuesto conocido en la técnica.

La detección de la expresión génica y la determinación de BCI pueden, por supuesto, realizarse en cualquier muestra que contenga células adecuada como se describe en el presente documento. Los ejemplos no limitantes de células para su uso en la divulgación incluyen aquellas recién aisladas del sujeto, las congeladas después del aislamiento, y los que están fijadas y/o embebidas, tal como fijadas en formalina, embebidas en parafina (FFPE). En la mayoría de las realizaciones, las células son células mamarias, tal como células de cáncer de mama.

Como se divulga en el presente documento, el BCI se utiliza para determinar el riesgo de reaparición de cáncer en una paciente afectado con cáncer de mama. Los ejemplos no limitantes de reaparición tardía incluyen después de 5 años de tratamiento con un inhibidor de la aromatasa, terapia dirigida o terapia endocrina, tales como el tamoxifeno, pero también incluye después de 4 años, después de 3 años, o después de 2 años o menos de tiempo de tratamiento. De forma similar, el BCI puede usarse para predecir la capacidad de respuesta a una terapia anti aromatasa, tal como anastrozol o letrozol, terapia dirigida o terapia antiestrógeno después de los períodos de tiempo anteriores.

En algunas realizaciones, los métodos divulgados en el presente documento pueden usarse ventajosamente en una muestra que contiene células de cáncer de mama de un sujeto, tal como una muestra de DCIS, aunque los métodos descritos en el presente documento pueden usarse con cualquier tipo de cáncer de mama, incluyendo cualquier cáncer de mama invasivo o no invasivo, tal como el carcinoma ductal invasivo, carcinoma lobulillar invasivo, cáncer de mama inflamatorio, cáncer de mama masculino, cáncer de mama metastásico, cáncer de mama recurrente, carcinoma papilar, cáncer de mama triple negativo, enfermedad de Paget del pezón, sarcoma de mama, carcinoma medular, carcinoma tubular, carcinoma mucinoso, carcinoma metaplásico, carcinoma adenocístico, tumor filodes y angiosarcoma.

Como se analiza en el Ejemplo 2 a continuación, el riesgo de reaparición usando BCI se puede clasificar como de riesgo bajo y de riesgo alto, sin una categoría de riesgo intermedia, sin pérdida de precisión. En este esquema, la clasificación intermedia, como se describe, p. ej., en la Publicación de Patente de EE. UU. 2013/0281502 (véase, p. ej., la Tabla 2 de la misma), se puede agrupar con el grupo de riesgo bajo cuando se clasifica el riesgo de reaparición a los 5 años o menos (por ejemplo, 4, 3, 2 o 1 año) y se puede agrupar con el grupo de riesgo alto cuando se clasifica el riesgo de reaparición a más de 5 años (por ejemplo, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 o más años).

Cuando BCI se escala a un intervalo de 1-10 (véase, p. ej., las FIG. 3, 6 y 8), el grupo intermedio se encuentra entre puntuaciones de BCI de aproximadamente 5-6,5. Como se analiza anteriormente, cuando el BCI se escala a un intervalo de 1-10 y se clasifica el riesgo de reaparición a los 5 años o menos, el grupo intermedio se puede unir con el grupo de riesgo bajo. El valor de punto de corte entre los grupos de riesgo bajo (+ intermedio) y de riesgo alto es, por tanto, donde el grupo intermedio se encuentra con el grupo de riesgo alto, es decir, aproximadamente 6,5, por ejemplo 5,5, 5,6, 5,7, 5,8, 5,9, 6,0, 6,1, 6,2, 6,3, 6,4, 6,6, 6,7, 6,8, 6,9, 7,0, 7,1, 7,2, 7,33, 7,4 o 7,5, o cualquier valor intermedio. Se entiende que el valor de punto de corte en estas circunstancias puede estar entre 5,5 y 7,5, dependiendo de cuántos años después se determine el riesgo de reaparición (menos años permite un punto de corte más alto) y cuán conservador desea ser el operador al declarar que un paciente tiene un riesgo bajo o alto de reaparición (es decir, el porcentaje de reaparición bajo el cual el índice tiene un riesgo bajo de reaparición). El experto en la materia podría determinar una puntuación de BCI adecuada, sin excesiva experimentación, para cualquier número particular de años y nivel de riesgo deseado.

De forma similar, cuando el BCI se escala a un intervalo de 1-10 y se clasifica el riesgo de reaparición a más de 5 años, el grupo intermedio se puede unir con el grupo de riesgo alto, haciendo que el punto de corte entre el grupo de riesgo bajo y el grupo alto (+ intermedio) sea el punto en el que el grupo intermedio se encuentra con el grupo de riesgo bajo, es decir, aproximadamente 5, por ejemplo 4,0, 4,1, 4,2, 4,3, 4,4, 4,5, 4,6, 4,7, 4,8, 4,9, 5,1, 5,2, 5,3, 5,4, 5,5, 5,6, 5,7, 5,8, 5,9 o 6,0, o cualquier valor intermedio.

Los escenarios específicos anteriores ilustran que, para pacientes de riesgo bajo después del tratamiento anticipado

(inicial) con tamoxifeno, terapia dirigida o inhibidor de la aromatasa después del diagnóstico de cáncer de mama y/o intervención quirúrgica, las pruebas de BCI pueden ofrecer la opción de no recibir más terapia sistémica. Por ejemplo, pacientes con riesgo alto de reaparición después del tamoxifeno adyuvante inicial (pacientes con alto *HOXB13/T117BR*) se benefician de la terapia hormonal prolongada con un cambio al inhibidor de la aromatasa letrozol. Esos pacientes exhibirían una relación H:I o puntuación de BCI alta. Pacientes con riesgo alto de reaparición después de 5 años de un inhibidor de la aromatasa inicial, terapia dirigida o terapia endocrina sin embargo, puede o no beneficiarse de la terapia hormonal adyuvante prolongada o de cualquier terapia sistémica. Esos pacientes también pueden ser candidatos para enfoques terapéuticos experimentales. Sin embargo, la divulgación inmediata proporciona BCI como un medio para ayudar a clasificar a esos pacientes de manera más apropiada.

En algunas realizaciones, el BCI calculado indica una reaparición distante de riesgo bajo en el sujeto si se trata con un inhibidor de la aromatasa, terapia dirigida o terapia endocrina, tales como el tamoxifeno, durante un período inicial de 5 años o menos porque el BCI calculado indica riesgo bajo o riesgo intermedio. Por tanto, el sujeto puede ser tratado durante el período inicial de 5 años con el inhibidor de la aromatasa, terapia dirigida o terapia endocrina con riesgo bajo de reaparición. Después del período inicial, y si el BCI calculado es de riesgo intermedio, la indicación es que el sujeto tiene un riesgo alto de reaparición de cáncer si no se trata con un inhibidor de la aromatasa, terapia dirigida o terapia endocrina durante un período adicional de cinco años.

En otras realizaciones, el BCI calculado indica una reaparición distante de riesgo alto en el sujeto si se trata con un inhibidor de la aromatasa, terapia dirigida o terapia endocrina, tales como el tamoxifeno, durante un período inicial de 5 años o menos porque el BCI calculado es de riesgo alto o riesgo intermedio. Por tanto, el sujeto puede ser tratado con el inhibidor de la aromatasa, terapia dirigida o terapia endocrina con una expectativa de éxito atenuada. En algunos casos, el sujeto puede ser tratado adicionalmente con terapia adicional, tal como quimioterapia o radioterapia como ejemplos no limitantes, durante el período inicial.

La presente divulgación incluye la identificación de un sujeto que se espera que se beneficie de una terapia adicional después de la supervivencia sin reaparición durante el curso de una anti aromatasa inicial, terapia dirigida o terapia endocrina, tal como por un período de cinco años o menos.

Por lo tanto, la divulgación incluye determinar el BCI como un indicador de una mayor probabilidad de reaparición de cáncer en el sujeto después de una terapia inicial anti aromatasa, terapia dirigida o terapia endocrina, tal como la terapia adyuvante con tamoxifeno. Por tanto, los métodos pueden incluir identificar al sujeto como probable o improbable, que experimente reaparición de cáncer local o distante. Como un ejemplo no limitante, la determinación de la probabilidad de reaparición en ausencia de un tratamiento posterior al inicial prolongado, la terapia se puede aplicar durante un período posterior de hasta cinco años o más.

La presente divulgación también proporciona la determinación del riesgo de reaparición de cáncer en un sujeto con cáncer de mama mediante la realización de una evaluación de riesgo individual como parte de, o en relación con, el BCI calculado como una variable continua. Como se divulga en el presente documento, la determinación de BCI en una población de muestras de cáncer de mama indica que BCI es una variable continua que se correlaciona con el riesgo de reaparición de cáncer en sujetos afectados por cáncer de mama. Por tanto, el riesgo de reaparición aumenta en una relación lineal con el aumento de los valores de BCI. En algunas realizaciones, el intervalo de valores de BCI como una variable continua se compara con un valor de BCI determinado para una muestra individual de cáncer de mama para evaluar el riesgo de reaparición de cáncer como de riesgo bajo, riesgo intermedio o riesgo alto.

El período de tratamiento o terapia adicional o posterior, como se divulga en el presente documento, puede producirse en cualquier momento después de la terapia de primera línea (inicial), tal como inmediatamente después, dentro de los tres meses posteriores a la finalización del tratamiento de primera línea, dentro de los seis meses posteriores a la finalización del tratamiento de primera línea, dentro de los nueve meses posteriores a la finalización del tratamiento de primera línea, dentro de los 12 meses posteriores a la finalización del tratamiento de primera línea, dentro de los 18 meses posteriores a la finalización de la terapia de primera línea, o dentro de los 24 meses (o más) posteriores a la finalización de la terapia de primera línea.

La capacidad de pronóstico para identificar un riesgo alto o bajo de reaparición proporciona información que puede ser fundamental para determinar un curso de tratamiento. Por ejemplo, cuando BCI indica que un sujeto (p. ej., un paciente) tiene un riesgo alto de reaparición distante si se trata con un inhibidor de la aromatasa, terapia dirigida o terapia endocrina, tales como el tamoxifeno, durante un período inicial de 5 años o menos, tales terapias estarían contraindicadas, y se pueden instituir otras terapias, tal como la quimioterapia.

En cambio, si BCI indica que el sujeto tiene riesgo bajo de reaparición distante si se trata con un inhibidor de la aromatasa, terapia dirigida o terapia endocrina, tales como el tamoxifeno, durante un período inicial de 5 años o menos, podría estar indicado un tratamiento con un inhibidor de la aromatasa, terapia dirigida o terapia endocrina.

Adicionalmente, si BCI indica que el sujeto tiene un riesgo alto de reaparición distante si no se trata con un inhibidor de la aromatasa, terapia dirigida o terapia endocrina durante un período adicional de 5 años, podría estar indicado un tratamiento con un inhibidor de la aromatasa, terapia dirigida o terapia endocrina y otra terapia adyuvante.

En otras realizaciones, se identifica que el sujeto se ha sometido a tratamiento con un inhibidor de la aromatasa, terapia dirigida o terapia endocrina durante un período de hasta cinco años sin reaparición de cáncer. A continuación, se clasifica el sujeto, basándose en el valor de BCI del tumor del sujeto, como que tiene o que no tiene un riesgo alto de reaparición distante del cáncer después de la finalización del tratamiento. Si el sujeto tiene un riesgo alto de reaparición, el sujeto se trata después durante un período adicional con inhibidor de la aromatasa, terapia dirigida o terapia endocrina.

Tal como se sabe, un alto índice *HOXB13/IL17BR* no solo es pronóstico, sino también predictivo del beneficio del tratamiento endocrino adyuvante. Véase la patente de EE.UU. 7.504.214 y la publicación de patente PCT WO/2012/079059. Como se analiza en el Ejemplo 2, la capacidad de BCI-C para predecir la capacidad de respuesta podría verse confundida por las propiedades predictivas del tratamiento endocrino y de pronóstico dual de *HOXB13/IL17BR*. Por el contrario, BCI-L contiene solo funciones aditivas de MGI y *HOXB13/IL17BR* y se desarrolló en un grupo no tratado de sujetos con cáncer de mama en el que el resultado clínico representaba la historia natural del cáncer de mama. Los hallazgos informados en el Ejemplo 2 sugieren que BCI-L se libró de cualquier efecto de confusión de las propiedades predictivas del tratamiento endocrino de *HOXB13/IL17BR* y, como resultado, BCI-L es una versión de pronóstico preferida de la combinación de *HOXB13/IL17BR* y MGI.

En un segundo aspecto, la divulgación proporciona un método que identifica la capacidad de respuesta al tratamiento, por ejemplo, un tratamiento adicional con un inhibidor de la aromatasa, terapia dirigida o terapia endocrina después de un tratamiento inicial con un inhibidor de la aromatasa, terapia dirigida o terapia endocrina durante cinco años o menos. En algunas realizaciones, el tratamiento adicional es durante un período de cinco años o más después del período de tratamiento inicial.

Por tanto, en realizaciones adicionales, la divulgación está dirigida a un método para determinar la necesidad de un tratamiento prolongado de un sujeto afectado con cáncer de mama. El método comprende determinar los niveles de expresión del ARNm de *HoxB13*, *IL17BR*, *Bub1B*, *CENPA*, *NEK2*, *RACGAP1* y *RRM2* en una muestra de células de cáncer de mama ER+ del sujeto;

sumar los niveles de expresión para formar un valor de Índice de Cáncer de Mama (BCI), donde un valor de BCI más bajo se correlaciona con no necesitar ningún tratamiento adicional más con un inhibidor de la aromatasa, terapia dirigida o terapia endocrina después de un tratamiento inicial con un inhibidor de la aromatasa, terapia dirigida o terapia endocrina durante cinco años o menos; y clasificar la muestra, basándose en el BCI, como indicación de la necesidad o no de un tratamiento adicional.

La capacidad de BCI para discernir el riesgo de reaparición y la respuesta al tratamiento se puede combinar para proporcionar opciones de tratamiento.

Por tanto, la divulgación también está dirigida a un método de tratamiento de un sujeto afectado con cáncer de mama. El método comprende

determinar los niveles de expresión del ARNm de *HoxB13*, *IL17BR*, *Bub1B*, *CENPA*, *NEK2*, *RACGAP1* y *RRM2* en una muestra de células de cáncer de mama ER+ del sujeto;

sumar los niveles de expresión para formar un valor de Índice de Cáncer de Mama (BCI), donde un valor de BCI más bajo se correlaciona con (a) la capacidad de respuesta al tratamiento adicional con un inhibidor de la aromatasa, terapia dirigida o terapia endocrina después de un tratamiento inicial con un inhibidor de la aromatasa, terapia dirigida o terapia endocrina durante cinco años o menos y (b) un menor riesgo de reaparición distante; y tratar al sujeto de manera consistente con la determinación del valor de BCI para el sujeto.

En algunas realizaciones, el análisis de BCI indica un riesgo alto de reaparición distante en el sujeto si se trata con un inhibidor de la aromatasa, terapia dirigida o terapia endocrina durante un período inicial de 5 años o menos. Dicho sujeto podría tratarse con quimioterapia.

En otras realizaciones, el análisis de BCI indica un riesgo bajo de reaparición distante en el sujeto si se trata con un inhibidor de la aromatasa, terapia dirigida o terapia endocrina durante 5 años o menos. Dicho sujeto podría tratarse con el inhibidor de la aromatasa, terapia dirigida o terapia endocrina durante el período inicial de 5 años, entonces no tendrá que tener una terapia prolongada.

En realizaciones adicionales, el análisis de BCI indica un riesgo alto de reaparición distante en el sujeto si no se trata con un inhibidor de la aromatasa, terapia dirigida o terapia endocrina durante un período adicional de 5 años. Dicho sujeto podría tratarse con un inhibidor de la aromatasa, terapia dirigida o terapia endocrina u otra terapia adyuvante. Cualquiera de los métodos anteriores puede ser útil para determinar un factor pronóstico o predictor de la capacidad de respuesta clínica en mujeres premenopáusicas y postmenopáusicas. Las mujeres postmenopáusicas se pueden definir como aquellas que tienen > 50 años, mientras que las mujeres premenopáusicas se pueden definir como aquellas que tienen menos de 50 años. En algunos aspectos, estas mujeres se han sometido a un tratamiento con anti aromatasa, terapia dirigida o terapia endocrina y permanecieron sin cáncer durante ese tiempo.

En una realización adicional, la divulgación proporciona la identificación de los patrones de expresión génica mediante

el análisis de la expresión génica global, o casi global, de células sueltas o poblaciones de células homogéneas que se han diseccionado o aislado o purificado de otro modo de, células contaminantes más allá de lo posible mediante una simple biopsia. Debido a que la expresión de numerosos genes fluctúa entre células de diferentes pacientes, así como entre células de la misma muestra de paciente, los niveles de expresión génica pueden determinarse en correspondencia con uno o más genes de "control" o "normalización", cuya o cuyas expresiones son relativamente constantes en las células de un paciente o entre pacientes.

Una ventaja de este enfoque es que las células contaminantes, no cancerosas (tal como los linfocitos infiltrantes u otras células del sistema inmunitario) no están presentes para afectar posiblemente los genes identificados o el análisis posterior de la expresión génica para identificar la reaparición de cáncer y/o los resultados de supervivencia de los pacientes. Dicha contaminación está presente cuando se usa una biopsia que contiene muchos tipos de células para analizar los perfiles de expresión génica.

Si bien la presente divulgación se describe principalmente en el contexto del cáncer humano, tales como cáncer de mama, se puede practicar en el contexto del cáncer de cualquier animal. Los animales preferidos para la aplicación de la presente divulgación son mamíferos, particularmente aquellos importantes para aplicaciones agrícolas (tales como, pero sin limitación, ganado bovino, oveja, caballos y otros "animales de granja"), modelos animales de cáncer, y animales para de compañía para los seres humanos (tal como, pero sin limitación, perros y gatos).

Los métodos proporcionados por la divulgación también pueden automatizarse en su totalidad o en parte. Kits

Los materiales para usar en los métodos de la presente divulgación son idóneos para la preparación de kits producidos de acuerdo con 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 de cáncer. Dichos kits comprenden opcionalmente el agente con una descripción o etiqueta de identificación o instrucciones relacionadas con su uso en los métodos de la presente divulgación. Dicho kit puede comprender envases, cada uno con uno o más de los diversos reactivos (generalmente en forma concentrada) utilizados en los métodos, incluyendo, por ejemplo, micromatrices prefabricadas, tampones, los trifosfatos de nucleótidos apropiados (p. ej., 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 (p. ej., cebadores de poli(T) o aleatorios de longitud apropiada unidos a un promotor reactivo con la ARN polimerasa). Por lo general, también se incluye un conjunto de instrucciones.

Las realizaciones preferidas se describen en los siguientes ejemplos. Otras realizaciones dentro del alcance de las reivindicaciones en el presente documento serán evidentes para un experto en la materia a partir de la consideración de la memoria descriptiva o la práctica de la invención como se divulga en el presente documento.

