

MÉTODOS Y CLASIFICADORES GENÓMICOS PARA EL PRONÓSTICO DEL CÁNCER DE MAMA E IDENTIFICACIÓN DE SUJETOS QUE PROBABLEMENTE NO SE BENEFICIEN DE LA RADIOTERAPIA

5 REFERENCIA CRUZADA CON SOLICITUD RELACIONADA

Esta solicitud reivindica prioridad sobre la solicitud de patente provisional de EE. UU. N.º de serie 63/154,821, presentada el 1 de marzo de 2021, que se incorpora aquí como referencia en su totalidad.

10 CAMPO DE LA INVENCION

La presente divulgación se refiere a sistemas y métodos para proporcionar evaluaciones de pronóstico individualizadas de la recurrencia del cáncer de mama (por ejemplo, recurrencia locorregional). Los sistemas y métodos implican la medición de la expresión génica de una muestra de paciente para crear una expresión génica distintiva que identifique a los sujetos que
15 probablemente no se beneficiarán de la radioterapia después de una cirugía de cáncer de mama.

ANTECEDENTES DE LA INVENCION

A pesar de los avances en el tratamiento y diagnóstico del cáncer de mama, sigue siendo necesario mejorar las evaluaciones pronósticas de la recurrencia del cáncer de mama (p. ej., recurrencia
20 locorregional). Además, existe la necesidad de nuevas características genéticas que sean pronósticas de la recurrencia locorregional y la sensibilidad a la radiación.

La presente invención aborda esta necesidad.

COMPENDIO DE LA INVENCION

La presente divulgación se refiere a sistemas y métodos para proporcionar evaluaciones de pronóstico individualizadas de la recurrencia del cáncer de mama (BC) (por ejemplo, recurrencia
25 locorregional (LRR)). Los sistemas y métodos implican la medición de la expresión génica de una muestra de pacientes para crear una característica de expresión génica que identifique a los sujetos que probablemente no se benefician de la radioterapia (RT) tras la cirugía del cáncer de mama (BCS).
30

La presente divulgación se refiere a métodos, sistemas y kits para el diagnóstico, pronóstico y tratamiento de BC en un sujeto. La divulgación también proporciona biomarcadores y clasificadores para identificar sujetos con bajo riesgo de recurrencia del cáncer de mama y que probablemente no se beneficiarán de la radioterapia adyuvante. Además, en ciertos casos, se divulgan en el presente
35 documento conjuntos de sondas para su uso en la detección de dichos biomarcadores para

determinar el riesgo de recurrencia del cáncer de mama en un sujeto. La divulgación proporciona además biomarcadores y clasificadores para identificar sujetos en riesgo de recurrencia locorregional (LRR) y predecir la respuesta a la radioterapia. También se proporcionan métodos de tratamiento del cáncer de mama basados en el perfil de expresión y/o la edad para determinar el riesgo de recurrencia del cáncer de mama.

En determinadas modalidades, la presente invención proporciona métodos que comprenden: a) medir un nivel de expresión de uno o más genes en una muestra biológica de un paciente humano que tiene o está en riesgo de tener cáncer de mama (BC), en donde uno o más genes son seleccionados de AGR2, B4GALT1, CLDN7, EZR, GNG11, JUN, MMP11, PKIB, PRPS1, PSMD10, SH3BP5, SLC16A3, SLC7A11, SPP1, TNNT1 y UBE2E1; y b) determinar una probabilidad de recurrencia de BC para el paciente basándose en el nivel de expresión de uno o más genes seleccionados. En ciertas modalidades, el método puede comprender adicionalmente comparar los niveles de expresión medidos de uno o más genes seleccionados con los niveles de expresión génica predeterminados consistentes con el riesgo de recurrencia de BC o puede comprender normalizar los niveles de expresión de uno o más genes seleccionados, por ejemplo, para producir un nivel de expresión normalizado para uno o más genes seleccionados. En modalidades adicionales, la determinación de la probabilidad de recurrencia de BC puede basarse en el nivel de expresión comparado de uno o más genes seleccionados. En otra modalidad, la determinación de la probabilidad de recurrencia de BC puede basarse en el nivel de expresión normalizado para uno o más genes seleccionados.

En ciertas modalidades, la divulgación proporciona métodos para predecir una probabilidad de recurrencia de BC para un paciente que tiene BC o en riesgo de tener BC que comprende: (a) medir, en una muestra obtenida del paciente, un nivel de expresión de uno o más de los siguientes genes: AGR2, B4GALT1, CLDN7, EZR, GNG11, JUN, MMP11, PKIB, PRPS1, PSMD10, SH3BP5, SLC16A3, SLC7A11, SPP1, TNNT1 y UBE2E1; y (b) predecir una probabilidad de recurrencia de BC para el paciente basándose en el nivel de expresión de uno o más genes, en donde la expresión aumentada de AGR2, CLDN7, EZR, MMP11, PKIB, PRPS1, PSMD10, SLC16A3, SLC7A11, SPP1, TNNT1 y UBE2E1 se correlacionan con un mayor riesgo de recurrencia de BC, y en donde una mayor expresión de B4GALT1, GNG11, JUN y SH3BP5 se correlaciona con un riesgo reducido de recurrencia de BC. En determinadas modalidades, el método puede comprender adicionalmente normalizar el nivel de expresión de uno o más genes seleccionados para obtener un nivel de expresión normalizado para uno o más genes seleccionados. En modalidades adicionales, la probabilidad de recurrencia de BC se predice basándose en el nivel de expresión normalizado para uno o más genes seleccionados.

En algunas modalidades, la recurrencia de BC es recurrencia local o locorregional o recurrencia a

distancia (metástasis). En algunas modalidades, la muestra biológica es una biopsia o una muestra de tumor.

Como se utiliza en el presente documento, "riesgo de recurrencia de BC" o "probabilidad de recurrencia de BC" se refiere a una probabilidad estadística (p. ej., probabilidad) de recurrencia de BC durante un período prolongado de tiempo (p. ej., 1 mes, 1 año, 5 años, 10 años, etc.). En algunas modalidades, el riesgo de recurrencia de BC implica un valor inicial en donde un paciente ha tenido una intervención exitosa (por ejemplo, intervención quirúrgica) y se caracteriza por no tener BC y/o células cancerosas que progresan activamente. Como tal, un riesgo o probabilidad de recurrencia implica la probabilidad de que el cáncer recurra de alguna manera. Dichos riesgos pueden caracterizarse como riesgo muy bajo, riesgo bajo, riesgo moderadamente bajo, riesgo promedio, riesgo moderadamente alto, riesgo alto y riesgo muy alto.

En algunas modalidades, los niveles de expresión aumentados de AGR2, CLDN7, EZR, MMP11, PKIB, PRPS1, PSMD10, SLC16A3, SLC7A11, SPP1, TNNT1 y UBE2E1 se correlacionan con un mayor riesgo o probabilidad de recurrencia del cáncer de mama.

En algunas modalidades, los niveles de expresión disminuidos de AGR2, CLDN7, EZR, MMP11, PKIB, PRPS1, PSMD10, SLC16A3, SLC7A11, SPP1, TNNT1 y UBE2E1 se correlacionan cada uno con un menor riesgo o probabilidad de una recurrencia del cáncer de mama.

En algunas modalidades, los niveles de expresión aumentados de B4GALT1, GNG11, JUN y SH3BP5 se correlacionan con un menor riesgo o probabilidad de recurrencia del cáncer de mama.

En algunas modalidades, la disminución de los niveles de expresión de B4GALT1, GNG11, JUN y SH3BP5 se correlaciona con un aumento o probabilidad de riesgo de recurrencia del cáncer de mama.

En algunas modalidades, los métodos comprenden además tratar al paciente con radioterapia adyuvante si el paciente se caracteriza por tener un riesgo elevado o alto de recurrencia del BC. En ciertas modalidades, un paciente puede ser tratado con radioterapia adyuvante si el paciente se caracteriza por un riesgo moderadamente alto, alto o muy alto de recurrencia del BC. En algunas modalidades, los métodos comprenden además no tratar al paciente con radioterapia adyuvante si el paciente se caracteriza por tener un riesgo o probabilidad reducido o bajo de recurrencia del BC. En determinadas modalidades, un paciente puede no ser tratado con radioterapia adyuvante o ser identificado como un paciente que no se beneficiaría de la radioterapia adyuvante si el paciente se caracteriza por tener un riesgo moderadamente bajo, un riesgo bajo o un riesgo muy bajo de recurrencia de BC.

En algunas modalidades, los niveles de expresión de todos los genes se miden en la muestra biológica. En algunas modalidades, el número de uno o más genes se selecciona del grupo que consiste en: 1 gen, 1-2 genes, 1-3, 1-4, 1-5, 1-6, 1-7, 1-8, 1-9, 1-10, 1-11, 1-12, 1-13, 1-14, 1-15, 1-

16, 2-3, 2-4, 2-5, 2-6, 2-7, 2-8, 2-9, 2-10, 2-11, 2-12, 2-13, 2-14, 2-15, 2-16, 3-4, 3-5, 3-6, 3-7, 3-8, 3-9, 3-10, 3-11, 3-12, 3-13, 3-14, 3-15, 3-16, 4-5, 4-6, 4-7, 4-8, 4-9, 4-10, 4-11, 4-12, 4-13, 4-14, 4-15, 4-16, 5-6, 5-7, 5-8, 5-9, 5-10, 5-11, 5-12, 5-13, 5-14, 5-15, 5-16, 6-7, 6-8, 6-9, 6-10, 6-11, 6-12, 6-13, 6-14, 6-15, 6-16, 7-8, 7-9, 7-10, 7-11, 7-12, 7-13, 7-14, 7-15, 7-16, 8-9, 8-10, 8-11, 8-12, 8-13, 8-14, 8-15, 8-16, 9-10, 9-11, 9-12, 9-13, 9-14, 9-15, 9-16, 10-11, 10-12, 10-13, 10-14, 10-15, 10-16, 11-12, 11-13, 11-14, 11-15, 11-16, 12-13, 12-14, 12-15, 12-16, 13-14, 13-15, 13-16, 14-15, 14-16, 15-16, y 16 genes.

En algunas modalidades, medir los niveles de expresión comprende realizar uno o más de: hibridación *in situ*, un método basado en PCR, un método basado en matrices, un método inmunohistoquímico, un método de ensayo de ARN o un método de inmunoensayo. En algunas modalidades, medir los niveles de expresión comprende usar un reactivo seleccionado del grupo que consiste en una sonda de ácido nucleico, uno o más cebadores de ácido nucleico y un anticuerpo. En algunas modalidades, medir el nivel de expresión comprende medir el nivel de un transcrito de ARN.

En ciertas modalidades, la divulgación proporciona un método para pronosticar y/o predecir el beneficio de la radioterapia adyuvante en un sujeto que tiene BC, el método comprende: a) obtener o haber obtenido un nivel de expresión en una muestra de un sujeto durante uno o más genes seleccionados de AGR2, B4GALT1, CLDN7, EZR, GNG11, JUN, MMP11, PKIB, PRPS1, PSMD10, SH3BP5, SLC16A3, SLC7A11, SPP1, TNNT1 y UBE2E1; y b) determinar que el sujeto tiene un riesgo bajo o alto de recurrencia del cáncer en función del nivel de expresión, y/o es probable que se beneficie o no es probable que se beneficie de la radioterapia adyuvante en función del nivel de expresión, pronosticando y/o prediciendo de ese modo el beneficio de Radioterapia adyuvante en el sujeto. En algunas modalidades, el método comprende además suspender la terapia de radioterapia adyuvante si se identifica que el sujeto no se beneficiará probablemente de la radioterapia adyuvante y/o administrar un tratamiento contra el cáncer distinto de la radioterapia adyuvante. En otras modalidades, la recurrencia del cáncer es recurrencia local o locorregional o recurrencia a distancia (metástasis). En algunas modalidades, los niveles de expresión de uno o más genes pueden aumentar o disminuir en comparación con un control. En algunas modalidades, los niveles de expresión de todos de uno o más genes se miden en la muestra biológica. En algunas modalidades, el método comprende además tratar al sujeto con radioterapia adyuvante. En algunas modalidades, el método comprende además tratar al sujeto con una terapia contra el cáncer distinta de la radioterapia adyuvante. En otras modalidades más, el método comprende además tratar al sujeto con mastectomía, refuerzo de radiación o terapia sistémica adyuvante. En algunas modalidades, la radioterapia se suspende del sujeto después de la cirugía de conservación de la mama (BCS). En algunas modalidades, el método comprende además determinar que el sujeto tiene un riesgo bajo

de recurrencia del cáncer en función de la edad del sujeto, o determinar que el sujeto no tiene un riesgo bajo de recurrencia del cáncer en función de la edad del sujeto. En otras modalidades, la recurrencia del cáncer es recurrencia local o locorregional o recurrencia a distancia (metástasis). En otras modalidades, el sujeto tiene un alto riesgo de recurrencia del cáncer y se trata con radioterapia.

5 En algunas modalidades, la divulgación proporciona un método que comprende: a) obtener o haber obtenido un nivel de expresión en una muestra de un sujeto para uno o más genes seleccionados de AGR2, B4GALT1, CLDN7, EZR, GNG11, JUN, MMP11, PKIB, PRPS1, PSMD10, SH3BP5, SLC16A3, SLC7A11, SPP1, TNNT1 y UBE2E1; y b) determinar que el sujeto tiene un riesgo bajo de recurrencia del cáncer y no es probable que se beneficie del tratamiento con radioterapia adyuvante en función del nivel de expresión, o determinar que el sujeto no tiene un riesgo alto de recurrencia del cáncer y es probable que se beneficie del tratamiento con radioterapia adyuvante según el nivel de expresión. En otras modalidades, el método comprende además administrar una terapia contra el cáncer distinta de la terapia de radioterapia adyuvante si se identifica que el sujeto no es probable que se beneficie de la radioterapia adyuvante. En otras modalidades, la recurrencia del cáncer es recurrencia local o locorregional o recurrencia a distancia (metástasis). En algunas modalidades, los niveles de expresión de uno o más genes pueden aumentar o disminuir en comparación con un control. En algunas modalidades, los niveles de expresión de todos los genes seleccionados de la Tabla 5 se miden en la muestra biológica. En aun otras modalidades, el método comprende además tratar al sujeto con mastectomía, refuerzo de radiación o terapia sistémica adyuvante. En algunas modalidades, la radioterapia se suspende del sujeto después de la cirugía de conservación de la mama (BCS). En algunas modalidades, el método comprende además determinar que el sujeto tiene un riesgo bajo de recurrencia del cáncer en función de la edad del sujeto, o determinar que el sujeto no tiene un riesgo bajo de recurrencia del cáncer en función de la edad del sujeto. En algunas modalidades, la recurrencia del cáncer es recurrencia local o locorregional o recurrencia a distancia (metástasis).