Ejemplo 1

Diseño del estudio y pacientes

Para un estudio de comparación prospectivo, se obtuvieron muestras de tejido del proyecto TransATAC, iniciado en 2002 para establecer un banco de tejidos de bloques de tumores primarios fijados en formalina y embebidos en parafina (FFPE) de pacientes postmenopáusicas con cáncer de mama receptor de estrógeno positivo de los grupos de monoterapia del ensayo ATAC para ayudar con la investigación traslacional (Paik et al., 2004; Dowsett et al., 2010). Se solicitaron bloques tumorales de archivo para todas las pacientes para las que ya se había calculado la puntuación de reaparición de 21 genes y la IHC4, excepto los que se sabía que eran receptor de estrógeno y receptor de progesterona negativos de acuerdo con las pruebas locales y las asignados al azar al grupo de tratamiento de combinación del ensayo ATAC. El estudio fue aprobado por el South-East London Research Ethics Committee y la Massachusetts General Hospital Institutional Review Board. Las pacientes habían dado su consentimiento por escrito para que su tejido se utilizara en ensayos adicionales.

Procedimientos

Anteriormente, se realizó un estudio en el que se extrajo ARN de bloques FFPE del banco de tejidos TransATAC de pacientes del Reino Unido (cuyas muestras constituían el 79 % de la colección) para calcular y probar la puntuación de reaparición de 21 genes (Paik et al., 2004). Posteriormente, se realizaron análisis inmunohistoquímico para receptores de estrógenos, receptores de progesterona, HER2, Ki-67 y evaluaciones del grado del tumor, y se calcularon IHC4 y la puntuación del tratamiento clínico (un modelo de pronóstico que utiliza las variables clásicas de tamaño y grado del tumor, estado de los ganglios linfáticos, edad y tratamiento) utilizando muestras de tejido de los mismos pacientes cuyo tejido se utilizó para calcular la puntuación de reaparición de 21 genes (para los que se disponía de tejido adicional suficiente).

En este estudio, se utilizaron las mismas muestras coincidentes que se utilizaron en los estudios anteriores con suficiente ARN residual para realizar el análisis de BCI. Los genes se probaron, se analizaron las secuencias de cebadores y sondas, y se realizaron procedimientos de RT-PCR para calcular *HOXB13/IL17BR* y *MGI* como se informó anteriormente (Patentes de EE. UU. 7.930.105 y 7.504.214, Publicaciones de Patente de EE.UU. 2011/0136680 y

2013/0281502, y Publicación de Patente PCT WO/2012/079059). Se probaron dos modelos de BCI especificados previamente. Los modelos fueron cúbico (BCI-C) y lineal (BCI-L), basándose en combinaciones cúbicas y lineales de las variables.

La puntuación de BCI se ajustó a escala linealmente hasta una puntuación final (0-10). Los grupos se identificaron como de riesgo bajo, riesgo intermedio y riesgo alto con puntos de corte especificados previamente para cada modelo: BCI-C de riesgo bajo (< 5,0 puntos), BCI-C de riesgo intermedio (5,0 a 6,4) y BCI-C de riesgo alto (> 6,4); y BCI-L de riesgo bajo (<5,0825 puntos), BCI-L de riesgo intermedio (5,0825 a 6,5025) y BCI-L de riesgo alto (> 6,5025). Los grupos principales de riesgo de reaparición de 21 genes se identificaron como se informa anteriormente (Paik et al., 2004). Se establecieron tres grupos de riesgo IHC4 utilizando dos puntos de corte que correspondían a una tasa de reaparición distante en 10 años del 10 % y 20 % (es decir, < 10 %, ≥ 10 % a 20 % y > 20 %) en la cohorte TransATAC, respectivamente. Los puntos de corte de IHC4 no se han validado de forma independiente.

La reaparición distante se definió potencialmente como el criterio principal de valoración, que se refiere a todas las reapariciones en órganos distantes, excluyendo enfermedad contralateral, reapariciones locorregionales e ipsilaterales y otros segundos cánceres primarios. También se incluyó la reaparición distante que tuvo lugar después de la reaparición locorregional como un evento en el momento de la reaparición distante. Se excluyeron las pacientes que murieron antes de la reaparición distante. Todas las reapariciones, muertes por cáncer de mama y supervivencia general (tiempo hasta la muerte por cualquier causa) se definieron como criterios de valoración secundarios. La población de análisis principal fueron pacientes con cáncer de mama receptor de estrógeno positivo, NO, mientras que las poblaciones de análisis secundario incluyeron pacientes con cáncer de mama receptor de estrógeno positivo, NO, HER2 negativo y aquellas con cáncer de mama receptor de estrógeno positivo, ganglios positivos. El objetivo principal del estudio se definió potencialmente como la evaluación de la capacidad pronóstica general (0 a 10 años) del modelo BCI-C para la reaparición distante en pacientes cáncer de mama receptor de estrógeno positivo, NO. Los objetivos secundarios fueron evaluar la capacidad pronóstica del modelo BCI-L y su componente, *HOXB13/IL17BR* y *MGI*, para la reaparición distante general (0 a 10 años), temprana (0 a 5 años) y tardía (5 a 10 años), así como comparar la capacidad de BCI-L con la de la puntuación de reaparición e IHC4.

Análisis estadístico

El comité directivo aprobó un plan de análisis estadístico para los ensayos ATAC y LATIE (Efectos del Tratamiento con Anastrozol a Largo Plazo Frente a tamoxifeno) antes del inicio del estudio. Las reapariciones distantes tempranas se evaluaron censurando el seguimiento de todas los pacientes 5 años después del diagnóstico. Las reapariciones distantes tardías se evaluaron dentro del subconjunto de pacientes que permanecieron sin reaparición distante durante al menos 5 años para evaluar si la firma genética seguía siendo pronóstica después de eliminar su efecto pronóstico de reaparición temprana. Se utilizaron pruebas de razón de verosimilitudes basadas en modelos de regresión de riesgos proporcionales de Cox para probar una diferencia significativa entre un modelo de riesgos proporcionales reducidos basándose en la puntuación del tratamiento clínico y un modelo de riesgos proporcionales completo, incluyendo BCI, puntuación de reaparición de 21 genes, o IHC4. La mejora en la predicción se cuantificó por el cambio en el valor de la razón de verosimilitudes χ^2 (LR- $\Delta\chi^2$), que mide la cantidad de información añadida al modelo de riesgos proporcionales por firmas de genes mosaico en comparación con la puntuación del tratamiento clínico. Debido a que IHC4 se desarrolló en un subconjunto de muestras TransATAC, se realizó la división de la muestra, como se describe anteriormente, para ajustar por posible sobrellenado. Se utilizó el análisis de supervivencia de Kaplan-Meier para presentar gráficamente la proporción de pacientes con reaparición distante en los tres grupos de riesgo especificados previamente de BCI, y se probó la calidad de las curvas con una prueba de rango logarítmico.

El riesgo de reaparición distante se calculó en función del BCI como una covariable lineal de los modelos de riesgos proporcionales de Cox para la reaparición distante general (0-10 años), temprana (0-5 años) y tardía (5-10 años). Para comparar el BCI, la puntuación de reaparición de 21 genes e IHC4, la razón de riesgos instantáneos (HR, por sus siglas en inglés) intercuartílica se estimó comparando el percentil 75 frente al percentil 25 de las puntuaciones continuas de los biomarcadores y el IC del 95 % asociado de los modelos de riesgos proporcionales de Cox. Un valor P doble menor de 0,05 se consideró estadísticamente significativo. Debido a que la puntuación de reaparición ya se había estudiado en TransATAC y se desarrolló IHC4 en un subconjunto de estas pacientes, se especificó previamente la capacidad de la puntuación de reaparición de 21 genes y de IHC4 como puntuaciones continuas, y no se realizó ningún ajuste de prueba múltiple. Los análisis estadísticos se realizaron con STATA versión 12.1.

Ejemplo 2

Pacientes y muestras

Se calcularon valores utilizando la puntuación de reaparición de 21 genes, IHC4 y BCI se calcularon para 915 mujeres, de las cuales 665 tenían cáncer de mama receptor de estrógeno positivo, NO (FIG. 1). Las características clínicas de estas 665 pacientes se enumeran en la Tabla 2 y se comparan con las características de 561 pacientes del Reino Unido con cáncer de mama receptor de estrógeno positivo, NO quienes participaron en el ensayo ATAC pero que no formaron parte de TransATAC. No hay diferencia significativa entre estos dos grupos, excepto que la cohorte no TransATAC tenía significativamente más tumores bien diferenciados y tumores menos moderadamente diferenciados

que las pacientes TransATAC, y significativamente menos reapariciones distantes tardías.

Tabla 1. Características demográficas y clínicas del paciente

	BCI de la cohorte TransATAC NO (n=665)	BCI de la cohorte TransATAC NO HER2neg (n=597)	Pacientes de RU NO No TransATAC* (n=561)	Valor de P
Edad, media (DT)	63,3 (8,1)	63,4 (8,0)	62,6 (7,8)	0,12
IMC, media (DT)	27,1 (4,8)	27,2 (4,8)	26,8 (5,1)	0,28
Tamaño tumoral				0,13
< 2 cm	486 (73,1 %)	442 (74,1 %)	432 (77,0 %)	
2-3 cm	144 (21,7 %)	125 (20,9 %)	95 (16,9 %)	
> 3 cm	35 (5,2 %)	30 (5 %)	29 (5,2 %)	
Desconocido	0	0	5 (0,9 %)	
Grado del tumor				0,0051
Bueno	143 (21,5 %)	138 (23,1 %)	155 (27,6 %)	
Moderado	395 (59,4 %)	357 (59,8 %)	300 (53,5 %)	
Malo	127 (19,1 %)	102 (17,1 %)	78 (13,9 %)	
Desconocido	0	0	28 (5,0 %)	
Radioterapia				0,95
N.º	220 (33,1 %)	189 (31,7 %)	187 (33,3 %)	
Sí	445 (66,9 %)	408 (68,3 %)	374 (66,7 %)	
Mastectomía				0,86
N.º	439 (66,0 %)	404 (67,7 %)	374 (66,7 %)	
Sí	226 (34,0 %)	193 (32,3 %)	187 (33,3 %)	
Tratamiento				
Anastrozol	337 (50,7 %)	309 (51,8 %)	285 (50,8 %)	0,95
Tamoxifeno	328 (49,3 %)	288 (48,2 %)	276 (49,2 %)	
Reaparición distante				
Temprana (0-5 años)	33 (5,0 %)	21 (3,5 %)	23 (4,1 %)	0,56
Tardía (5-10 años)	39 (5,9 %)	36 (6,6 %)	12 (2,3 %)	0,0022

* Estos son pacientes del Reino Unido en el ensayo ATAC que no tienen bloques tumorales disponibles para el estudio traslacional.

la comparación es entre cohortes NO TransATAC y NO no TransATAC, se utilizaron pruebas t para la edad y el IMC, Se utilizó una prueba proporcional basada en la aproximación normal para la reaparición distante, todos los demás utilizaron la prueba exacta de Fisher.

Abreviaturas: ER, receptor de estrógeno; NO, ganglio negativo; HER2neg, receptor 2 del factor de crecimiento epidérmico humano negativo; IMC, índice de masa corporal; RU, Reino Unido

- 5 En mujeres NO en la cohorte de BCI TransATAC, hubo 106 reapariciones, incluidas 72 reapariciones distantes y siete reapariciones locales después de la mastectomía. La mediana de seguimiento en la cohorte de BCI TransATAC fue de 9,97 años (IQR 8,5 a 10).

Cálculo de H/I y MGI

- 10 Generalmente, y con respecto a MGI, se prefiere que los niveles de expresión de los genes divulgados se combinen para formar un índice único que sirva como un factor pronóstico fuerte y predictor de los resultados clínicos. El índice es una suma de los niveles de expresión de los genes usados y usa coeficientes determinados a partir del análisis de componentes principales (PCA, por sus siglas en inglés) para combinar casos de más de un gen divulgado en un solo índice. Los coeficientes se determinan mediante factores tales como la desviación típica de los niveles de expresión de cada gen en un conjunto de datos representativo y el valor de expresión de cada gen en cada muestra. El conjunto de datos representativo tiene un control de calidad basándose en los valores de expresión promedio para los genes de referencia como se divulga en el presente documento.
- 15
- 20 Dicho de manera diferente, se normalizaron los niveles de expresión normalizados para los cinco genes de, p. ej., micromatrices, la secuenciación de próxima generación o RT-PCR a una media de 0 y una desviación típica de 1 en las muestras dentro de cada conjunto de datos y después se combinaron en un solo índice por muestra a través de PCA utilizando el primer componente principal. La normalización de los datos de expresión primaria dentro de cada

conjunto de datos fue necesaria para tener en cuenta las diferentes plataformas (micromatrices, secuenciación y rtPCR) y tipos de muestras (congeladas y FFPE). Como resultado, y siguiendo los parámetros de escala, se genera una fórmula para la suma de valores de expresión que define el índice. La precisión de los parámetros de escala se puede probar basándose en los medios, errores estándar y desviaciones típicas (con intervalos de confianza) de los niveles de expresión de los genes en el conjunto de datos. Por lo tanto, la generación de la fórmula para el índice depende del conjunto de datos, gen de referencia y genes del MGI.

La relación *HOXB13:IL17BR* se calculó como la diferencia en los niveles de expresión normalizados entre *HOXB13* y *IL17BR* como se describe anteriormente (Ma et al., 2006). Las medias y las desviaciones típicas para *HOXB13* e *IL17BR* utilizados para normalizar la cohorte de la Tabla 2 puede derivar de un análisis de 190 secciones de tejido FFPE de una cohorte separada basada en la población de pacientes con cáncer de mama receptor estrógeno positivo, ganglios linfáticos negativo.

Para el MGI, obviamente, los valores de CT en bruto anómalos se eliminaron antes de realizar el promedio de los valores sobre duplicados para cada gen y cada muestra. El valor de CT sin procesar promediado para cada gen se normalizó luego por el valor de CT promedio de cuatro genes de referencia (ACTB, HMBS, SDHA y UBC). Los niveles de expresión normalizados (CT) se compararon con un valor de corte predeterminado, tale como 0, donde el MGI alto está por encima del punto de corte y el MGI bajo está por debajo del punto de corte.

Índice de cáncer de mama (BCI)

BCI se construye combinando H:I y MGI como variables continuas. La linealidad de estas dos variables se verificó ajustando un modelo de regresión de riesgos proporcionales de Cox con splines cúbicos restringidos, y H:I demostró una no linealidad significativa. Se usó una función polinómica de H:I para aproximar los modelos restringidos usando el criterio de información de Akaike. El predictor resultante del modelo de regresión de Cox final se volvió a ajustar a escala en el intervalo de 0 a 10, que se conoce como el BCI.

El BCI se clasifica además en tres niveles: riesgo bajo, riesgo intermedio y riesgo alto como se describe en el presente documento.

PUNTO DE CORTE H/I: Se utilizó en este estudio el punto de corte de 0,06 para la relación *HOXB13: IL17BR*, definido previamente para estratificar a los pacientes tratados con tamoxifeno adyuvante en riesgo bajo y alto de reaparición.

En las 665 pacientes receptor de estrógeno positivo, NO, el análisis de Kaplan-Meier del modelo BCI-C mostró diferencias significativas en la reaparición distante absoluta durante un período de 10 años ($p < 0,001$) en los grupos de riesgo BCI-C categórico especificado previamente, y diferencias en los HR entre el grupo de riesgo bajo y los otros grupos de riesgo, después de ajustar los efectos del tamaño y grado del tumor, edad y tratamiento (según lo determinado por la puntuación del tratamiento clínico; véase la FIG. 2A). BCI-C analizado como variable continua, en lugar de como subgrupos con puntos de corte definidos, no se asoció significativamente con el riesgo general (0 a 10 años) de reaparición distante cuando se ajustó por la puntuación del tratamiento clínico (HR intercuartílico 1,39; LR- $\Delta\chi^2 = 3,70$; $p = 0,054$).

La evaluación de BCI-L en la misma población de pacientes mostró que esta versión se asociaba mucho más fuertemente con el riesgo general de reaparición distante que BCI-C cuando se ajustó por la puntuación del tratamiento clínico (HR intercuartílico 2,30; LR- $\Delta\chi^2 = 22,69$; $p < 0,0001$; Tabla 2).

Tabla 2.

	Todas las reapariciones (0-10 años)		Reaparición temprana (0-5 años)		Reaparición tardía (5-10 años)	
	HR ^x (IC de 95 %)	LR- $\Delta\chi^2$ (valor de p)	HR ^x (IC de 95 %)	LR- $\Delta\chi^2$ (valor de p)	HR* (IC del 95%)	LR- $\Delta\chi^2$ (valor de p)
Univariado						
<u>BCI</u>						
NO	3,12 (2,25-4,32)	49,07 ($p < 0,0001$)	4,11 (2,52-6,70)	34,58 ($p < 0,0001$)	2,47(1,59-3,83)	17,37 ($p < 0,0001$)
NO HER2-negativo	3,30 (2,30-4,73)	46,03 ($p < 0,0001$)	4,22 (2,32-7,64)	25,86 ($p < 0,0001$)	2,84(1,80-4,48)	22,66 ($p < 0,0001$)
<u>Puntuación de reaparición de 21 genes</u>						
NO	1,64 (1,39-1,94)	27,37 ($p < 0,0001$)	1,96 (1,60-2,41)	28,09 ($p < 0,0001$)	1,28(0,95-1,72)	2,99 ($p = 0,21$)

(continuación)						
	Todas las reapariciones (0-10 años)		Reaparición temprana (0-5 años)		Reaparición tardía (5-10 años)	
	HR ^x (IC de 95 %)	LR-Δχ ² (valor de p)	HR ^x (IC de 95 %)	LR-Δχ ² (valor de p)	HR* (IC del 95%)	LR-Δ χ ² (valor de p)
NO HER2-negativo IHC4	1,89 (1,45-2,47)	19,55 (p < 0,0001)	2,38 (1,61-3,53)	16,18 (p < 0,0001)	1,59(1,09-2,31)	6,65 (p=0,014)
NO	2,30 (1,80-2,95)	40,90 (p < 0,0001)	3,38 (2,39-4,78)	42,46 (p < 0,0001)	1,55(1,06-2,26)	5,58 (p=0,022)
NO HER2-negativo	2,66 (1,85-3,81)	27,04 (p < 0,0001)	4,08 (2,26-7,36)	22,13 (p < 0,0001)	2,06(1,29-3,28)	9,32 (p=0,0034)
Multivariante, incluida la puntuación del tratamiento clínico						
BCI						
NO	2,30 (1,62-3,27)	22,69 (p < 0,0001)	2,77 (1,63-4,70)	15,42 (p < 0,0001)	1,95(1,22-3,14)	7,97 (p=0,0048)
NO HER2-negativo	2,49 (1,68-3,68)	21,99 (p < 0,0001)	3,26 (1,96-6,30)	13,65 (p=0,00023)	2,12(1,30-3,47)	9,453 (p=0,0021)
Puntuación de reaparición de 21 genes						
NO	1,48 (1,22-1,78)	13,68 (p=0,0002)	1,80 (1,42-2,29)	18,48 (p < 0,0001)	1,13(0,82-1,56)	0,48 (p = 0,47)
NO HER2-negativo	1,52 (2,15-2,02)	7,65 (p=0,0055)	1,93 (1,26-2,96)	8,37 (p=0,0041)	1,28(0,87-1,88)	1,33 (p = 0,28)
NO	1,69 (1,51-2,56)	22,83 (p=0,0001)	2,90 (2,01-4,18)	29,14 (p < 0,0001)	1,30(0,88-1,94)	1,59 (p = 0,20)
NO HER2-negativo	2,13 (1,45-3,14)	13,75 (p=0,0002)	3,41 (1,83-6,39)	13,83 (p < 0,0001)	1,61(0,98-2,66)	3,30 (p=0,086)

HR = razón de riesgos instantáneos, LR-Δχ² = cambio en el valor χ² basándose en la estadística de razón de verosimilitudes. BCI = ensayo de índice de cáncer de mama. NO = nódulo negativo, IHC4 = cuatro marcadores inmunohistoquímicos (receptor de estrógeno, receptor de progesterona, HER2 y Ki-67). * La HR se calculó entre el IQR de las puntuaciones continuas de cada biomarcador; se utilizó la división de la muestra para calcular las HR y χ² para IHC4.