25 En algunas modalidades, la descripción proporciona un método para tratar el cáncer de mama en un sujeto, que comprende: a) obtener o haber obtenido un nivel de expresión en una muestra de un sujeto para uno o más genes seleccionados de AGR2, B4GALT1, CLDN7, EZR, GNG11, JUN, MMP11, PKIB, PRPS1, PSMD10, SH3BP5, SLC16A3, SLC7A11, SPP1, TNNT1 y UBE2E1; b) 30 determinar que el sujeto tiene un bajo riesgo de recurrencia del cáncer y no es probable que se beneficie del tratamiento con radioterapia adyuvante en función del nivel de expresión; y c) administrar un tratamiento contra el cáncer distinto de la radioterapia adyuvante si se identifica que el sujeto no es probable que se beneficie de la radioterapia adyuvante en función del nivel de expresión. En otras modalidades, la recurrencia del cáncer es recurrencia local o locorregional o recurrencia a distancia (metástasis). En algunas modalidades, los niveles de expresión de uno o 35

más de los genes pueden aumentar o disminuir en comparación con un control. En algunas modalidades, los niveles de expresión de todos los genes se miden en la muestra biológica. En otras modalidades más, el método comprende además tratar al sujeto con mastectomía, refuerzo de radiación o terapia sistémica adyuvante. En algunas modalidades, la radioterapia se suspende del sujeto después de la cirugía de conservación de la mama (BCS). En algunas modalidades, el método comprende además determinar que el sujeto tiene un riesgo bajo de recurrencia del cáncer en función de la edad del sujeto, o determinar que el sujeto no tiene un riesgo bajo de recurrencia del cáncer en función de la edad del sujeto. En otras modalidades, la recurrencia del cáncer es recurrencia local o locorregional o recurrencia a distancia (metástasis).

En algunas modalidades, la pluralidad de genes usados en los métodos y clasificadores genómicos de la presente divulgación se seleccionan del grupo que consiste en gradiente anterior 2, proteína disulfuro isomerasa (AGR2), beta-1,4-galactosiltransferasa 1 (B4GALT1), Claudin 7 (CLDN7), Ezrin (EZR), Subunidad de proteína G Gamma 11 (GNG11), Protooncogén Jun (JUN), Matriz metalopeptidasa 11 (MMP11), Inhibidor beta de la proteína quinasa dependiente de CAMP (PKIB), Fosforribosilpirofosfato Sintetasa 1 (PRPS1), subunidad del proteasoma 26S, no ATPasa 10 (PSMD10), proteína de unión al dominio SH3 5 (SH3BP5), miembro 3 de la familia de portadores de solutos 16 (SLC16A3), miembro 11 de la familia de portadores de solutos 7 (SLC7A11), fosfoproteína 1 secretada (SPP1), troponina T1, tipo esquelético lento (TNNT1) y enzima conjugadora de ubiquitina E2 E1 (UBE2E1).

En determinadas modalidades, el sujeto tiene cáncer de mama con receptor de estrógeno positivo (ER+), cáncer de mama con receptor 2 del factor de crecimiento epidérmico humano negativo (HER2-), cáncer de mama en estadio I-II o cáncer de mama con ganglios negativos y/o es posmenopáusica.

Dichos métodos no se limitan a una muestra específica o un tipo de muestra biológica. Por ejemplo, en algunas modalidades la muestra o muestra biológica es una muestra de tejido, muestra de fluido corporal, muestra de sangre, muestra de secreción de órganos, muestra de CSF, muestra de saliva, muestra de plasma, muestra de suero o muestra de orina. En algunas modalidades, la muestra puede comprender tejido mamario o tejido circundante, una biopsia de mama, una muestra de tumor o tejido que contiene células mamarias o células de cáncer de mama.

En determinadas modalidades, los ácidos nucleicos que comprenden secuencias de genes seleccionados de la Tabla 5, o complementos de los mismos, se aíslan de la muestra biológica y/o se purifican y/o amplifican antes del análisis. En algunas modalidades, los ácidos nucleicos pueden comprender transcritos de ARN.

En otras modalidades, los niveles de expresión de biomarcadores se determinan mediante hibridación *in situ*, métodos basados en PCR, métodos basados en matrices, métodos

inmunohistoquímicos, métodos de ensayo de ARN o métodos de inmunoensayo. En otras modalidades, los niveles de expresión génica se determinan usando uno o más reactivos. En determinadas modalidades, uno o más reactivos son sondas de ácido nucleico, cebadores de ácido nucleico y/o anticuerpos. En otras modalidades, determinar el nivel de expresión de un biomarcador comprende medir el nivel de un ácido nucleico. En algunas modalidades, el ácido nucleico es un transcrito de ARN.

En algunas modalidades, el nivel de expresión de al menos un gen se reduce en comparación con un control. En otras modalidades, el nivel de expresión de al menos un gen aumenta en comparación con un control.

En determinadas modalidades, los métodos descritos en el presente documento se realizan antes del tratamiento del sujeto con radioterapia adyuvante. En determinadas modalidades, los métodos descritos en el presente documento se realizan antes del tratamiento del sujeto con mastectomía, refuerzo de radiación o terapia sistémica adyuvante.

En otras modalidades, el método comprende además calcular una puntuación de riesgo para el sujeto, en donde se retiene la radioterapia adyuvante del sujeto si se identifica que el sujeto tiene un riesgo bajo de recurrencia del cáncer y que no es probable que se beneficie de la radioterapia adyuvante basándose tanto en el riesgo puntuación y los niveles de expresión de uno o más genes seleccionados de AGR2, B4GALT1, CLDN7, EZR, GNG11, JUN, MMP11, PKIB, PRPS1, PSMD10, SH3BP5, SLC16A3, SLC7A11, SPP1, TNNT1 y UBE2E1 en la muestra biológica, y administrar al sujeto una terapia contra el cáncer distinta de la radioterapia adyuvante. En otras modalidades, la recurrencia del cáncer es recurrencia local o locorregional o recurrencia a distancia (metástasis).

La importancia de los niveles de expresión de uno o más genes biomarcadores se puede evaluar usando, por ejemplo, una prueba T, valor P, valor P KS (Kolmogorov Smirnov), exactitud, valor P de exactitud, valor predictivo positivo (PPV), valor predictivo negativo (VPN), sensibilidad, especificidad, AUC, valor P de AUC (valor Auc.p), valor P de la prueba de Wilcoxon, diferencia de pliegues de la mediana (MFD), curvas de Kaplan Meier (KM), AUC de supervivencia (survAUC), valor P de Kaplan Meier (valor P de KM), valor P del índice de probabilidades del análisis univariable (uvaORPval), valor P del índice de probabilidades del análisis multivariable (mvaORPval), valor P del índice de riesgo del análisis univariable (uvaHRPval) y índice de riesgo del análisis multivariable Valor p (mvaHRPval). La importancia del nivel de expresión de una o más dianas puede basarse en dos o más métricas seleccionadas del grupo que comprende AUC, valor P de AUC (valor Auc.p), valor P de la prueba de Wilcoxon, diferencia de pliegues de mediana (MFD), Curvas de Kaplan Meier (KM), supervivencia AUC (survAUC), análisis univariable Odds Ratio valor P (uvaORPval), análisis multivariable Odds Ratio valor P (mvaORPval), Kaplan Meier valor P (KM P-valor), análisis univariable Riesgo Valor P de relación (uvaHRPval) o Valor P de relación de riesgo de análisis

multivariable (mvaHRPval).

En otro aspecto, la divulgación incluye un conjunto de sondas para determinar un pronóstico de un sujeto que tiene BC y si se debe tratar o no al sujeto con radioterapia, el conjunto de sondas comprende una pluralidad de sondas para detectar una pluralidad de ácidos nucleicos diana, en donde la pluralidad de ácidos nucleicos diana comprende una o más secuencias genéticas, o complementos de las mismas, de uno o más genes seleccionados de AGR2, B4GALT1, CLDN7, EZR, GNG11, JUN, MMP11, PKIB, PRPS1, PSMD10, SH3BP5, SLC16A3, SLC7A11, SPP1, TNNT1 y UBE2E1. Las sondas pueden estar marcadas de forma detectable para facilitar la detección. En algunas modalidades, el pronóstico comprende un pronóstico de recurrencia del cáncer. En modalidades adicionales, la recurrencia del cáncer es recurrencia local o locorregional o recurrencia a distancia (metástasis).

En otro aspecto, la divulgación incluye un sistema para determinar un pronóstico de un sujeto que tiene BC y si tratar o no al sujeto con radioterapia, el sistema comprende: a) un conjunto de sondas descrito en el presente documento; y b) un modelo informático o algoritmo para analizar un nivel de expresión o perfil de expresión de la pluralidad de ácidos nucleicos diana hibridados con la pluralidad de sondas en una muestra biológica de un sujeto que tiene BC y determinar si el sujeto tiene un riesgo bajo de recurrencia del cáncer basándose en el nivel de expresión o en el perfil de expresión y debe ser tratado con radioterapia. En otras modalidades, la recurrencia del cáncer es recurrencia local o locorregional o recurrencia a distancia (metástasis).

En algunas modalidades, la divulgación incluye un kit para determinar un pronóstico de un sujeto que tiene cáncer de mama y si tratar o no al sujeto con radioterapia adyuvante, el kit comprende agentes para medir los niveles de expresión de uno o más genes seleccionados de AGR2, B4GALT1, CLDN7, EZR, GNG11, JUN, MMP11, PKIB, PRPS1, PSMD10, SH3BP5, SLC16A3, SLC7A11, SPP1, TNNT1 y UBE2E1. En otras modalidades, el kit puede incluir uno o más agentes (por ejemplo, sondas de hibridación, cebadores de PCR o micromatriz) para medir niveles de expresión de una pluralidad de genes, un recipiente para contener una muestra biológica que comprende células de cáncer de mama aisladas de un ser humano. sujeto para prueba, y/o instrucciones impresas para hacer reaccionar los agentes con la muestra biológica o una porción de la muestra biológica para determinar si el sujeto tiene un riesgo bajo de recurrencia del cáncer de mama y es probable que se beneficie del tratamiento con radioterapia adyuvante. En otras modalidades, la recurrencia del cáncer es recurrencia local o locorregional o recurrencia a distancia (metástasis). En otras modalidades, los agentes se envasan en contenedores separados. En otras modalidades más, el kit comprende además una o más muestras de referencia de control u otros reactivos para medir la expresión génica (por ejemplo, reactivos para realizar PCR, RT-PCR, análisis de micromatrices, una transferencia Northern, un inmunoensayo o inmunohistoquímica). En otra

modalidad, el kit comprende agentes para medir los niveles de expresión de todos los siguientes genes: AGR2, B4GALT1, CLDN7, EZR, GNG11, JUN, MMP11, PKIB, PRPS1, PSMD10, SH3BP5, SLC16A3, SLC7A11, SPP1, TNNT1 y UBE2E1. En ciertas modalidades, el kit comprende un conjunto de sondas, como se describe en el presente documento, para detectar una pluralidad de ácidos nucleicos diana, en donde la pluralidad de ácidos nucleicos diana comprende una o más secuencias genéticas, o complementos de las mismas, de uno o más genes seleccionados de AGR2, B4GALT1, CLDN7, EZR, GNG11, JUN, MMP11, PKIB, PRPS1, PSMD10, SH3BP5, SLC16A3, SLC7A11, SPP1, TNNT1 y UBE2E1, o cualquier combinación de los mismos.

En otras modalidades, el kit comprende además un sistema, en donde el sistema comprende: a) un conjunto de sondas que comprende una pluralidad de sondas para detectar una pluralidad de ácidos nucleicos diana, en donde la pluralidad de ácidos nucleicos diana comprende una o más secuencias genéticas, o complementos de los mismos, de uno o más genes seleccionados de AGR2, B4GALT1, CLDN7, EZR, GNG11, JUN, MMP11, PKIB, PRPS1, PSMD10, SH3BP5, SLC16A3, SLC7A11, SPP1, TNNT1 y UBE2E1; y b) un modelo informático o algoritmo para analizar un nivel de expresión o perfil de expresión de la pluralidad de ácidos nucleicos diana hibridados con la pluralidad de sondas en una muestra biológica de un sujeto que tiene cáncer de mama y determinar si el sujeto tiene un riesgo bajo de recurrencia del cáncer según el nivel de expresión o el perfil de expresión y no es probable que se beneficien del tratamiento con radioterapia adyuvante. En otras modalidades, la recurrencia del cáncer es recurrencia local o locorregional o recurrencia a distancia (metástasis). Estas y otras modalidades de la presente divulgación se les ocurrirán fácilmente a los expertos en la técnica en vista de la presente divulgación.

BREVE DESCRIPCIÓN DE LOS DIBUJOS

FIG. 1: Diagrama de selección de cohortes de formación y validación en SweBCG91-RT.

FIG. 2: Diagrama para la selección de la cohorte de validación de Princess Margaret.

FIG. 3: Diagrama de selección de genes a incluir en el modelo.

FIG. 4: Incidencia acumulada de recurrencia locorregional con o sin radioterapia (RT) adyuvante en la cohorte de validación SweBCG91-RT para pacientes clasificados por POLAR como de bajo riesgo (A) o alto riesgo (B). Los índices de riesgo y los valores p se calculan utilizando un modelo de regresión de riesgos proporcionales de Cox específico para la causa.

FIG. 5: Incidencia acumulada de recurrencia locorregional en la cohorte Princess Margaret con o sin radioterapia adyuvante (RT) para pacientes clasificados por POLAR como de bajo riesgo (A) o de alto riesgo (B). Los índices de riesgo y los valores p se calculan utilizando un modelo de regresión de riesgos proporcionales de Cox específico para la causa.

DESCRIPCIÓN DETALLADA DE LA INVENCIÓN

Múltiples ensayos clínicos aleatorizados de fase III han demostrado consistentemente el beneficio de la radioterapia (RT) de mama completa después de la cirugía conservadora de mama (BCS) para reducir la recurrencia locorregional (LRR), y la RT se considera el estándar de atención para las mujeres con cáncer de mama invasivo en estadio temprano (consulte, Malmström, P. et al. European journal of cancer (Oxford, Inglaterra: 1990) 39, 1690-1697 (2003); Fyles, A. W. et al. The New England journal of medicine 351, 963-970 (2004); Holli, K., et al., Br J Cancer 84, 164-169 (2001); Forrest, A. P. et al. Lancet 348, 708-713 (1996); Fisher, B. et al. Journal of clinical oncology: Official journal of the American Society of Clinical Oncology 20, 4141-4149 (2002); Liljegren, G. et al. J Clin Oncol 17, 2326-2333 (1999); Veronesi, U. et al. Annals of oncology: official journal of the European Society for Medical Oncology / ESMO 12, 997-100 (2001); Clark, R. M. et al. J Natl Cancer Inst 88, 1659-1664 (1996)). El metanálisis del Oxford Early Breast Cancer Trialists' Group (EBCTCG) demostró una reducción del riesgo relativo de recurrencia local de dos tercios a los 10 años para pacientes que reciben RT, del 30 % al 10 %, lo que se tradujo en un beneficio de supervivencia de ~ 5 % a los 15 años (consulte, Clarke, M. et al. Lancet 366, 2087-2106 (2005); Early Breast Cancer Trialists' Collaborative, Lancet 378, 1707-1716 (2011)). Además de la mejora en el control de la enfermedad locorregional, estos datos también demostraron la heterogeneidad en la respuesta y el beneficio de la radiación adyuvante. Incluso en la era anterior de quimioterapia sistémica menos efectiva y uso limitado de terapia endocrina, hasta el 70 % de las mujeres no recurrieron localmente sin RT, pero todavía faltan herramientas clínicas para identificar qué mujeres pueden renunciar de manera segura a la radioterapia adyuvante.