Las curvas de Kaplan-Meier muestran claras diferencias en las tasas de reaparición distante absoluta según los grupos de riesgo BCI-L especificados previamente (p < 0,001; FIG. 2B). El riesgo general de reaparición distante en 10 años aumentó linealmente con el aumento de BCI-L (FIG. 3).

5 En el subconjunto de 597 pacientes HER2 negativo, NO, tanto BCI-C como BCI-L se asociaron significativamente con el riesgo general de reaparición distante (HR intercuartil de BCI-C 1,65, LR-Δχ²= 6,61, p=0,0001; HR intercuartil de BCI-L 2,49, LR-Δχ²= 21,9, p < 0,0001; Tabla 2). Las curvas de Kaplan-Meier de los grupos especificados previamente para ambas versiones de BCI mostraron diferencias claras en la reaparición distante absoluta (FIG. 4A - BCI-C; 4B - BCI-L).

15 La comparación de la capacidad pronóstica de BCI-L con la de BCI-C mostró que, a diferencia de BCI-C, BCI-L fue un predictor significativo del riesgo de reaparición como variable tanto continua como categórica, y la HR, después del ajuste por puntuación de tratamiento clínico, fue 2,19 frente a 4,86 entre los grupos de riesgo alto y riesgo bajo para BCI-C y BCI-L, respectivamente. EL análisis posterior a continuación utiliza el modelo lineal (denominado BCI en el mismo).

Grupos basados en BCI

20 El BCI se asoció significativamente con el riesgo de reaparición distante temprana (0-5 años) (Tabla 3) cuando se ajustó por la puntuación del tratamiento clínico. Las curvas de Kaplan-Meier (IG. 5A) mostraron diferencias en la tasa absoluta de reaparición distante a los 5 años. Aunque se especificaron previamente tres grupos de riesgo, los resultados del análisis especificado previamente de Kaplan-Meier mostraron que los pacientes de riesgo bajo y riesgo intermedio tenían tasas similares de reaparición distante y constituyen un grupo que es claramente diferente del grupo

de pacientes de riesgo alto.

Un análisis post-hoc de Kaplan-Meier mostró poca diferencia en la reaparición distante a los 5 años entre los grupos de BCI de riesgo bajo y riesgo intermedio, que contenía 556 (84 %) de 665 pacientes (PI) con una tasa combinada de reaparición distante en 5 años del 2-6 %; (Tabla 3).

Tabla 3. Riesgo absoluto de reaparición distante temprana y tardía en subconjuntos clínicamente relevantes de pacientes ER+ NO.

Reaparición Temprana (0-5 años)		
Subconjuntos de riesgo	N(%)	Riesgo de DR temprana a los 5 años (IC del 95 %)
P1 (BCI de riesgo bajo e intermedio)	556 (84 %)	2,6 % (1,5 % - 4,3 %)
P2 (BCI de riesgo alto)	109 (16 %)	18,1 % (12,0 % -27,0 %)
Reaparición Tardía (5-10 años)		
Subconjuntos de riesgo	N (%)	Riesgo de DR temprana a los 5 años (IC del 95 %)
P3 (BCI de riesgo bajo)	366 (61 %)	3,5 % (2,0 % - 6,1 %)
P4 (BCI de riesgo intermedio y alto)	230 (39 %)	13,4 % (9,3 % - 19,0 %)

El grupo BCI de riesgo alto (P2) que contenía 109 (16 %) de 665 pacientes, tuvo una tasa de reaparición distante en 5 años del 18,1 %. Cuando se ajustó por la puntuación del tratamiento clínico, la HR entre P1 y P2 fue de 4,61.

Para reaparición tardía (5-10 años), el BCI se asoció significativamente con el riesgo de reaparición distante cuando se ajustó por la puntuación del tratamiento clínico (Tabla 2). Las curvas de Kaplan-Meier mostraron diferencias en las tasas de reaparición distante absoluta para 5-10 años para los grupos de BCI de riesgo bajo, de riesgo intermedio y de riesgo alto (FIG. 5B). Los resultados del análisis de Kaplan-Meier especificado previamente mostraron que los pacientes de riesgo intermedio y riesgo alto tenían tasas de reaparición muy similares, constituyendo una población que era claramente diferente de la población de pacientes de riesgo bajo. Los análisis adicionales post-hoc de Kaplan-Meier (Tabla 3) mostraron que el grupo de BCI de riesgo bajo (P3) tenía una tasa de reaparición distante del 3,5 % para 5-10 años, sustancialmente diferente de la tasa combinada de grupos de BCI de riesgo intermedio y riesgo alto (P4) del 13,4 %. Ajustando de la puntuación del tratamiento clínico, la HR entre P3 y P4 fue de 2,94. El riesgo de reaparición distante aumentó linealmente con el aumento de los valores de BCI para la reaparición tanto temprana como tardía (FIG. 6A y 6B).

Estado de HER2

Debido a que la historia natural del cáncer de mama receptor de estrógeno positivo, HER2 positivo difiere del cáncer de mama receptor de estrógeno positivo, HER2 negativo, se realizó un análisis de subconjuntos para evaluar si la capacidad pronóstica de BCI en toda la cohorte TransATAC NO receptor de estrógeno positivo se vio indebidamente afectada por la inclusión del subconjunto de pacientes HER2 positivo. En el subconjunto HER2 negativo NO de 597 pacientes (90 % del total del grupo de estudio analizado), el BCI se asoció significativamente con el riesgo de reaparición distante temprana y reaparición distante tardía (Tabla 2), así como con diferencias claras en la reaparición distante absoluta de acuerdo con el grupo de riesgo de BCI (FIG. 7). Tanto para la reaparición temprana como para la tardía, el riesgo de reaparición distante aumentaba con el aumento de los valores de BCI (FIG. 8).

Inhibidores de la aromataza y terapia endocrina

Las curvas de Kaplan-Meier de reaparición distante general (0-10 años) para los grupos de riesgo de puntuación de reaparición de 21 genes y de IHC4 para todos los pacientes, y por separado de acuerdo el grupo de tratamiento (anastrozol o tamoxifeno), se muestran en la FIG. 9. Para todos los pacientes combinados (es decir, aquellos que recibieron anastrozol o tamoxifeno), el grupo de BCI de riesgo bajo tuvo la proporción más baja de pacientes con reaparición distante en 10 años (4,8 %) en comparación con el grupo de riesgo bajo de puntuación de reaparición de 21 genes (6,5 %) y el grupo de riesgo bajo de IHC4 (6,2 %), mientras que el grupo de BCI de riesgo alto tuvo la mayor proporción de reaparición distante (29,0 %) en comparación con el grupo de riesgo alto de puntuación de reaparición de 21 genes (27,1 %) y el grupo de riesgo alto de IHC4 (21,8 %); FIG. 9).

Adicionalmente, como se muestra en las FIG. 9D-9F, BCI estratificó el riesgo de reaparición distante entre los grupos de anastrozol de riesgo alto y bajo mucho mejor que los sistemas de 21 genes e IHC4, ya que BCI tuvo el % de reaparición más alto en su grupo de riesgo alto (21,6 % frente a 13,5 % y 15,6 %), y también tuvo el % de reaparición más bajo en su grupo de riesgo bajo (4,8 % frente a 9,4 % y 8,0 %).

Comparación con otras evaluaciones

El cambio en los valores de la razón de verosimilitudes $LR-\Delta\chi^2$ se utilizó para proporcionar una comparación directa cabeza a cabeza de BCI con el IHC4 y la puntuación de reaparición de 21 genes. La capacidad de pronóstico relativa de cada biomarcador varió dependiendo del período de tiempo de reaparición distante (Tabla 2). Para una reaparición temprana, BCI, IHC4 y la puntuación de reaparición de 21 genes fueron todos pronósticos de reaparición distante tanto en análisis univariados como multivariados (Tabla 2). En todas las pacientes NO, IHC4 fue más pronóstico que la puntuación de reaparición y el BCI después de ajustar la puntuación del tratamiento clínico. Sin embargo, en las pacientes NO HER2 negativo, BCI e IHC4 tenían capacidades de pronóstico similares que eran mejores que las de la puntuación de reaparición de 21 genes después de ajustar la puntuación del tratamiento clínico (Tabla 2). En el análisis multivariado de reaparición tardía, solo el BCI mantuvo un fuerte pronóstico en todas las pacientes NO y NO HER2 negativo, mientras que tanto IHC4 como la puntuación de reaparición de 21 genes no fueron pronósticos en ninguna de las poblaciones (Tabla 2). Se observaron resultados similares considerando todas las reapariciones, muertes por cáncer de mama y supervivencia global como criterios de valoración (Tabla 4).

Tabla 4. Rendimiento de pronóstico comparativo para eventos secundarios de enfermedad temprana y tardía de BCI, RS (puntuación de reaparición de 21 genes) e IHC4 en todas las pacientes NO receptor de hormonas positivo y el subconjunto NO HER2.

		Reaparición Temprana (0-5 años)		Reaparición tardía (5-10 años)	
		HR (IC de 95 %)*	LR- $\Delta\chi^2$ (Valor p)	HR (IC de 95 %)*	LR- $\Delta\chi^2$ (valor p)
UNIVARIADO					
BCI		Todas las reapariciones			
	NO	2,58 (1,74-3,82)	23,08 (< 0,0001)	1,79 (1,26-2,54)	10,53 (0,0012)
	NO/HER2-	2,20 (1,40-3,45)	11,92 (0,00061)	1,95 (1,37-2,79)	13,60 (0,00021)
RS	NO	1,90 (1,57-2,29)	32,11 (< 0,0001)	1,28 (0,95-1,72)	2,02 (0,15)
	NO/HER2-	2,03 (1,46-2,82)	14,62 (0,00012)	1,48 (1,09-2,02)	5,52 (0,018)
IHC4	NO	2,52 (1,88-3,40)	33,37 (< 0,0001)	1,55 (1,06-2,26)	4,47 (0,034)
	NO/HER2-	2,48 (1,54-3,98)	13,23 (0,00031)	1,91 (1,31-2,78)	10,50 (0,0012)
BCI	NO	Muerte por cáncer de mama 5,82 (3,10-10,92)	34,36 (< 0,0001)	2,23 (1,29-3,86)	8,42 (0,0037)
	NO/HER2-	7,30 (3,22-16,51)	26,42 (< 0,0001)	2,52 (1,42-4,45)	10,38 (0,0013)
RS	NO	2,05 (1,62-2,60)	24,15 (< 0,0001)	1,41 (1,01-1,99)	3,24 (0,072)
	NO/HER2-	2,90 (1,80-4,65)	15,35 (0,0001)	1,78 (1,14-2,77)	5,36 (0,021)
IHC4	NO	3,66 (2,40-5,56)	33,84 (< 0,0001)	1,73 (1,09-2,75)	5,03 (0,024)
	NO/HER2-	4,91 (2,32-10,41)	16,62 (< 0,0001)	2,26 (1,27-4,01)	7,13 (0,0076)
Supervivencia global					
BCI	NO	2,25 (1,54-3,28)	18,26 (< 0,0001)	1,96 (1,43-2,68)	17,87 (< 0,0001)
	NO/HER2-	2,04 (1,34-3,12)	11,07 (0,00091)	2,04 (1,47-2,82)	18,63 (< 0,0001)
RS	NO	1,54 (1,26-1,89)	13,59 (0,00020)	1,19 (0,94-1,50)	1,93 (0,16)
	NO/HER2-	1,58 (1,12-2,24)	5,86 (0,016)	1,33 (0,99-1,78)	3,17 (0,075)
IHC4	NO	1,84 (1,37-2,48)	14,45(0,00010)	1,25 (0,94-1,67)	2,21 (0,14)
	NO/HER2-	1,57 (0,98-2,51)	3,36(0,066)	1,45 (1,01-2,07)	3,97 (0,046)
MULTIVARIADO INCLUYENDO CTS					
BCI		Todas las reapariciones			
	NO	1,99 (1,27-2,99)	9,71 (0,0018)	1,49 (1,02-2,17)	4,25 (0,039)
	NO/HER2-	1,83 (1,12-3,01)	5,93 (0,014)	1,57 (1,07-2,30)	5,34 (0,021)

(continuación)					
Reaparición Temprana (0-5 años)			Reaparición tardía (5-10 años)		
		HR (IC de 95 %)*	LR- $\Delta\chi^2$ (Valor p)	HR (IC de 95 %)*	LR- $\Delta\chi^2$ (valor p)
MULTIVARIADO INCLUYENDO CTS					
RS	NO	1,76 (1,43-2,17)	22,21 (< 0,0001)	1,10 (0,83-1,46)	0,44 (0,51)
	NO/HER2-	1,80 (1,26-2,56)	9,41 (0,0022)	1,27 (0,92-1,74)	2,01 (0,16)
IHC4	NO	2,22 (1,62-3,03)	22,48 (< 0,0001)	1,25 (0,90-1,73)	1,65 (0,19)
	NO/HER2-	2,16 (1,31-3,57)	8,60 (0,0034)	1,61 (1,08-2,40)	5,27 (0,022)
BCI	NO	Muerte por cáncer de mama 3,92 (1,98-7,76)	17,41 (< 0,0001)	1,68 (0,94-3,01)	3,10 (0,078)
	NO/HER2-	7,18 (2,77-18,64)	20,17 (< 0,0001)	1,78 (0,97-3,26)	3,60 (0,057)
RS	NO	1,92 (1,46-2,52)	16,62 (< 0,0001)	1,26 (0,86-1,84)	1,29 (0,25)
	NO/HER2-	2,54 (1,49-4,31)	10,70 (0,0011)	1,40 (0,88-2,23)	1,90 (0,16)
IHC4	NO	3,19 (2,05-4,98)	24,02 (< 0,0001)	1,45 (0,88-2,37)	2,04 (0,15)
	NO/HER2-	4,31 (1,96-9,47)	12,50 (0,00041)	1,73 (0,93-3,22)	2,83 (0,092)
BCI	NO	Supervivencia global 1,75 (1,16-2,64)	7,25 (0,0071)	1,51 (1,08-2,11)	5,86 (0,015)
	NO/HER2-	1,77 (1,11-2,84)	5,91 (0,015)	1,54 (1,08-2,18)	5,98 (0,014)
RS	NO	1,40 (1,12-1,75)	7,37 (0,0066)	1,04 (0,81-1,33)	0,07 (0,78)
	NO/HER2-	1,39 (0,97-2,01)	2,97 (0,084)	1,08 (0,80-1,46)	0,25 (0,61)
IHC4	NO	1,58 (1,15-2,17)	7,39 (0,0066)	1,02 (0,75-1,39)	0,02 (0,89)
	NO/HER2-	1,33 (0,81-2,20)	1,23 (0,27)	1,11 (0,76-1,63)	0,30 (0,58)

* La HR se calculó entre el intervalo intercuartílico de las puntuaciones continuas de cada biomarcador.

Abreviaturas: BCI, Índice de Cáncer de Mama; RS, Puntuación de reaparición de OncotypeDX; IHC4, cuatro marcadores inmunohistoquímicos (receptor de estrógeno, receptor de progesterona, factor de crecimiento epidérmico humano 2 y Ki-67; HR, razón de riesgos instantáneos; LR- $\Delta\chi^2$, valor χ^2 basado en la estadística de razón de verosimilitudes; CTS, puntuación de tratamiento clínico; NO, ganglio negativo; HER2-, receptor del factor de crecimiento epidérmico negativo.

Cáncer de mama ganglios positivos

- 5 Aunque el análisis principal de estos ejemplos se centró en pacientes NO, un análisis de pacientes con ganglios positivos mostró que el BCI también era un pronóstico de reaparición distante en estos pacientes (rango logarítmico $p = 0,0045$; FIG. 10). Asimismo, un análisis comparativo mostró que BCI, IHC4, y la puntuación de reaparición de 21 genes tuvieron una capacidad de pronóstico muy similar en esta población de pacientes, aunque menos fuerte que el observado en el subconjunto NO (Tabla 5).

- 10 Tabla 5. Rendimiento de pronóstico comparativo para la reaparición distante en 0-10 años de BCI, RS, IHC4 en pacientes receptor de hormonas positivo, ganglios positivos

		HR* (IC del 95 %)	LR- $\Delta\chi^2$ (Valor p)
UNIVARIADO			
	BCI	1,70 (1,21-2,40)	9,48 (0,0021)
	RS	1,30 (1,07-1,58)	5,97 (0,014)
	IHC4	1,40 (1,07-1,85)	5,56 (0,018)
MULTIVARIADO INCLUYENDO CTS			
	BCI	1,42 (1,02-1,97)	4,49 (0,034)
	RS	1,25 (1,02-1,52)	4,24 (0,039)
	IHC4	1,40 (1,05-1,87)	4,92 (0,027)

La tabla 5 muestra el análisis multivariado en relación con la reaparición de cáncer.

Como se muestra en lo anterior, el BCI es pronóstico de reapariciones de cáncer tardías en pacientes ER+ después de 5 años de tratamiento con tamoxifeno. La expresión de *HoxB13* también es un pronóstico de reapariciones de cáncer tardías en pacientes ER+ después de 5 años de tratamiento con tamoxifeno.