Los factores de riesgo clínicos como la edad avanzada, el tamaño del tumor más pequeño y los cánceres de mama con receptores de estrógeno positivos (ER+) se han asociado con un riesgo bajo de LRR (consulte, Fredriksson, I. et al. The British Journal of Surgery 90, 1093-1102 (2003)). Los intentos de utilizar estas características para identificar a las mujeres que pueden omitir la radioterapia después de la BCS han tenido resultados mixtos. En el ensayo CALGB 9343 se inscribieron mujeres de 70 años o más con cánceres de mama pequeños ER+ resecados con márgenes negativos que luego fueron asignadas al azar para recibir RT de mama completa y tamoxifeno durante 5 años o solo tamoxifeno (consulte, Hughes, K. S. et al. Journal of clinical oncology: official journal of the American Society of Clinical Oncology 31, 2382-2387 (2013)). En este ensayo de más de 600 mujeres, la RT disminuyó el riesgo de LRR a los 10 años del 10 % al 2 % sin afectar la tasa de preservación mamaria. De manera similar, el ensayo PRIME II incluyó pacientes de 65 años o más con cánceres de mama ER+ pequeños resecados con márgenes negativos que fueron tratados con terapia endocrina y asignados al azar a RT o no. Los resultados de 10 años informados recientemente demostraron una disminución en IBTR del 9,8 % al 0,9 % con la adición

de RT (consulte, Kunkler, I.H. et al. *The lancet oncology* 16, 266-273 (2015)). Como la tasa de recurrencia local a los 10 años para aquellos que no fueron tratados con RT fue del 10 % en ambos ensayos, se puede argumentar que la reducción del riesgo relativo del 80-90 % a los 10 años con la adición de radioterapia puede ser demasiado significativa para omitir RT en mujeres que tienen una
 5 esperanza de vida media superior a 80 años en los Estados Unidos. Hasta el momento, no se ha incorporado ningún subtipo o variable clínico-patológica en la práctica clínica que prediga la falta de beneficio de la radiación mamaria adyuvante, lo que hace que las decisiones de omisión de radiación sean un desafío clínico (consulte, Killander, F. et al. *European Journal of Cancer* (Oxford, Inglaterra: 1990) 67, 57-65 (2016); Sjöström, M. et al. *Journal of Clinical Oncology* 35, 3222-3229 (2017); Liu, FF et al. *Journal of clinical oncology: Official journal of the American Society of Clinical Oncology* 33, 2035-2040 (2015)).
 10

El interés en el uso de estratificadores de riesgo basados en información genómica ha aumentado en los últimos años dado el éxito de dichos enfoques para determinar la probabilidad de beneficio de la quimioterapia sistémica en mujeres con cáncer de mama. Se ha demostrado que dichas
 15 pruebas con información molecular pronostican el resultado en mujeres con cáncer de mama y/o predicen el beneficio de la quimioterapia en ensayos clínicos anteriores (consulte, Albain, K. S. et al. *The Lancet Oncology* 11, 55-65, (2010); Cardoso, F. et al. *New England Journal of Medicine* 375, 717-729, (2016); Dowsett, M. et al. *Journal of Clinical Oncology* 28, 1829-1834 (2010); Dowsett, M. et al. *Journal of Clinical Oncology* 31, 2783-2790 (2013); Paik, S. et al. *The New England journal of medicine* 351, 2817-2826, (2004); Sparano, J. A. et al. *New England Journal of Medicine* 373, 2005-2014 (2015)). Si bien se han realizado intentos previos de crear características de recurrencia local o sensibilidad a la radiación, no existen características ampliamente utilizadas para seleccionar
 20 pacientes que pueden omitir la radiación en el cáncer de mama invasivo en etapa temprana (véase, Cui, Y., Li, B., Pollom, E. L., Horst, K. C. & Li, R. *Clinical cancer research: an official journal of the American Association for Cancer Research* 24, 4754-4762 (2018); Eschrich, S. A. et al. *International journal of radiation oncology, biology, physics* 75, 489-496 (2009); Sjöström, M. et al. *Journal of Clinical Oncology* 37, 3340-3349, (2019); Speers, C. et al. *Clinical cancer research: An official journal of the American Association for Cancer Research* 21, 3667-3677 (2015); Tramm, T. et al. *Clinical cancer research: An official journal of the American Association for Cancer Research* 20, 5272-5280 (2014)).
 25 ARTIC, una característica clínicogenómica de 27 genes, se desarrolló y validó previamente como pronóstico de LRR y predictivo del beneficio de la RT. Esta característica pudo identificar a los pacientes con mayor riesgo de LRR, con un beneficio reducido de la radioterapia y, por lo tanto, puede usarse para identificar a aquellos que requieren un tratamiento intensificado. Los pacientes identificados como de bajo riesgo por ARTIC tuvieron un mayor beneficio de la RT y no eran
 30 candidatos adecuados para la omisión de la RT. Se ha sugerido el uso de características diseñadas
 35

para estimar el riesgo de recurrencia a distancia después de la terapia sistémica para evitar la RT, y éstas se están probando en el contexto de ensayos clínicos en curso, pero estos ensayos no madurarán hasta dentro de varios años (consulte, Speers, C. & Pierce, LJ Current Breast Cancer Reports 12, 255-265 (2020)).

5 Los experimentos realizados durante el curso del desarrollo de modalidades para la presente invención (consulte el Ejemplo I) realizaron un análisis de SweBCG91-RT, un ensayo que aleatorizó a mujeres con cáncer de mama invasivo (BC) en estadio I-II con ganglios negativos +/- RT después de BCS y la cohorte Princess Margaret que aleatorizó a mujeres de 50 años o más con BC con ganglios T1-T2 negativos a +/- RT después de BCS. Los pacientes del ensayo SweBCG91-RT no
10 fueron tratados con terapia sistémica adyuvante, mientras que los pacientes del ensayo Princess Margaret fueron tratados con tamoxifeno. El perfil de tumores de todo el transcriptoma se realizó utilizando micromatrices Affymetrix Human Exon 1.0 ST. La cohorte SweBCG91-RT se dividió en una cohorte de entrenamiento de 243 pacientes y una cohorte de validación de 354 pacientes, y 16 genes distintivos (AGR2, B4GALT1, CLDN7, EZR, GNG11, JUN, MMP11, PKIB, PRPS1, PSMD10,
15 SH3BP5, SLC16A3, SLC7A11, SPP1, TNNT1 y UBE2E1; consulte la Tabla 5 y el Ejemplo I) fue entrenado para predecir LRR mediante regresión neta elástica, denominada Perfil para la omisión de radiación adyuvante local (POLAR). Se demostró que los pacientes clasificados como POLAR de bajo riesgo y no tratados con RT tenían una LRR a 10 años del 6 %, que no hubo un beneficio significativo de la RT en los pacientes POLAR de bajo riesgo y los pacientes clasificados como
20 POLAR de alto riesgo tenían una disminución significativa del riesgo de LRR cuando se tratan con RT. Estos resultados indican que la nueva característica genómica POLAR basada en la biología de LRR es capaz de identificar pacientes con un riesgo bajo de LRR a pesar de no recibir RT y, por lo tanto, los principales candidatos para la omisión de RT.

Por consiguiente, la presente divulgación se refiere a sistemas y métodos para proporcionar
25 evaluaciones de pronóstico individualizadas de la recurrencia de BC (por ejemplo, LRR). Los sistemas y métodos implican medir la expresión génica de una muestra de un paciente para crear una característica de expresión génica que identifique a los sujetos que no es probable que se beneficien de la RT.

En determinadas modalidades, la presente invención proporciona métodos que comprenden: a)
30 medir un nivel de expresión de uno o más genes en una muestra biológica de un paciente humano que tiene o está en riesgo de tener cáncer de mama (BC), en donde uno o más genes son seleccionados de AGR2, B4GALT1, CLDN7, EZR, GNG11, JUN, MMP11, PKIB, PRPS1, PSMD10, SH3BP5, SLC16A3, SLC7A11, SPP1, TNNT1 y UBE2E1; y b) determinar una probabilidad de recurrencia de BC para el paciente basándose en el nivel de expresión de uno o más genes
35 seleccionados.

En determinadas modalidades, la descripción proporciona métodos para predecir una probabilidad de recurrencia de BC para un paciente con BC en riesgo de tener BC que comprende: (a) medir, en una muestra obtenida del paciente, un nivel de expresión de uno o más de los siguientes genes: AGR2, B4GALT1, CLDN7, EZR, GNG11, JUN, MMP11, PKIB, PRPS1, PSMD10, SH3BP5, SLC16A3, SLC7A11, SPP1, TNNT1 y UBE2E1; y (b) predecir una probabilidad de recurrencia de BC para el paciente en función del nivel de expresión del uno o más genes, en donde el aumento de la expresión de AGR2, CLDN7, EZR, MMP11, PKIB, PRPS1, PSMD10, SLC16A3, SLC7A11, SPP1, TNNT1 y UBE2E1 se correlaciona con un mayor riesgo de recurrencia de BC, y en donde el aumento de la expresión de B4GALT1, GNG11, JUN y SH3BP5 se correlaciona con un riesgo reducido de recurrencia de cáncer de mama.

En ciertas modalidades, la presente divulgación divulga sistemas y métodos para diagnosticar, predecir y/o controlar el estado o resultado de un BC en un sujeto usando análisis basado en la expresión de uno o más genes/diana genéticos. Generalmente, el método comprende (a) obtener un nivel de expresión en una muestra de un sujeto para uno o más genes; y (b) determinar el riesgo del sujeto de recurrencia del cáncer en función del nivel de expresión de uno o más genes. En algunas modalidades, el método también puede comprender administrar RT si se identifica que el sujeto tiene riesgo de recurrencia del cáncer en función del nivel de expresión de uno o más genes, o retener RT si se identifica que el sujeto tiene un riesgo bajo de padecer cáncer. recurrencia del cáncer basada en el nivel de expresión de uno o más genes.

En determinadas modalidades, se proporcionan métodos para determinar si un sujeto tiene un riesgo alto o bajo de recurrencia del cáncer de mama. Generalmente, el método comprende: (a) proporcionar una muestra que comprende células de cáncer de mama de un sujeto; (b) analizar el nivel de expresión de uno o más genes en la muestra; y (c) determinar si el sujeto tiene un riesgo bajo de recurrencia del cáncer de mama en función del nivel de expresión de la pluralidad de dianas. En ciertas modalidades, el método puede comprender adicionalmente determinar si se trata o no al sujeto con radioterapia, quimioterapia o terapia endocrina adyuvante. Por ejemplo, un sujeto identificado como de bajo riesgo de recurrencia del cáncer de mama según los métodos de la presente divulgación, puede tener menos probabilidades o no de responder a la radioterapia adyuvante, mientras que un sujeto identificado como de mayor riesgo de recurrencia del cáncer de mama pueden tener más probabilidades de responder a la radioterapia adyuvante.

En determinadas modalidades, esta divulgación se refiere a sistemas y métodos para proporcionar evaluaciones de pronóstico individualizadas de la recurrencia de BC y la identificación de sujetos que probablemente no se beneficiarán de la RT. Dichos sistemas y métodos de la presente invención no se limitan a una manera particular de evaluar el riesgo de recurrencia de BC. En algunas

modalidades, la evaluación del riesgo de recurrencia de BC se proporciona en el formato de expresiones génicas distintivas que brindan información importante y fácil de entender sobre los tumores del paciente.

En determinadas modalidades, los sistemas y métodos de la presente invención implican medir y analizar la expresión génica de una muestra de un paciente para crear una expresión génica distintiva. En algunas modalidades, la expresión génica distintiva se proporciona como una puntuación o categoría de riesgo fácil de entender. Por ejemplo, la expresión génica distintiva puede proporcionarse como una puntuación de riesgo numérica, o la puntuación de riesgo puede usarse para asignar al paciente a una categoría de riesgo de recurrencia de BC. Por ejemplo, un paciente puede clasificarse como de alto o bajo riesgo de recurrencia de BC. En otras modalidades, la categoría de riesgo puede caracterizarse como riesgo muy bajo, riesgo bajo, riesgo moderadamente bajo, riesgo medio o intermedio, riesgo moderadamente alto, riesgo alto o riesgo muy alto de recurrencia de BC. La puntuación de riesgo o la categoría de riesgo pueden hacer que hablar con los pacientes sobre los resultados de sus pruebas sea fácil y eficiente y también puede ayudar al médico a tomar decisiones de tratamiento más rápidamente al reducir la cantidad de tiempo necesario para interpretar los resultados del paciente.

Por lo tanto, los métodos de la invención implican, en determinadas modalidades, la creación de puntuaciones de riesgo útiles para el tratamiento clínico del cáncer de mama. Las expresiones génicas distintivas pueden predecir, por ejemplo, un riesgo de recurrencia de la enfermedad y, como tal, las expresiones génicas distintivas pueden usarse para seleccionar un curso de tratamiento óptimo. Por ejemplo, las puntuaciones de riesgo pueden usarse para identificar pacientes que tienen un alto riesgo de recurrencia y, por lo tanto, buenos candidatos para RT. Alternativamente, las puntuaciones de riesgo pueden usarse para identificar pacientes que tienen un riesgo bajo de recurrencia y, por lo tanto, buenos candidatos para omitir la RT. En consecuencia, las puntuaciones de riesgo pueden resultar útiles para clasificar a un paciente y seleccionar un tratamiento adecuado. Por ejemplo, en algunas modalidades, las mediciones de expresión para genes seleccionados o combinaciones de los mismos pueden formularse en modelos o algoritmos lineales o no lineales y convertirse en una puntuación de probabilidad o riesgo.

En determinadas modalidades, la puntuación de probabilidad o riesgo se puede calcular usando un algoritmo lineal en donde a cada gen analizado se le asigna un peso o coeficiente basado en la correlación individual del gen con la recurrencia de BC (véase la Tabla 8). Los niveles de expresión ponderados para uno o más genes ensayados pueden luego sumarse para producir una puntuación de probabilidad o riesgo. En una modalidad, si se analiza el nivel de expresión de un único gen, el nivel de expresión ponderado para ese gen puede considerarse la puntuación de riesgo y puede usarse para determinar la categoría de riesgo del paciente.

En determinadas modalidades, la puntuación resultante puede comprender un valor numérico y puede estar en un rango entre -5,0 y 10,0, o entre -4,0 y 9,0, o entre -3,0 y 8,0, o entre -2,0 y 7,0, o entre -1,0 y 6,0, o entre -0,5 y 5,0, o entre -0,5 y 4,0, o entre -0,5 y 3,0, o entre 0 y 2,5, o entre 0,5 y 2,0 o entre 1,0 y 1,5. En modalidades adicionales, la puntuación puede comprender, por ejemplo, un valor numérico tal como -5,0, -4,5, -4,0, -3,5, -3,0, -2,5, -2,0, -1,5, -1,0, -0,95, -0,90, -0,85, -0,80, -0,75, -0,70, -0,65, -0,60, -0,55, -0,50, -0,45, -0,40, -0,35, -0,30, -0,25, -0,20, -0,15, -0,10, -0,05, 0, 0,05, 0,10, 0,15, 0,20, 0,25, 0,30, 0,35, 0,40, 0,45, 0,50, 0,55, 0,60, 0,65, 0,70, 0,75, 0,80, 0,85, 0,90, 0,95, 1,0, 1,5, 2,0, 2,5, 3,0, 3,5, 4,0, 4,5, 5,0, 5,5, 6,0, 6,5, 7,0, 7,5, 8,0, 8,5, 9,0, 9,5 o 10,0.