Referencias

- 10 Cuzick et al., Lancet Oncol. 11:1135-1141 (2010).
Cuzick et al., J. Clin. Oncol. 29:4273-4278 (2011).
Dowsett et al., J. Clin. Oncol. 28:1829-1834 (2010).
Goetz et al., Clin Cancer Res. 12:2080-7 (2006).
Jansen et al., J. Clin. Oncol. 25:662-8 (2007).
Jerevall et al., Breast Cancer Res.Treat (2007).
- 15 Ma et al., Cancer Cell, 5:607-16 (2004).
Ma et al., J. Clin. Oncol., 24:4611-9 (2006).
Paik et al., N. Engl. J. Med. 351:2817-26 (2004).
Sgroi, et al., Proc SABVS; Abstract S1-9 (2012).
Patente de EE.UU. N.º 6.291.170
- 20 Patente de EE.UU. N.º 6.794.141.
Patente de EE.UU. N.º 7.930.105.
Patente de EE.UU. N.º 7.504.214.
Publicación de Solicitud de Patente de EE.UU. 2005/0239079.
Publicación de Solicitud de Patente de EE.UU. 2005/0239083.
- 25 Publicación de Solicitud de Patente de EE.UU. 2006/0154267.
Publicación de Solicitud de Patente de EE.UU. 2011/0136680.
Publicación de Solicitud de Patente de EE.UU. 2013/0281502.
Publicación de Patente PCT WO 2012/079059.
- 30 En vista de lo anterior, se verá que se logran varios objetivos de la invención y se obtienen otras ventajas.

REIVINDICACIONES

1. Un método para evaluar cáncer de mama de un sujeto, comprendiendo el método medir los niveles de expresión del ARNm de *HoxB13*, *IL17BR*, *Bub1B*, *CENPA*, *NEK2*, *RACGAP1* y *RRM2* en una muestra de células de cáncer de mama ER+ del sujeto; calcular una relación de niveles de expresión de *HoxB13/IL17BR* (H:I); sumar los niveles de expresión de *Bub1B*, *CENPA*, *NEK2*, *RACGAP1* y *RRM2* usando coeficientes para obtener un índice MGI; combinar linealmente H:I y MGI como variables continuas para formar un valor de Índice de Cáncer de Mama (BCI); y clasificar la muestra, basándose en el valor de BCI, como se indica:
 - (i) un riesgo bajo o un riesgo alto de reaparición de cáncer en el sujeto, sin categoría de riesgo intermedio, en donde un valor de BCI mayor se correlaciona con un mayor riesgo de reaparición de cáncer y un valor de BCI menor se correlaciona con un menor riesgo de reaparición de cáncer; o
 - (ii) necesidad o falta de necesidad de tratamiento, en donde un valor de BCI menor se correlaciona con no necesitar tratamiento con un inhibidor de la aromatasa, terapia dirigida o terapia endocrina después de un tratamiento inicial con un inhibidor de la aromatasa, terapia dirigida o terapia endocrina durante cinco años o menos.
2. El método de la reivindicación 1, en donde los coeficientes se determinan a partir del análisis de componentes principales.
3. El método de la reivindicación 1, en donde dicho cáncer es carcinoma ductal *in situ* (DCIS) y dicha reaparición de cáncer comprende la reaparición local.
4. El método de la reivindicación 1, en donde dicha reaparición es una reaparición más de 5 años después de la terapia inicial.
5. El método de una cualquiera de las reivindicaciones 1-4, en donde:
 - a) la muestra se clasifica como que indica un riesgo alto de reaparición distante en el sujeto si se trata con un inhibidor de la aromatasa, terapia dirigida, o terapia endocrina, durante un período inicial de 5 años o menos; en donde opcionalmente el método comprende además recomendar el tratamiento al sujeto con quimioterapia;
 - b) la muestra se clasifica como que indica riesgo bajo de reaparición distante en el sujeto si se trata con un inhibidor de la aromatasa, terapia dirigida, o terapia endocrina, durante un período inicial de 5 años o menos; opcionalmente en donde el método comprende además recomendar el tratamiento al sujeto con dicho inhibidor de la aromatasa, terapia dirigida o terapia endocrina; o
 - c) la muestra se clasifica como que indica un riesgo alto de reaparición distante en el sujeto si no se trata con un inhibidor de la aromatasa, terapia dirigida, o terapia endocrina, durante un segundo período de 5 años después de un primer período de 5 años; opcionalmente en donde el método comprende además recomendar el tratamiento al sujeto con un inhibidor de la aromatasa, terapia dirigida, terapia endocrina o terapia adyuvante.
6. El método de la reivindicación 1, en donde el sujeto es un sujeto que se trató por cáncer de mama durante un período de cinco años o menos, y además en donde la necesidad o la falta de necesidad indicada para el tratamiento es un tratamiento de cinco años o más.
7. El método de cualquier reivindicación anterior, en donde el sujeto tiene cáncer de mama con ganglios linfáticos negativos.
8. El método de cualquier reivindicación anterior, en donde el sujeto tiene cáncer de mama HER2 negativo.
9. El método de cualquier reivindicación anterior, en donde la muestra se disecciona del tejido del sujeto.
10. El método de cualquier reivindicación anterior, en donde dicha muestra es una muestra fijada en formalina embebida en parafina.
11. El método de cualquier reivindicación anterior, en donde el inhibidor de la aromatasa, terapia dirigida o terapia endocrina es tamoxifeno.
12. El método de la reivindicación 1, en donde el riesgo de reaparición es un riesgo de reaparición distante, riesgo de reaparición local o riesgo de reaparición después de 5 años.
13. El método de la reivindicación 1, en donde el riesgo bajo es un riesgo de reaparición de cáncer después de recibir aproximadamente cinco años de terapia adyuvante que es menos de aproximadamente el 5 %.
14. El método de la reivindicación 1, en donde el riesgo bajo y riesgo alto de reaparición de cáncer se distinguen en relación con un valor de corte seleccionado, opcionalmente en donde el valor de corte se selecciona de manera que más del 50 %, más del 55 % o más del 60 % de los sujetos ER+ en una población de referencia se clasifican como de riesgo bajo.

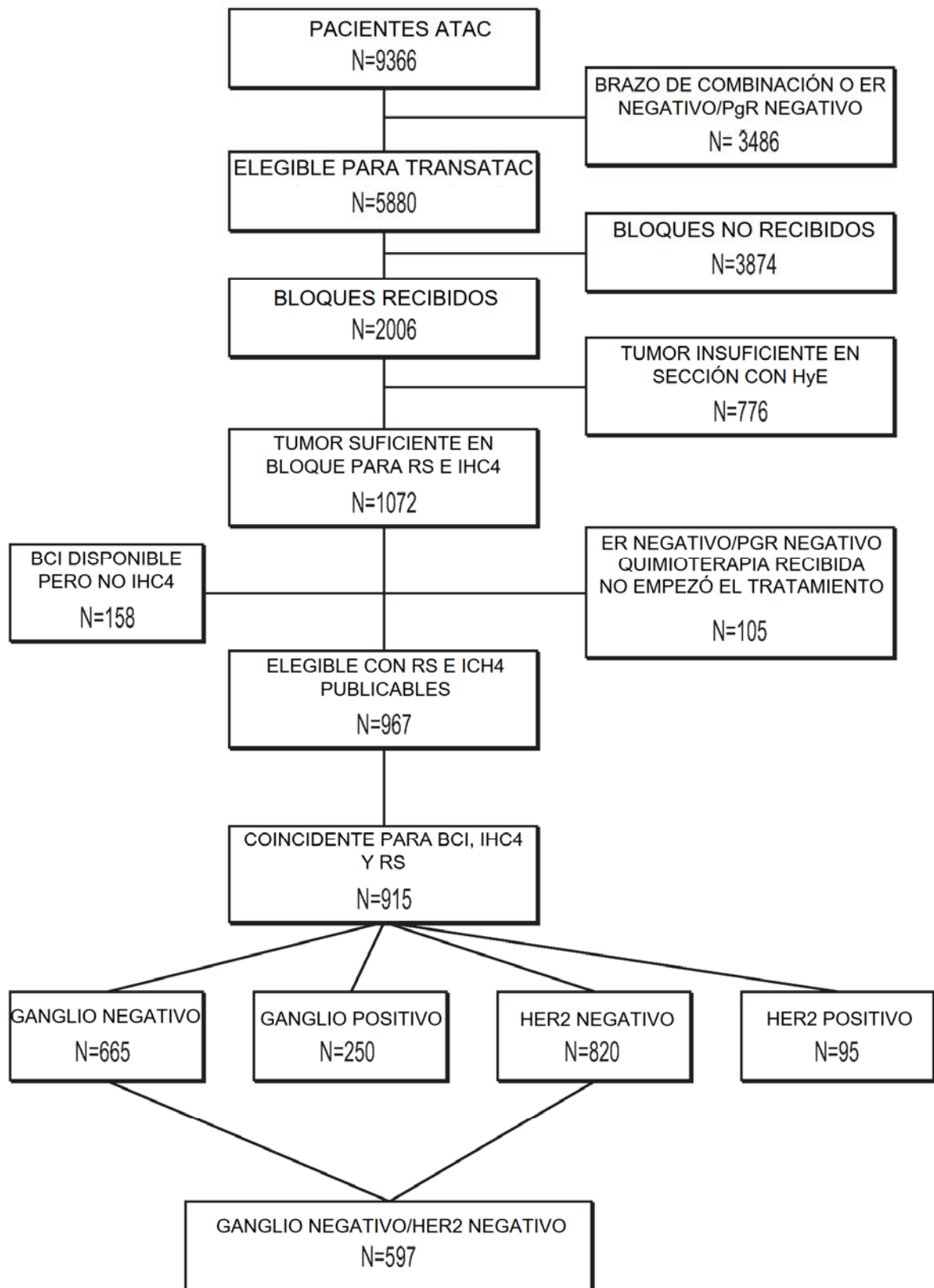


FIG. 1

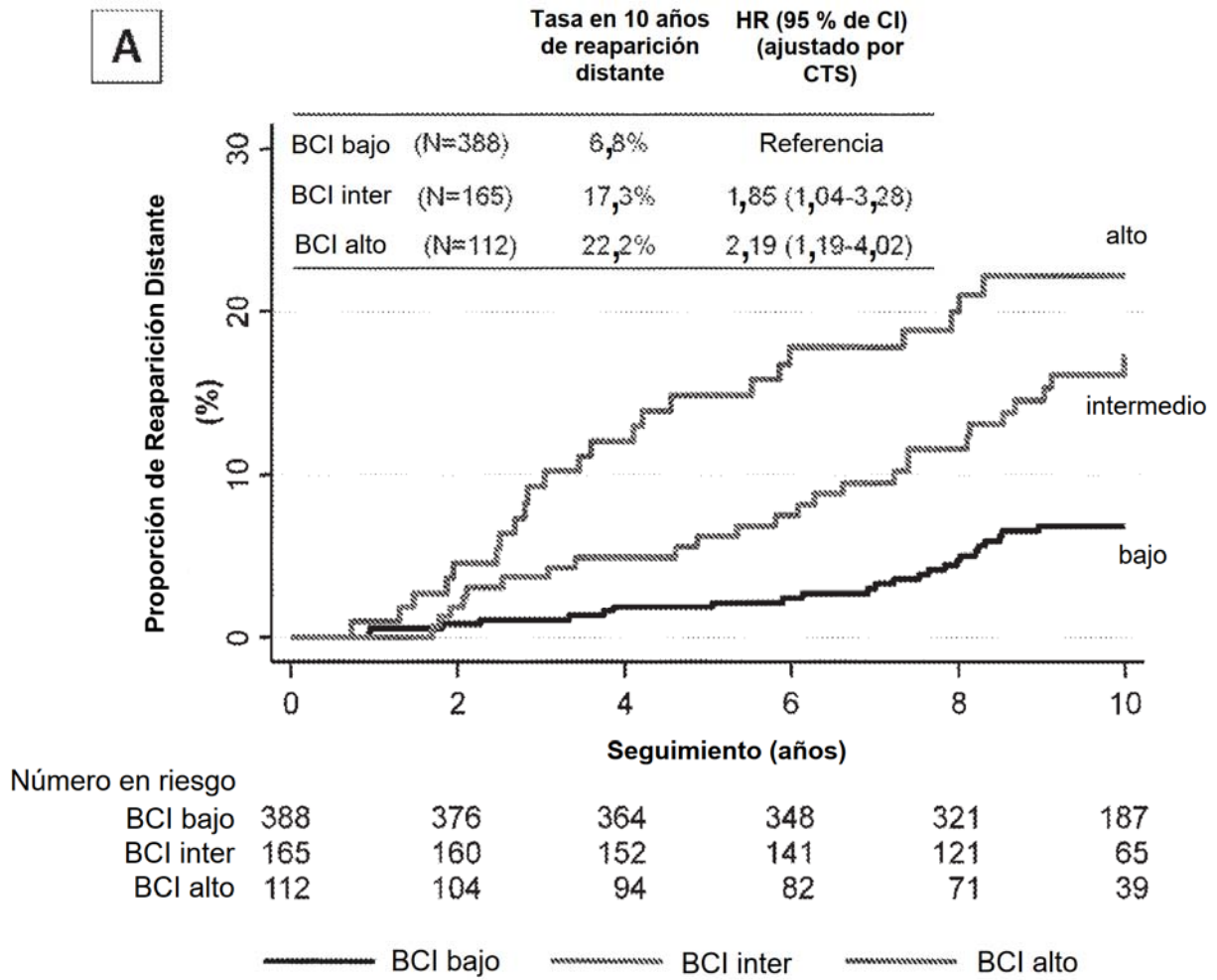


FIG. 2

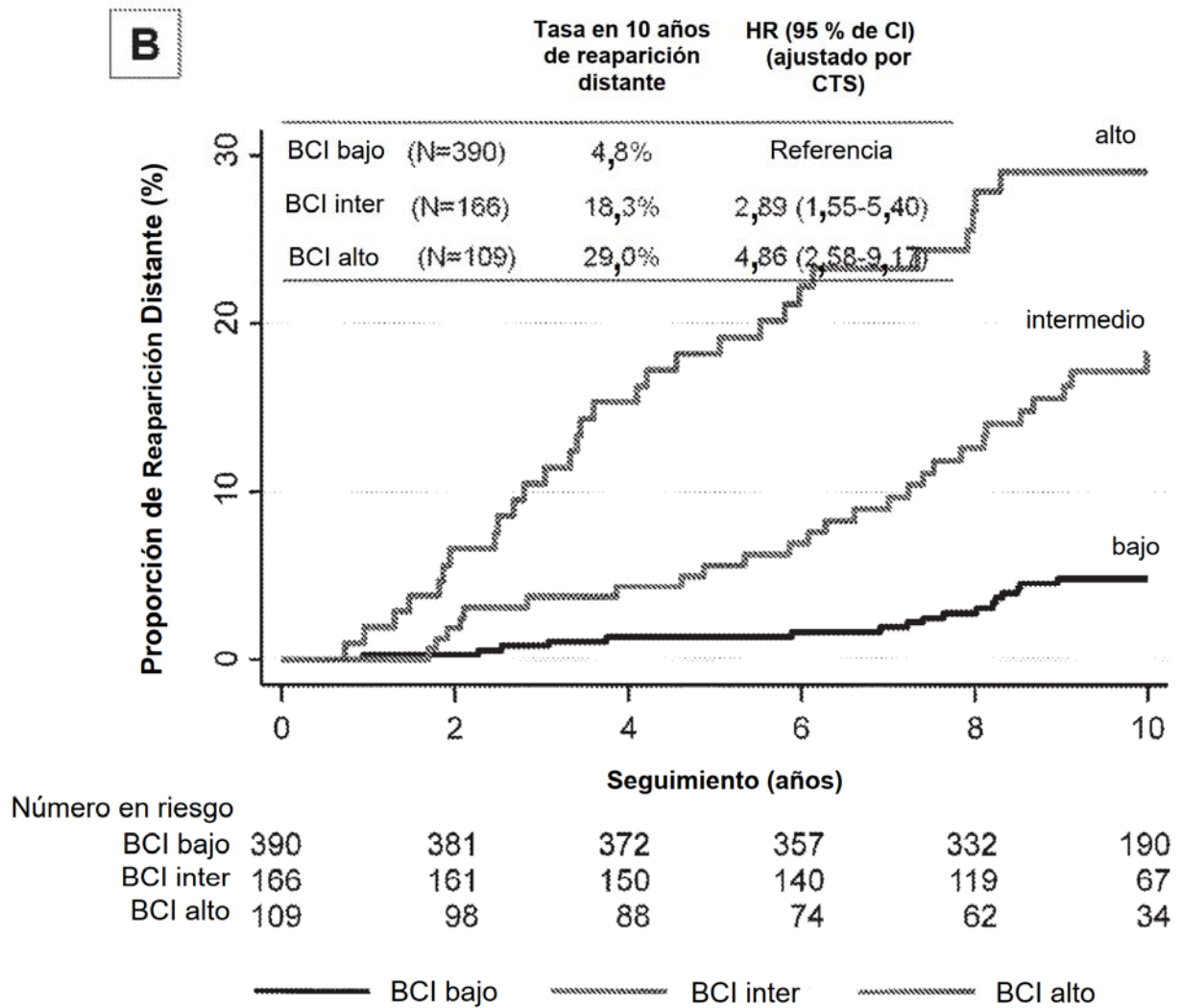


FIG. 2 (cont)

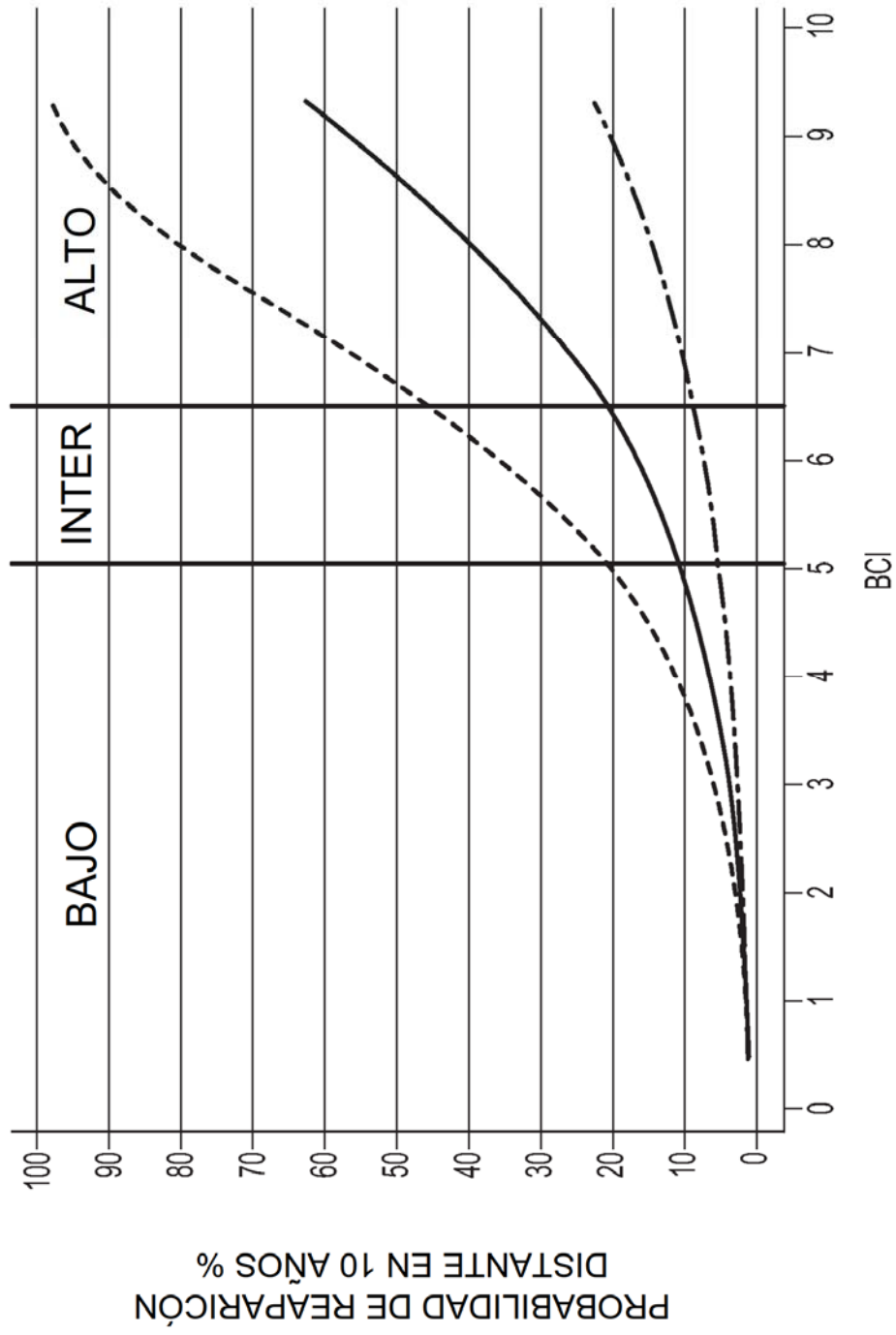


FIG. 3

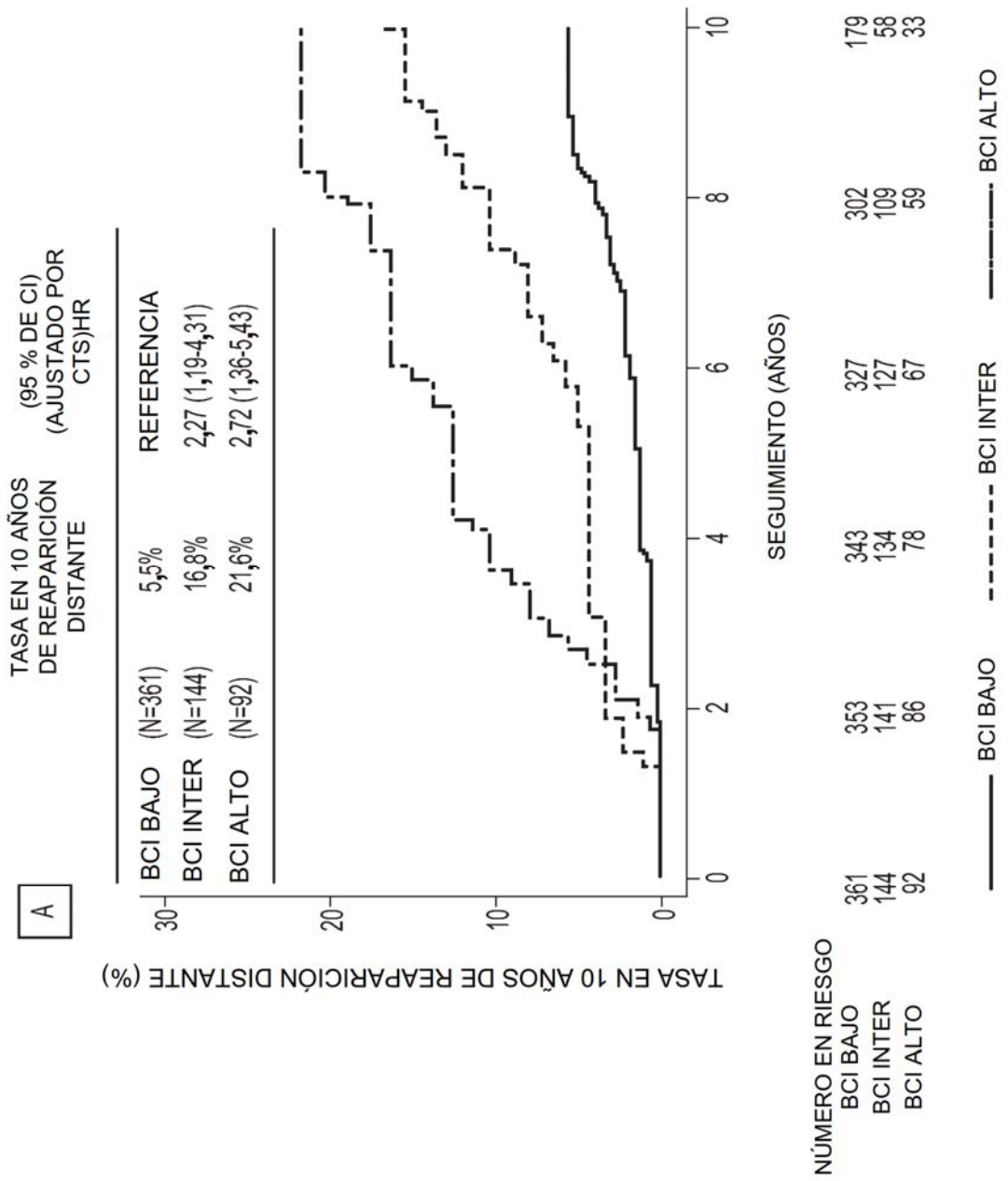


FIG. 4

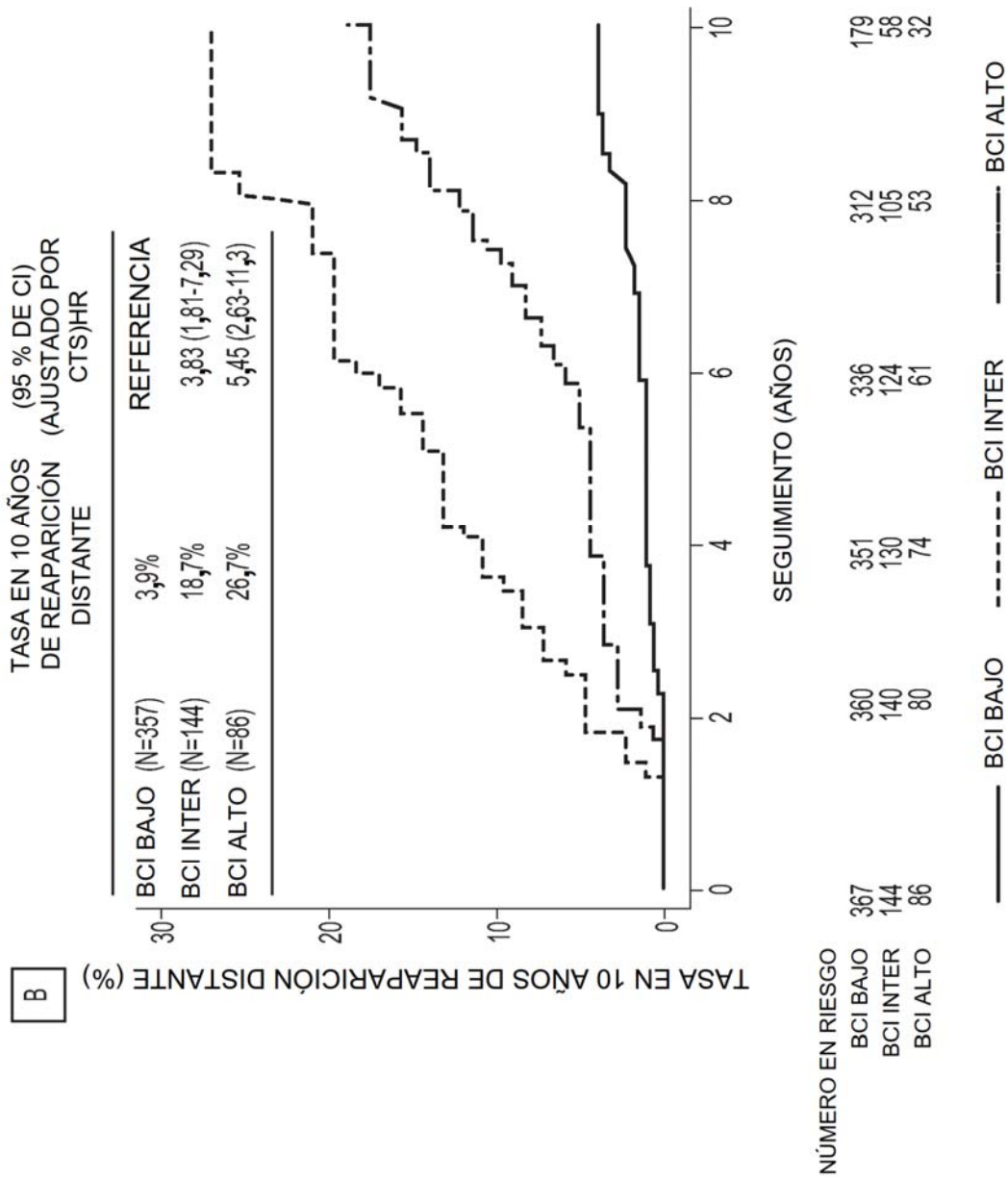


FIG. 4 (cont)

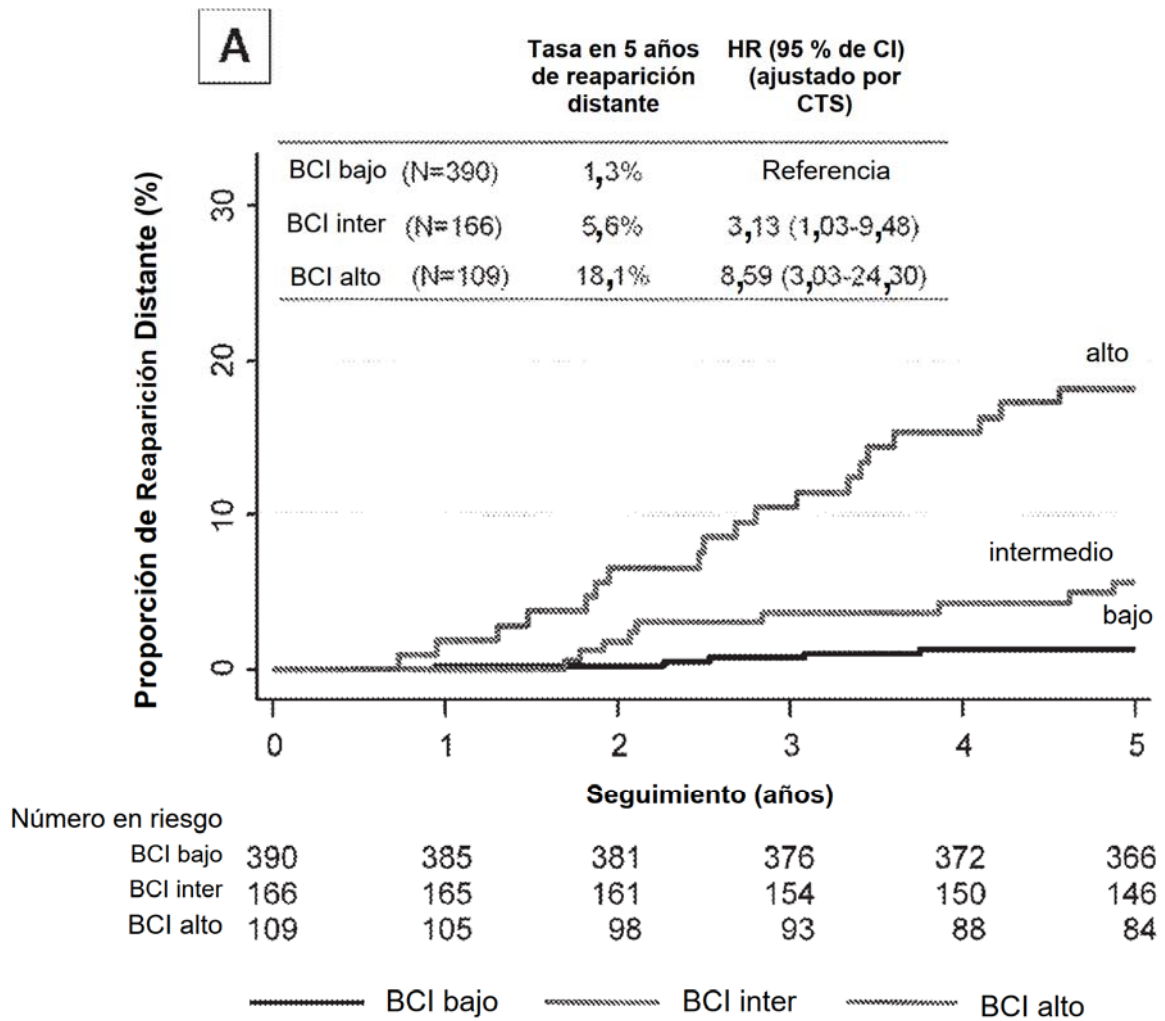


FIG. 5

B

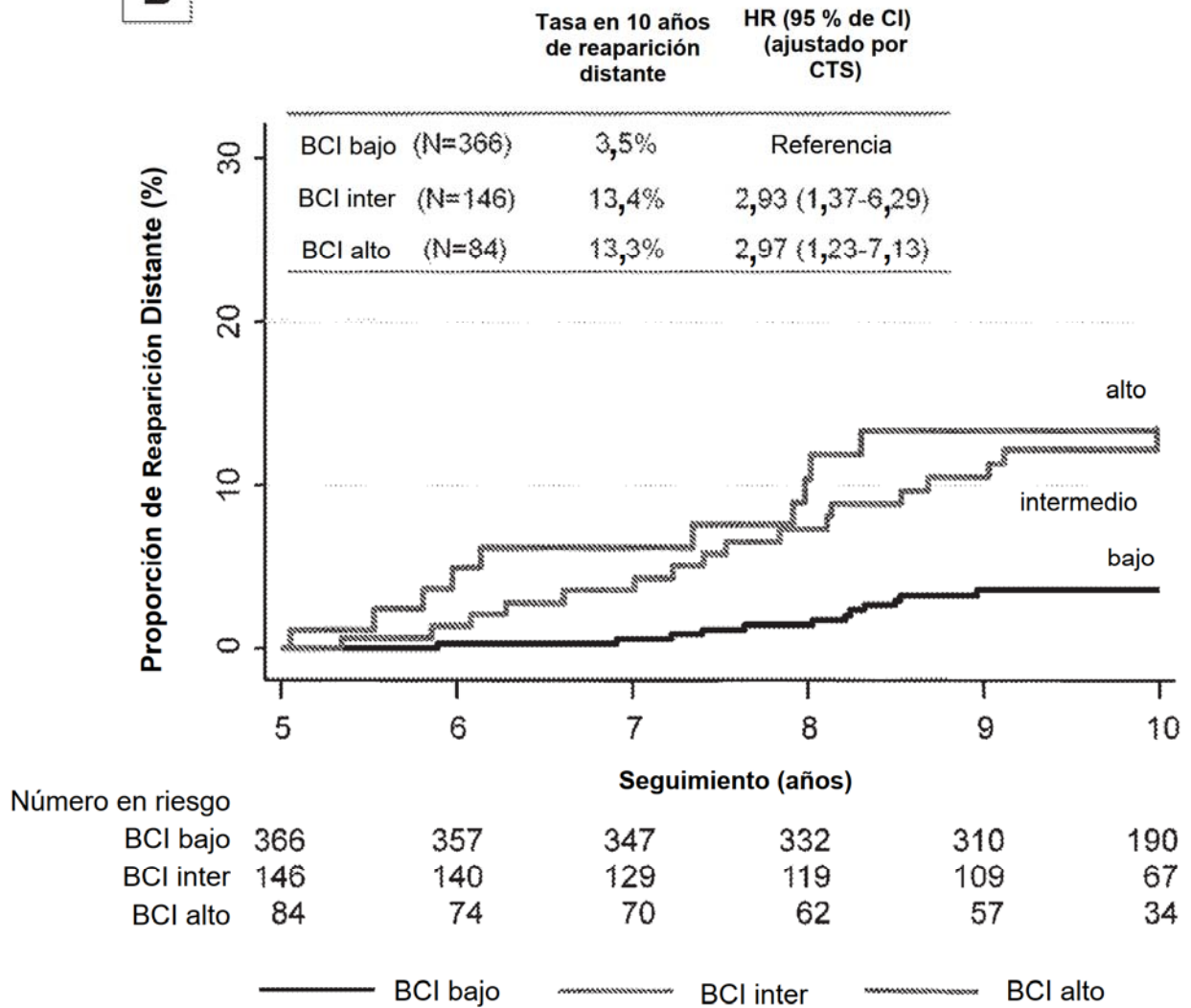


FIG. 5 (cont)