En otras modalidades, la puntuación de riesgo se puede comparar con un umbral predeterminado para determinar si un paciente tiene un riesgo alto (es decir, con una puntuación por encima del umbral predeterminado) o un riesgo bajo (es decir, con una puntuación por debajo del umbral predeterminado) para la recurrencia de BC y, por lo tanto, es probable o no que se beneficie del RT. En determinadas modalidades, el umbral predeterminado puede ser 0, de modo que un paciente con una puntuación de riesgo positiva se considera de alto riesgo y un paciente con una puntuación de riesgo negativa se considera de bajo riesgo. En otra modalidad, el umbral predeterminado puede estar comprendido en un intervalo entre -1,0 y 5,0, por ejemplo entre -0,5 y 3,0, entre 0 y 2,0, entre 0 y 1,0, entre 0,25 y 0,75, entre 0,30 y 0,70, entre 0,40 y 0,60, entre 0,50 y 0,60, entre 0,55 y 0,65, entre 0,55 y 0,60, y entre 0,60 y 0,65. En una modalidad adicional, el umbral predeterminado puede incluir -1,0, -0,90, -0,80, -0,70, -0,60, -0,55, -0,50, -0,45, -0,40, -0,35, -0,30, -0,25, -0,20, -0,15, -0,10, -0,05, 0, 0,05, 0,10, 0,15, 0,20, 0,25, 0,30, 0,35, 0,40, 0,41, 0,42, 0,43, 0,44, 0,45, 0,46, 0,47, 0,48, 0,49, 0,50, 0,51, 0,52, 0,53, 0,54, 0,55, 0,56, 0,57, 0,58, 0,59, 0,60, 0,61, 0,62, 0,63, 0,64, 0,65, 0,66, 0,67, 0,68, 0,69, 0,70, 0,75, 0,80, 0,85, 0,90, 0,95, 1,0, 1,5, 2,0, 2,5 o 3,0.

En modalidades adicionales, un riesgo o probabilidad de recurrencia puede caracterizarse como riesgo muy bajo, riesgo bajo, riesgo moderadamente bajo, riesgo promedio o intermedio, riesgo moderadamente alto, riesgo alto o riesgo muy alto. En dichas modalidades, la puntuación de riesgo se puede comparar con más de un umbral predeterminado para determinar si un paciente tiene un riesgo muy bajo, un riesgo bajo, un riesgo moderadamente bajo, un riesgo promedio o intermedio, un riesgo moderadamente alto, un riesgo alto o un riesgo muy alto de recurrencia de BC o cualquier subconjunto o combinación de estos niveles de riesgo. Por ejemplo, la puntuación de riesgo se puede comparar con más de un umbral predeterminado para determinar si un paciente tiene un riesgo bajo, un riesgo medio o intermedio o un riesgo alto de recurrencia de BC. En otra modalidad, la puntuación de riesgo se puede comparar con más de un umbral predeterminado para determinar si un paciente tiene un riesgo muy bajo, un riesgo bajo, un riesgo medio o intermedio, un riesgo alto o un riesgo muy alto de recurrencia de BC. En aquellos casos en donde la puntuación de riesgo deba compararse con más de un umbral predeterminado, los umbrales predeterminados pueden

estar dentro de un rango entre -1,0 y 5,0, por ejemplo entre -0,5 y 3,0, entre 0 y 2,0, entre 0 y 1,0, entre 0,25 y 0,75, entre 0,30 y 0,70, entre 0,40 y 0,60, entre 0,50 y 0,60, entre 0,55 y 0,65, entre 0,55 y 0,60 y entre 0,60 y 0,65. En una modalidad adicional, los umbrales predeterminados pueden incluir -1,0, -0,90, -0,80, -0,70, -0,60, -0,55, -0,50, -0,45, -0,40, -0,35, -0,30, -0,25, -0,20, -0,15, -0,10, -0,05, 0, 0,05, 0,10, 0,15, 0,20, 0,25, 0,30, 0,35, 0,40, 0,41, 0,42, 0,43, 0,44, 0,45, 0,46, 0,47, 0,48, 0,49, 0,50, 0,51, 0,52, 0,53, 0,54, 0,55, 0,56, 0,57, 0,58, 0,59, 0,60, 0,61, 0,62, 0,63, 0,64, 0,65, 0,66, 0,67, 0,68, 0,69, 0,70, 0,75, 0,80, 0,85, 0,90, 0,95, 1,0, 1,5, 2,0, 2,5, o 3,0 o cualquier combinación de los mismos. En ciertas modalidades, cada uno de los más de un umbral predeterminado o rango de umbrales puede ser distinto y puede estar asociado con un nivel de riesgo distinto.

En algunas modalidades, la puntuación de riesgo o probabilidad se puede escalar o ajustar a una escala fácilmente interpretable. Por ejemplo, la puntuación de riesgo se puede escalar o ajustar a un valor entre 1 y 100. Esta escala fácilmente interpretable puede hacer que las decisiones sobre tratamientos y la conversación con los pacientes sobre los resultados de sus pruebas y las opciones de tratamiento sean fáciles y eficientes.

En ciertas modalidades, los umbrales predeterminados para determinar los niveles de riesgo pueden escalarse o ajustarse de manera similar, por ejemplo, de manera que coincidan o correspondan con el ajuste realizado para obtener la escala final. Por ejemplo, si la puntuación de riesgo se escala o se ajusta a una escala con valores entre 1 y 100, el umbral predeterminado, o más de un umbral predeterminado, se puede escalar o ajustar en consecuencia de manera que los valores de dichos umbrales caigan dentro de la escala 1-100.

Los métodos de la invención pueden incluir además combinar expresiones distintivas o puntuaciones de riesgo con otros factores clínicos para proporcionar una única evaluación de riesgo.

Como se utiliza en el presente documento, "riesgo de recurrencia de BC" o "probabilidad de recurrencia de BC" se refiere a una probabilidad estadística (p. ej., probabilidad) de recurrencia de BC durante un período prolongado de tiempo (p. ej., 1 mes, 1 año, 5 años, 10 años, etc.). En algunas modalidades, el riesgo de recurrencia de BC implica un valor inicial en donde un paciente ha tenido una intervención exitosa (por ejemplo, intervención quirúrgica) y se caracteriza por no tener BC y/o células cancerosas que progresan activamente. Como tal, un riesgo o probabilidad de recurrencia implica la probabilidad de que el cáncer recurra de alguna manera.

En algunas modalidades, los niveles de expresión aumentados de AGR2, CLDN7, EZR, MMP11, PKIB, PRPS1, PSMD10, SLC16A3, SLC7A11, SPP1, TNNT1 y UBE2E1 se correlacionan con un mayor riesgo o probabilidad de recurrencia del cáncer de mama.

En algunas modalidades, los niveles de expresión disminuidos de AGR2, CLDN7, EZR, MMP11, PKIB, PRPS1, PSMD10, SLC16A3, SLC7A11, SPP1, TNNT1 y UBE2E1 se correlacionan cada uno con un menor riesgo o probabilidad de una recurrencia del cáncer de mama.