A

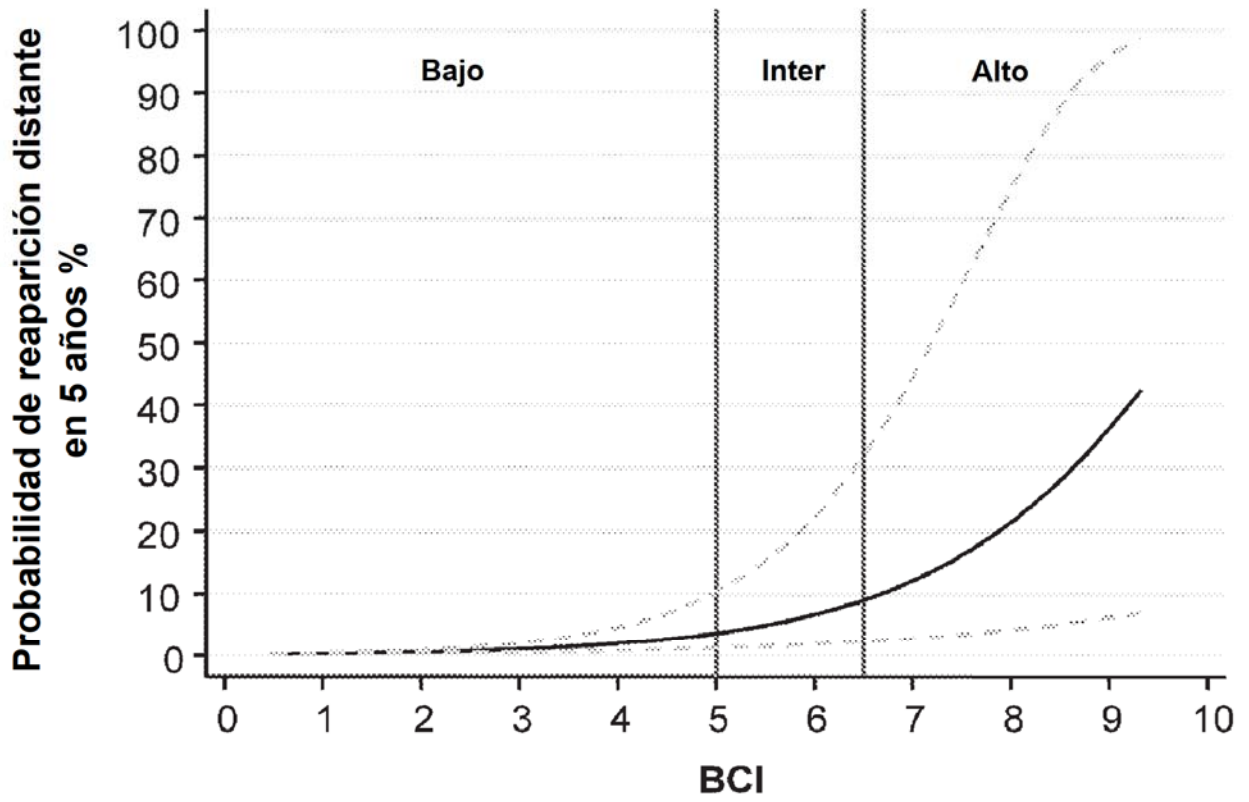


FIG. 6

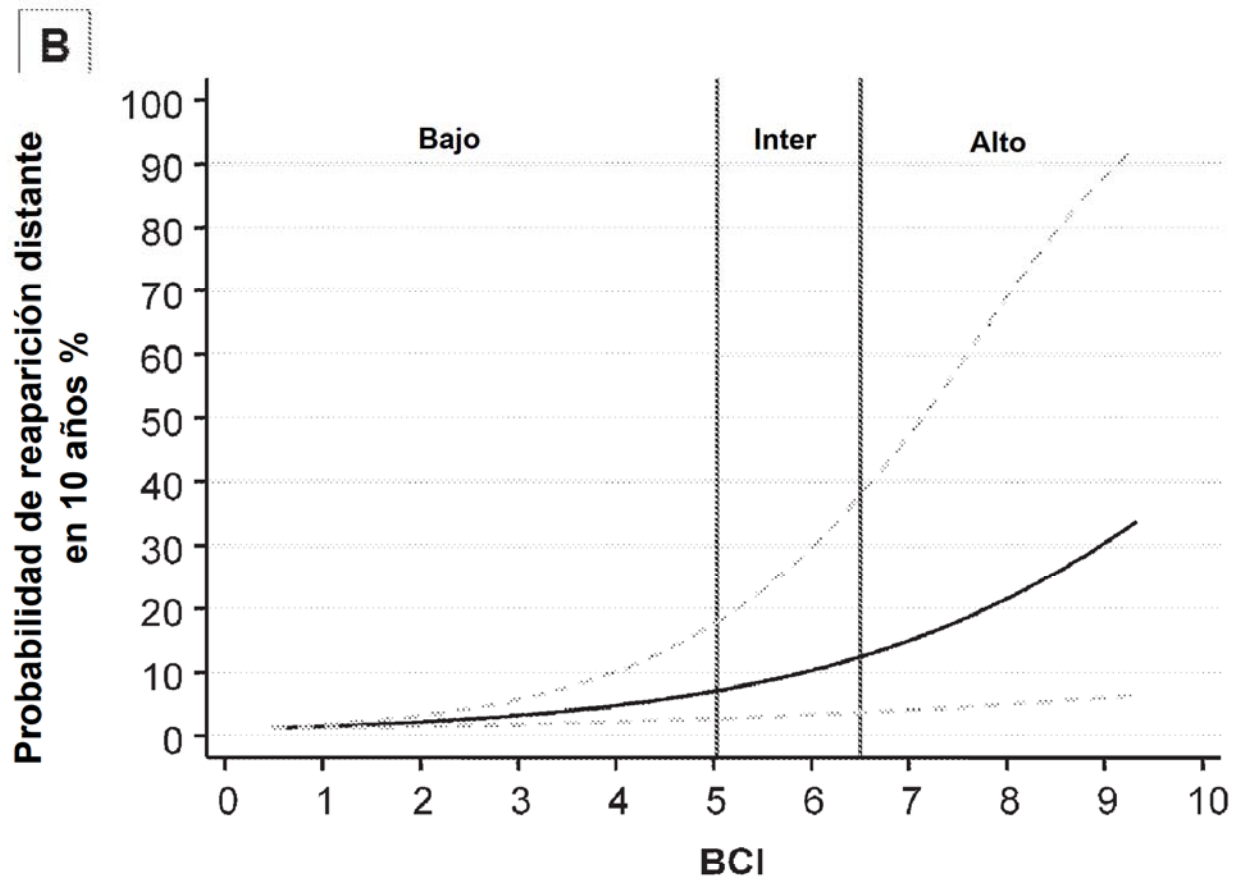
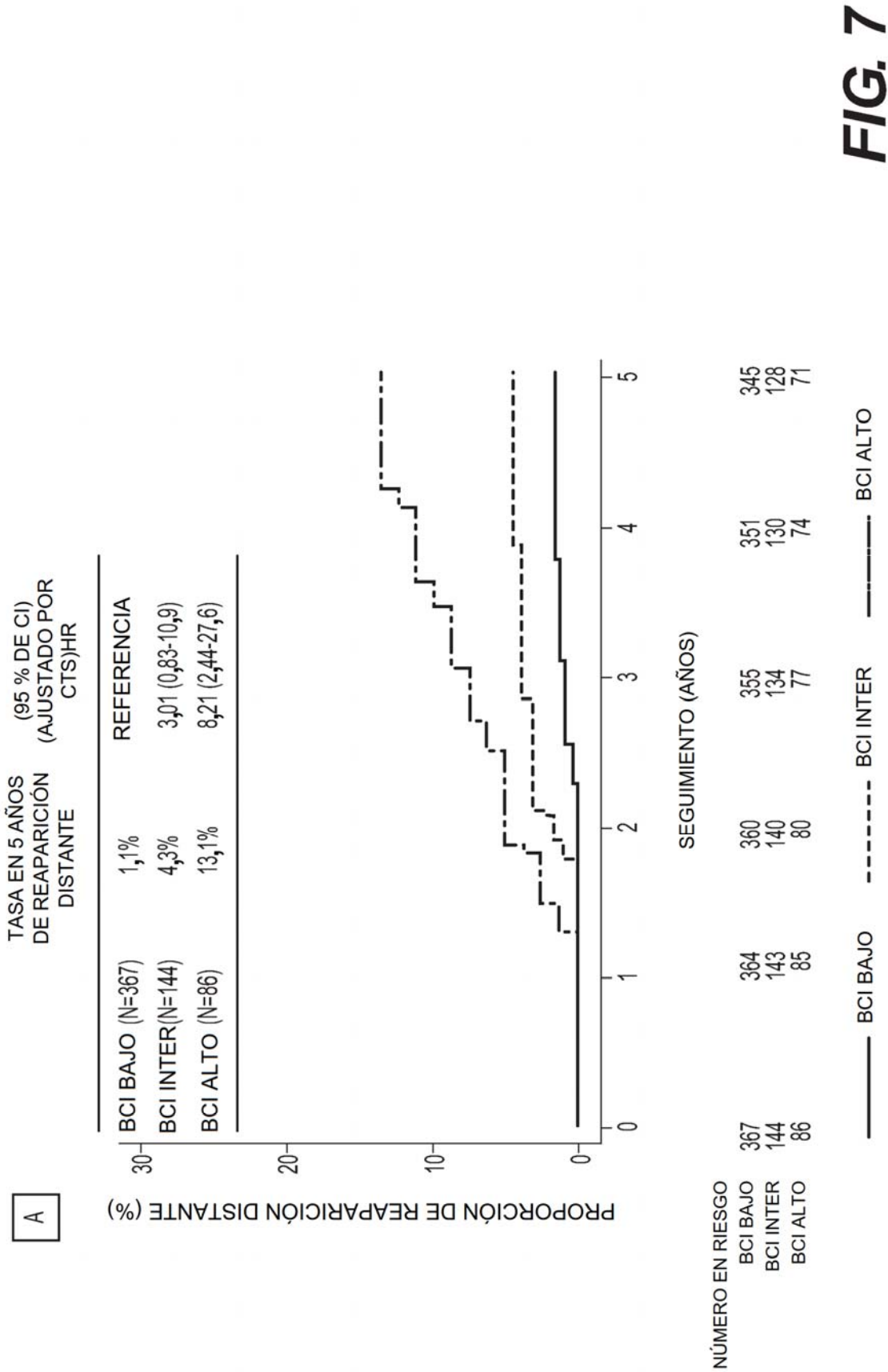


FIG. 6 (cont)



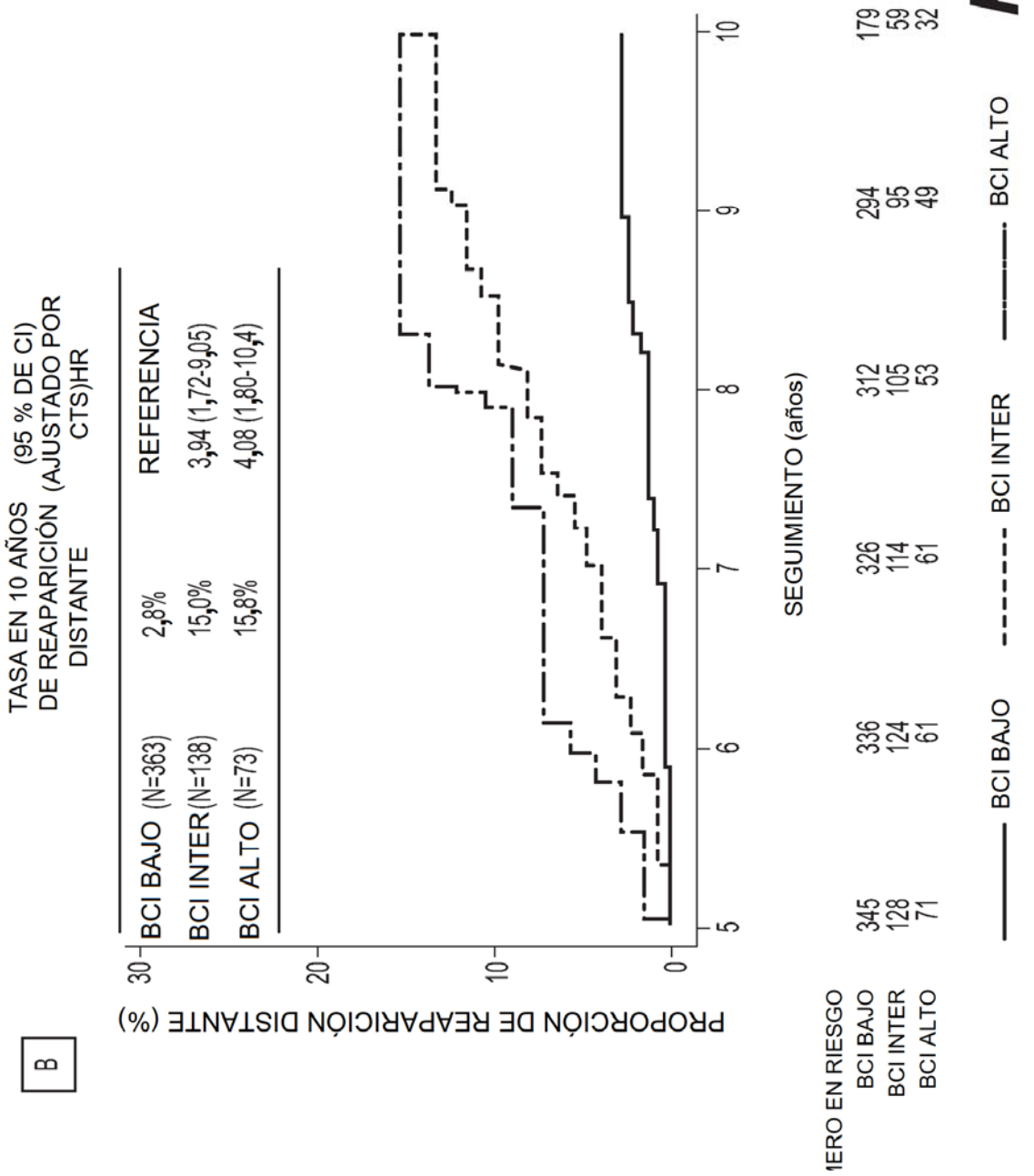


FIG. 7 (cont)

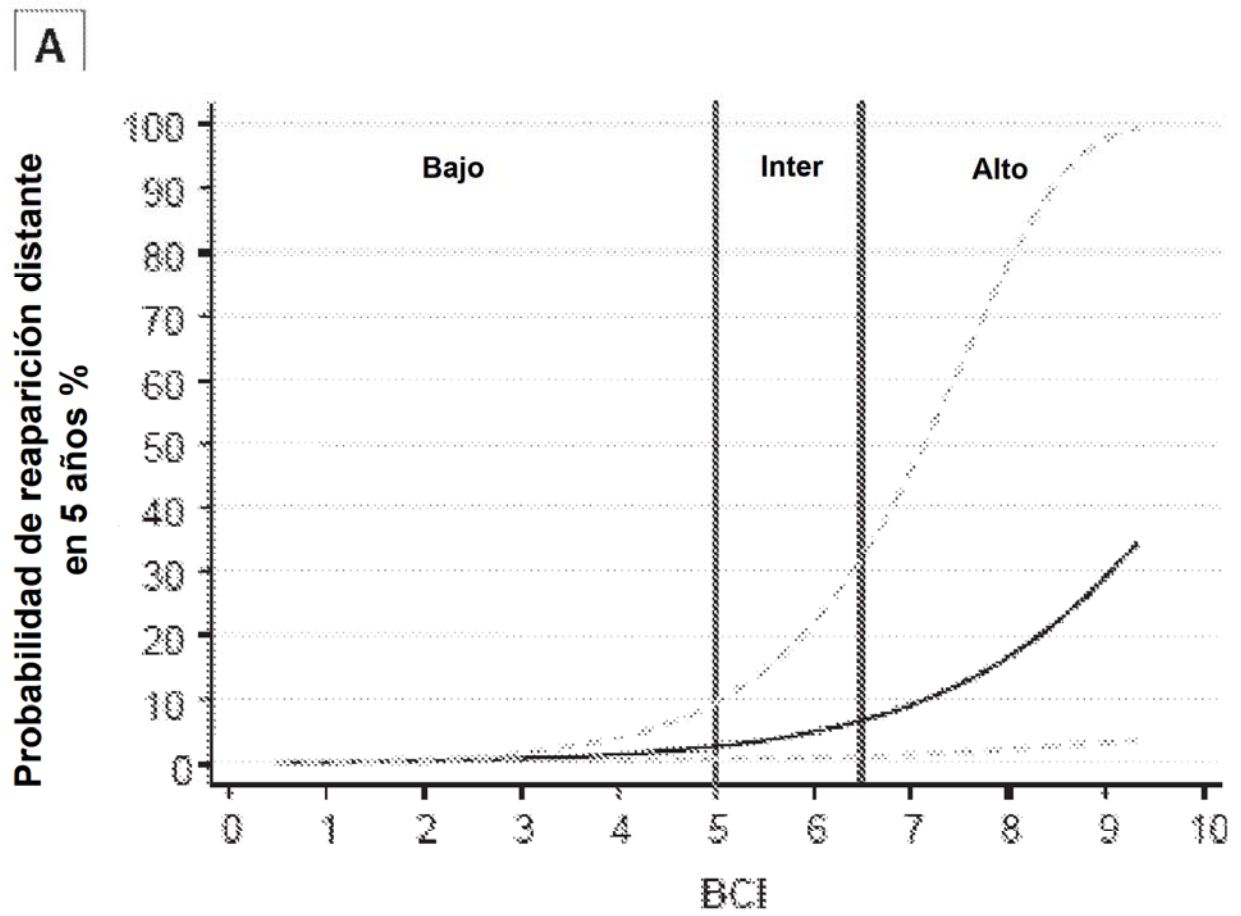


FIG. 8

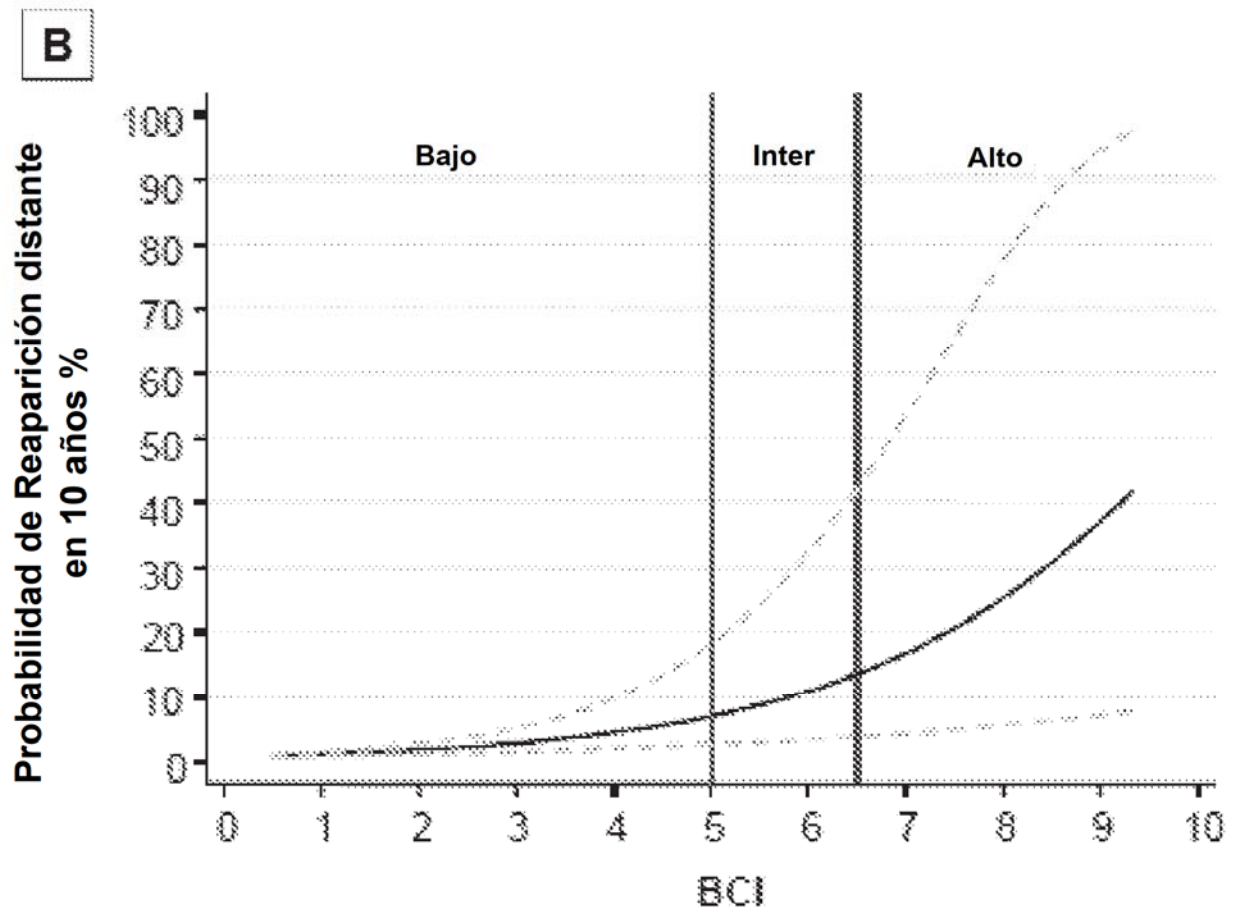


FIG. 8 (cont)

A

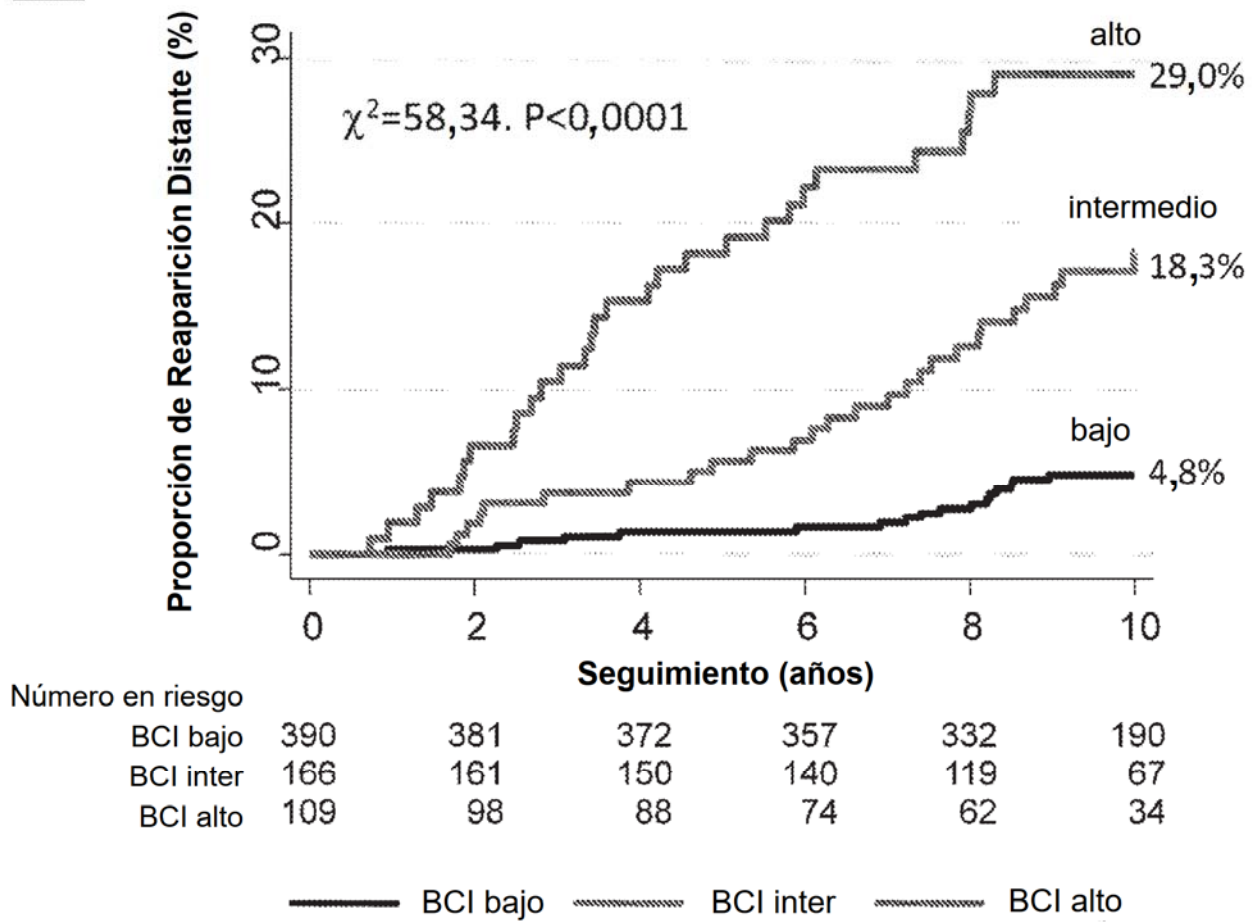


FIG. 9

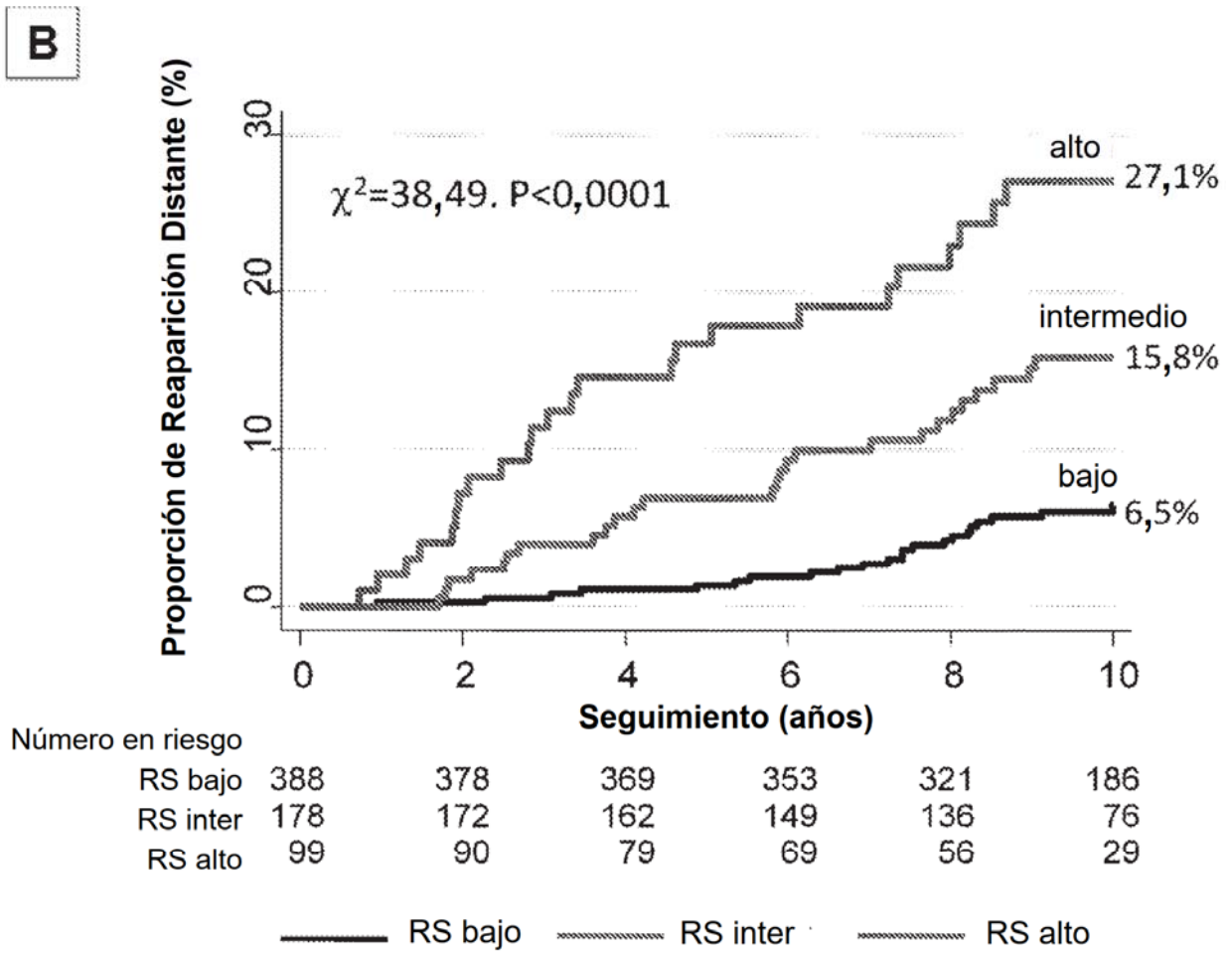


FIG. 9 (cont)

C

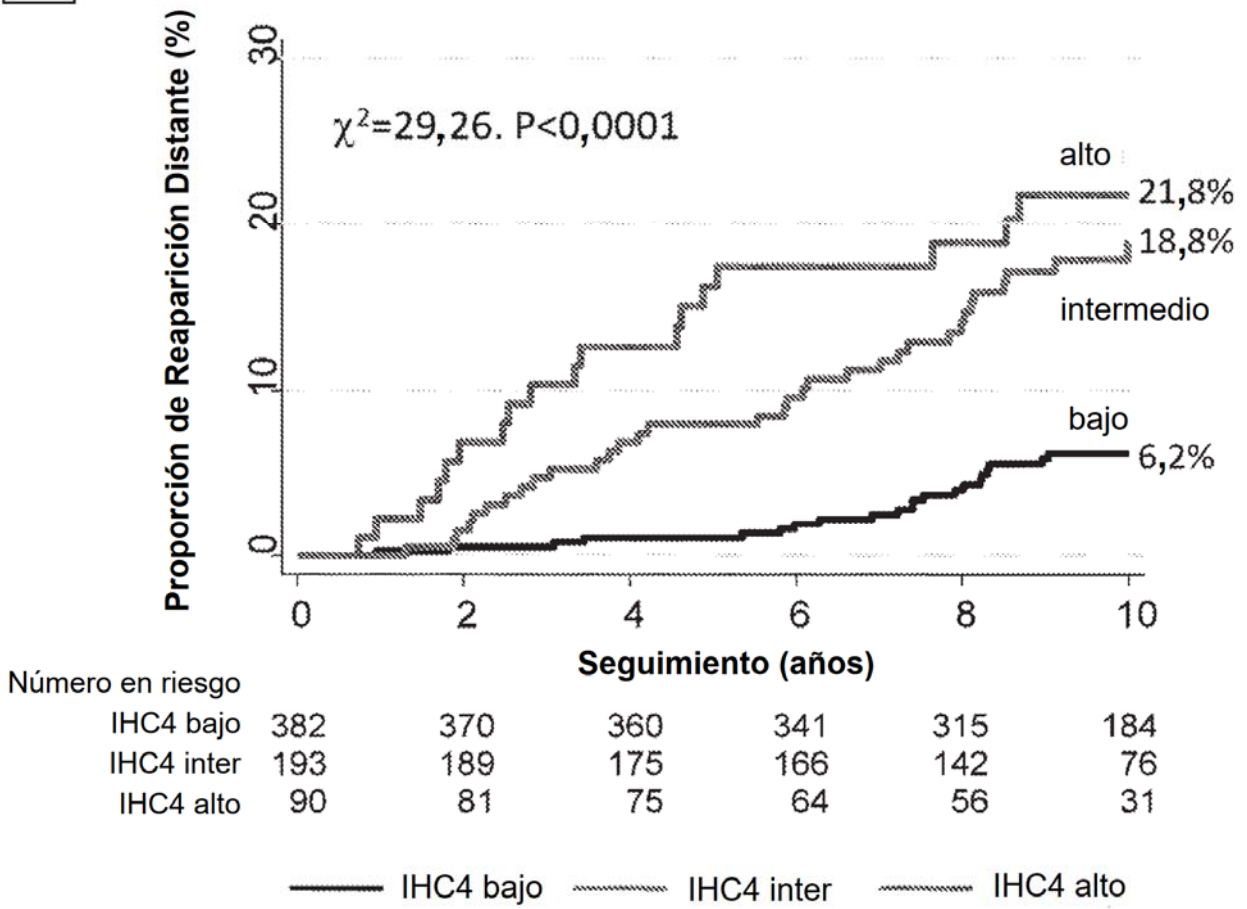


FIG. 9 (cont)

D

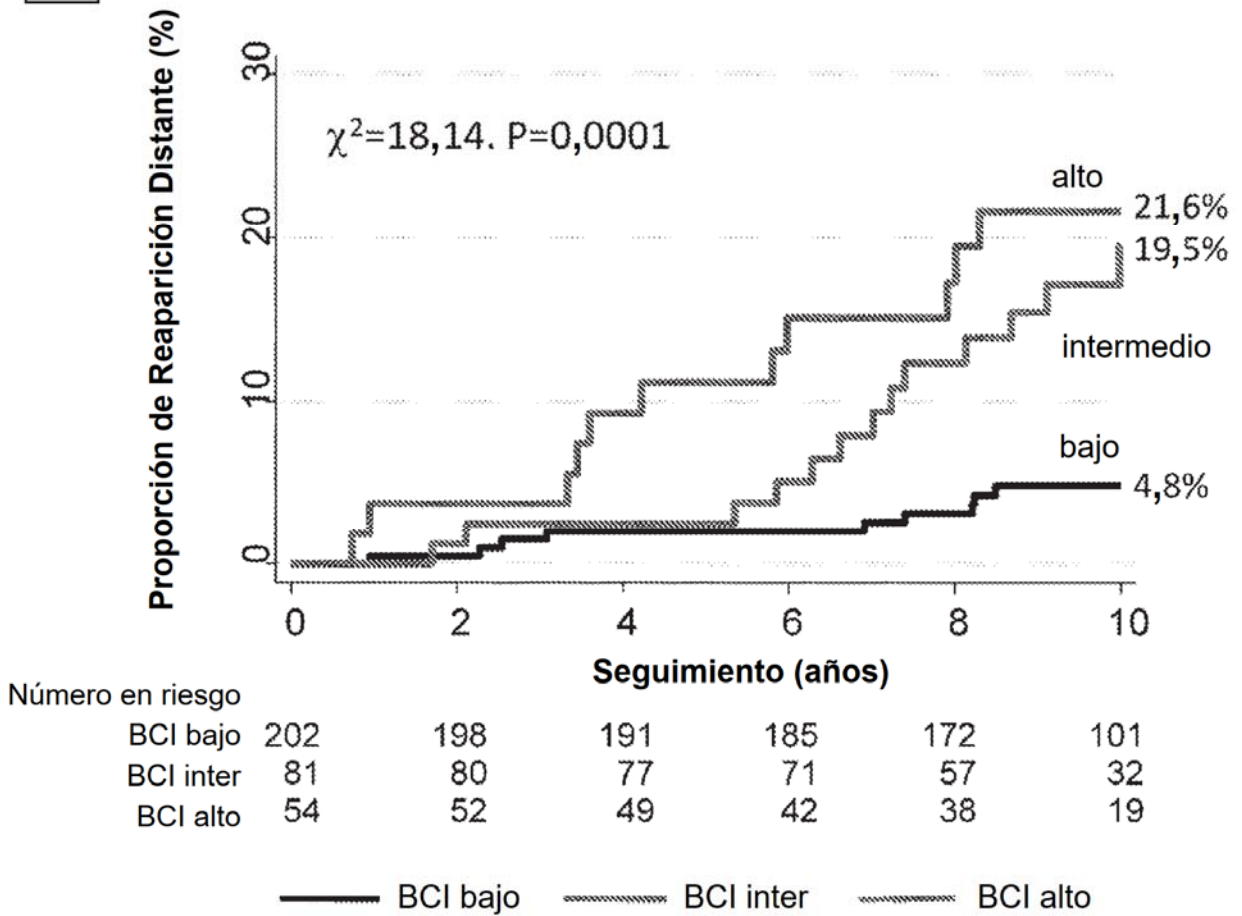


FIG. 9 (cont)

E

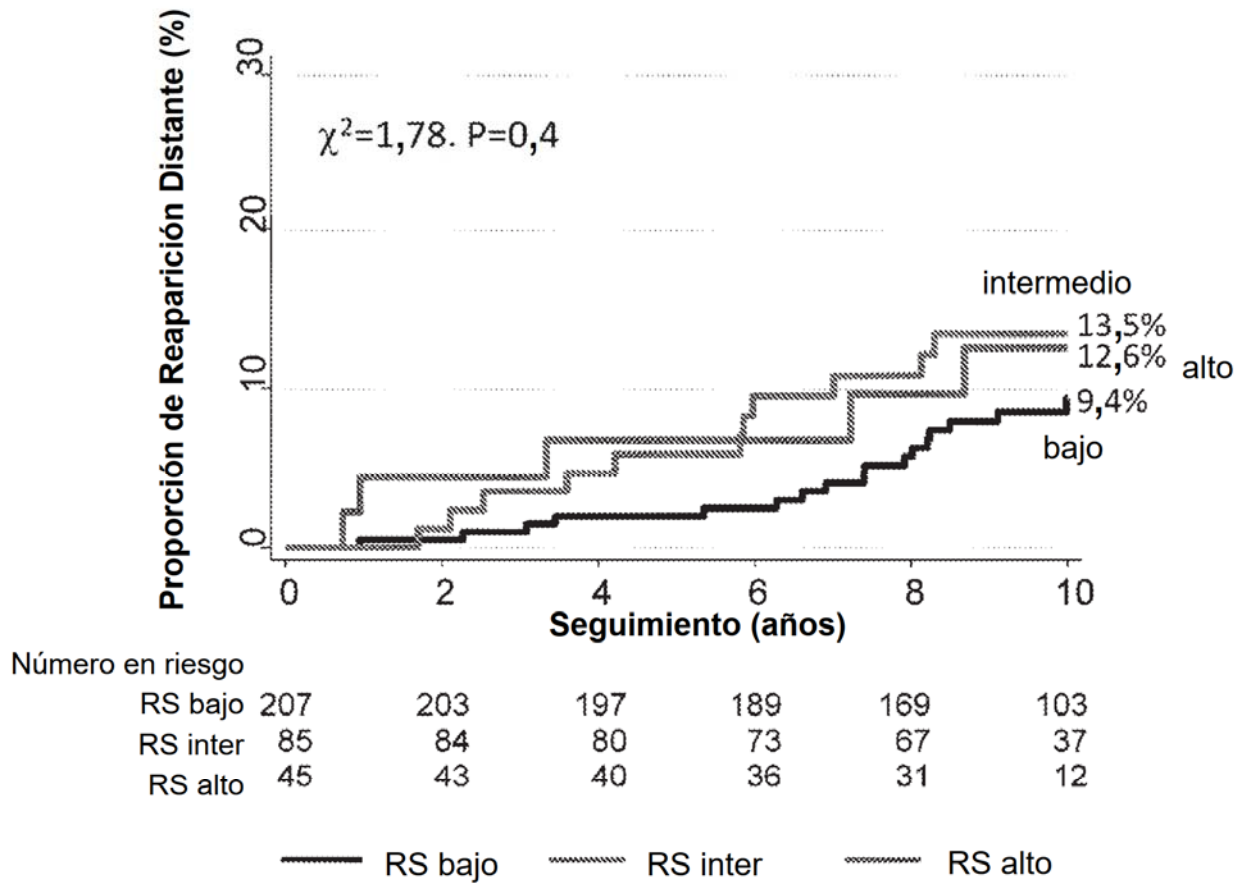


FIG. 9 (cont)

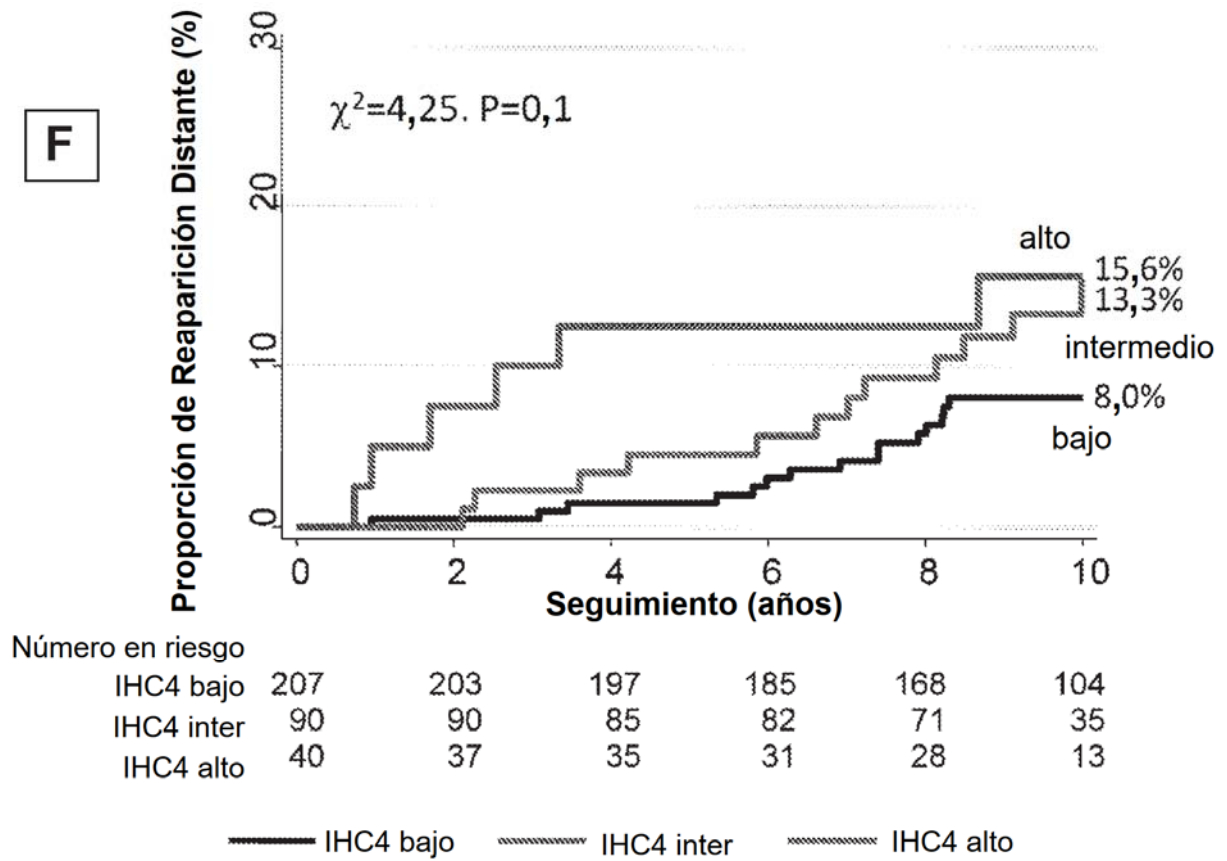


FIG. 9 (cont)

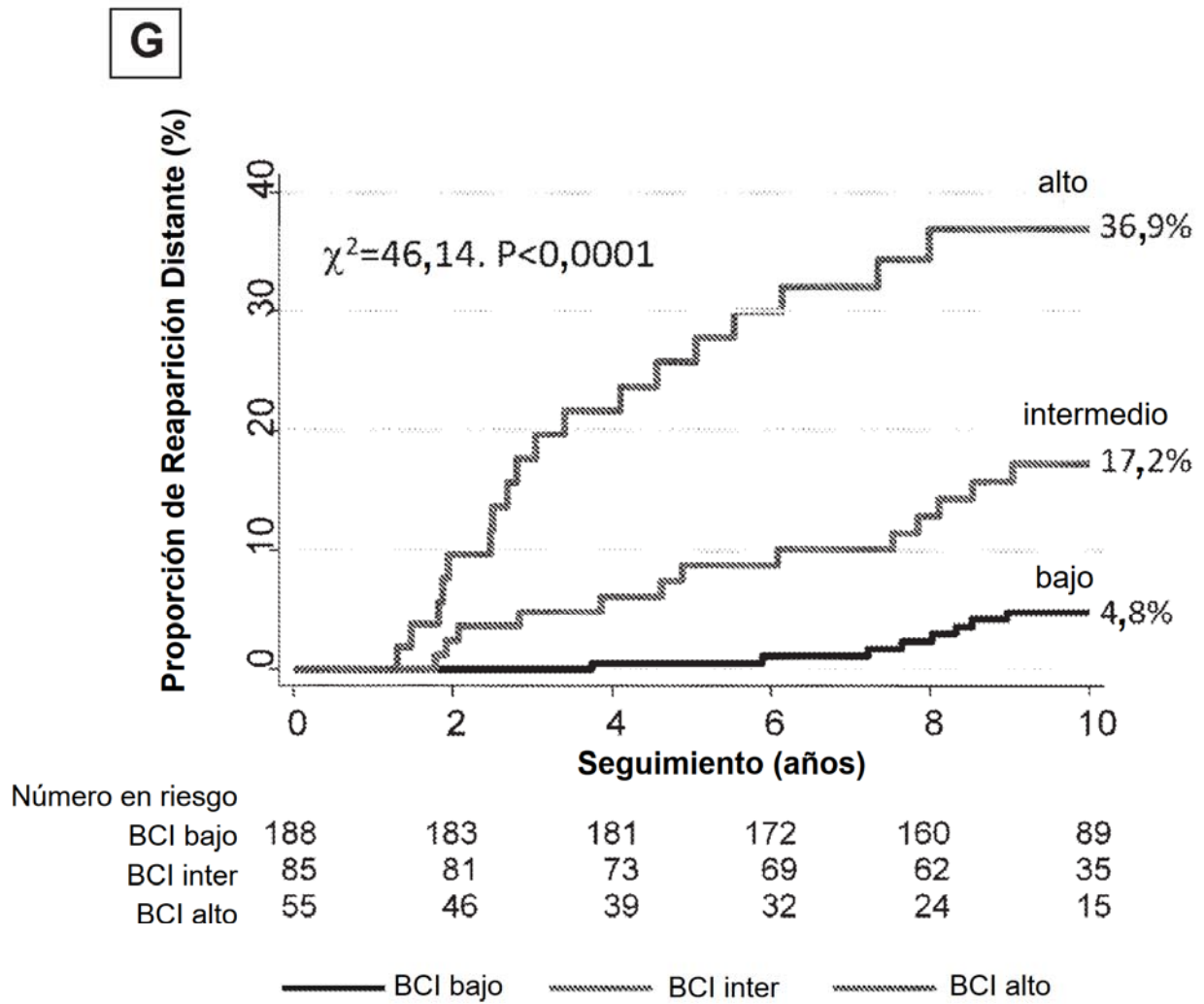


FIG. 9 (cont)

H

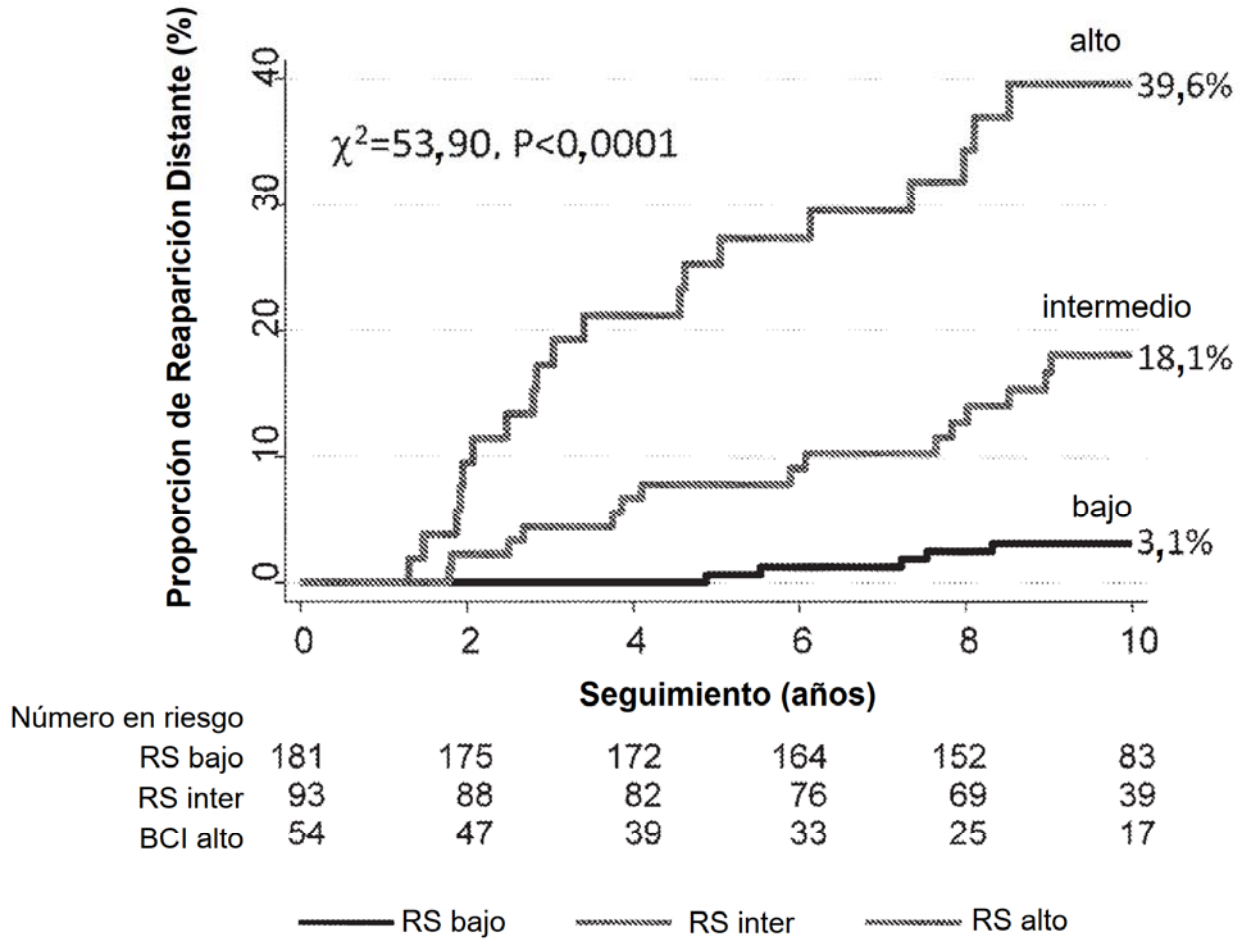


FIG. 9 (cont)

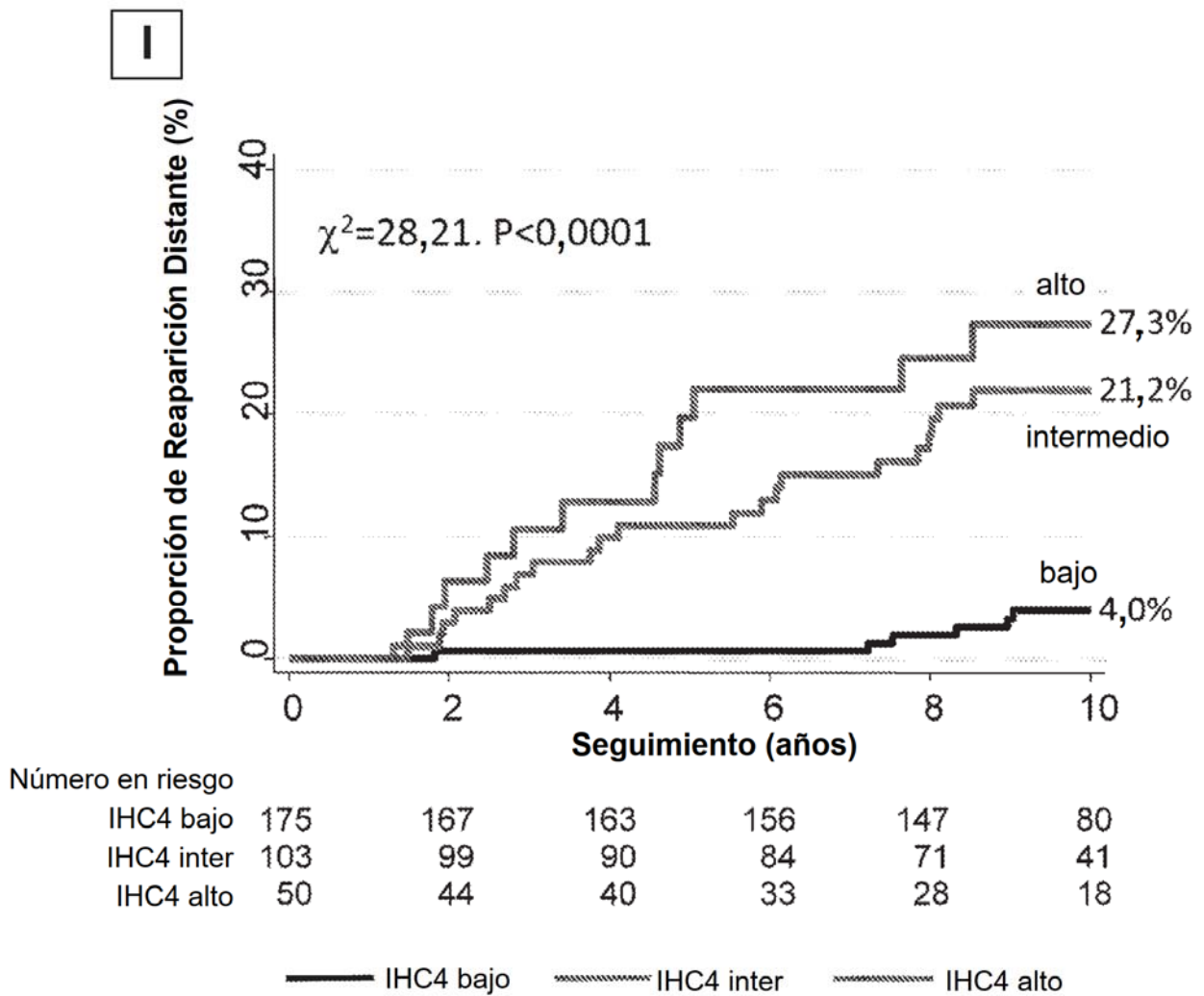


FIG. 9 (cont)

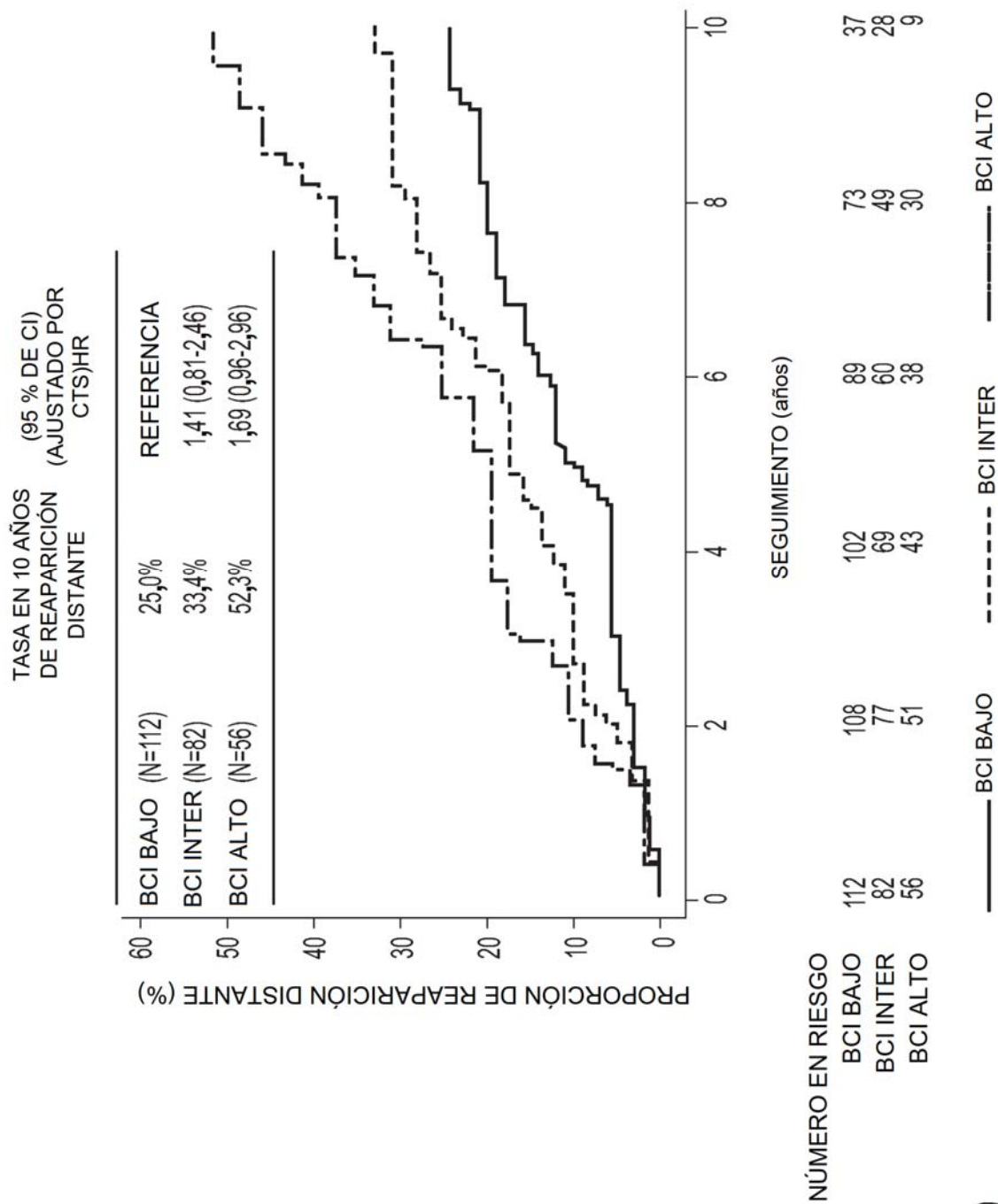


FIG. 10