En algunas modalidades, los niveles de expresión aumentados de B4GALT1, GNG11, JUN y SH3BP5 se correlacionan con un menor riesgo o probabilidad de recurrencia del cáncer de mama. En algunas modalidades, la disminución de los niveles de expresión de B4GALT1, GNG11, JUN y SH3BP5 se correlaciona con un aumento o probabilidad de riesgo de recurrencia del cáncer de mama.

Dichos sistemas y métodos no se limitan al uso de un tipo particular o clase de muestra de un sujeto. Por ejemplo, en algunas modalidades la muestra o muestra biológica es una muestra de tejido, muestra de fluido corporal, muestra de sangre, muestra de secreción de órganos, muestra de CSF, muestra de saliva, muestra de plasma, muestra de suero o muestra de orina. En algunas modalidades, la muestra puede comprender tejido mamario o tejido circundante, una biopsia de mama, una muestra de tumor o tejido que contiene células mamarias o células de cáncer de mama. En algunas modalidades, el sujeto es un sujeto humano. En algunas modalidades, el sujeto es un sujeto humano en riesgo de desarrollar BC. En algunas modalidades, el sujeto es un sujeto humano femenino en riesgo de desarrollar BC. En algunas modalidades, el sujeto es un sujeto humano diagnosticado con BC. En algunas modalidades, el sujeto es un sujeto humano femenino diagnosticado con BC. Se puede sospechar que el sujeto tiene cáncer debido a diversos síntomas, incluida la detección de un bulto o masa. En algunas modalidades, el cáncer es cáncer de mama en etapa temprana, es decir, cáncer que está contenido completamente dentro de la mama.

En algunas modalidades, el análisis o comparación de la expresión génica se puede realizar con un algoritmo de agrupamiento jerárquico no supervisado, tal como un algoritmo de agrupamiento de K-medias. Un algoritmo de agrupamiento es un algoritmo que agrupa o agrupa un conjunto de objetos de tal manera que los objetos del mismo grupo (llamado grupo) se parecen más entre sí que los de otros grupos (grupos). El algoritmo de agrupación puede agrupar los niveles de expresión de ARN de la muestra del paciente con los niveles de expresión de ARN esperados en una o más etapas del cáncer. Los niveles de expresión de ARN esperados en pacientes que tienen un riesgo bajo de recurrencia de BC o un riesgo alto de recurrencia de BC pueden provenir de una o más muestras de tumores asociadas con resultados conocidos. Los niveles de expresión de ARN pueden agruparse en función de sus similitudes de expresión. En algunas modalidades, el algoritmo de agrupación agrupa los niveles de expresión de ARN en distintos grupos asociados con los resultados conocidos. Los grupos pueden reflejar una serie de resultados que son indicativos de pronósticos.

Dichos sistemas y métodos no se limitan a una manera o técnica particular para medir y/o determinar la expresión génica. "Expresión génica", como se utiliza en el presente documento, se refiere a los niveles relativos de expresión y/o patrón de expresión de un gen en una muestra biológica.

Se conocen en la técnica técnicas para medir la expresión génica. De hecho, se contempla e incorpora en el presente documento cualquier técnica conocida para medir la expresión génica. La

expresión génica se puede determinar mediante cualquier técnica adecuada que incluye, entre otras, técnicas que comprenden técnicas basadas en PCR (por ejemplo, PCR en tiempo real), técnicas de electroforesis en gel, técnicas cromatográficas, técnicas basadas en anticuerpos, técnicas de centrifugación o combinaciones de las mismas. Los métodos para medir la expresión génica pueden comprender medir cantidades de ADNc elaborado a partir de ARN aislado de tejido. En algunas modalidades, las técnicas de medición de la expresión génica implican ensayos de expresión génica con o sin el uso de chips genéticos (véase, Onken et al., J Molec Diag 12(4): 461-468 (2010); y Kirby et al., Adv. Clin Chem. 44: 247-292 (2007). En algunas modalidades, las técnicas de medición de la expresión génica implican chips de genes de affymetrix y chips de ARN y kits de ensayo de expresión génica (por ejemplo, ensayos de expresión génica TaqMan® de Applied Biosystems™).

Las técnicas adicionales para determinar el nivel de expresión génica en una muestra biológica implican el proceso de amplificación de ácidos nucleicos, por ejemplo, mediante RT-PCR (patente estadounidense N.º 4,683,202), reacción en cadena de la ligasa (Barany, Proc. Natl. Acad. Sci. USA 88:189-93, 1991), replicación de secuencia autosostenida (Guatelli et al., Proc. Natl. Acad. Sci. USA 87:1874-78, 1990), sistema de amplificación transcripcional (Kwoh et al., Proc. Natl. Acad. Sci. USA 86:1173-77, 1989), Q-Beta Replicase (Lizardi et al., Bio/Technology 6:1197, 1988), replicación de círculo rodante (patente estadounidense N.º 5,854,033), o cualquier otro método de amplificación de ácido nucleico, seguido de la detección de las moléculas amplificadas usando técnicas bien conocidas por los expertos en la técnica.

En algunas modalidades, la expresión génica se determina con un ensayo cuantitativo de amplificación de señal y diana en tiempo real específico de alelo (QuARTS). En cada ensayo QuARTS se producen tres reacciones secuencialmente, incluida la amplificación (reacción 1) y la escisión de la sonda diana (reacción 2) en la reacción primaria; y escisión FRET y generación de señal fluorescente (reacción 3) en la reacción secundaria. Cuando el ácido nucleico diana se amplifica con cebadores específicos, una sonda de detección específica con una secuencia de solapa se une libremente al amplicón. La presencia del oligonucleótido invasivo específico en el sitio de unión diana provoca que una nucleasa 5', *por ejemplo*, una endonucleasa FEN-1, libere la secuencia de solapa cortando entre la sonda de detección y la secuencia de solapa. La secuencia de solapa es complementaria a una porción que no es de horquilla de un casete FRET correspondiente. En consecuencia, la secuencia flap funciona como un oligonucleótido invasivo en el casete FRET y efectúa una escisión entre el fluoróforo del casete FRET y un inactivador, lo que produce una señal fluorescente. La reacción de escisión puede cortar múltiples sondas por diana y así liberar múltiples fluoróforos por solapa, proporcionando una amplificación exponencial de la señal. QuARTS puede detectar múltiples dianas en un solo pocillo de reacción utilizando casetes FRET con diferentes tintes. Véase, *por ejemplo*, en Zou et al. (2010) "Sensitive quantification of

methylyated markers with a novel methylation specific technology" *Clin Chem* **56**: A199), y las patentes estadounidenses No. 8,361,720; 8,715,937; 8,916,344; y 9,212,392, cada una de las cuales se incorpora aquí como referencia para todos los propósitos.

En determinadas modalidades, los niveles de expresión génica pueden normalizarse para minimizar errores o variaciones entre muestras. Por lo tanto, la normalización es útil para realizar comparaciones precisas de la expresión génica entre muestras, ya que la expresión génica por muestra puede verse afectada por variables no biológicas que surgen durante la recolección y el procesamiento de muestras, lo que puede agregar ruido a la señal verdadera. Los ensayos pueden proporcionar la normalización incorporando la expresión de ciertos genes normalizadores, que no difieren significativamente en los niveles de expresión en las condiciones relevantes. Los genes de normalización ilustrativos divulgados en el presente documento incluyen genes de mantenimiento. (Consulte, por ejemplo, E. Eisenberg, et al., *Trends in Genetics* 19(7):362-365 (2003).) La normalización puede basarse en la señal media o mediana (Ct o Cp) de todos los genes analizados o un gran subconjunto de ellos (enfoque de normalización global). En general, los genes normalizadores, también denominados genes de referencia, deben ser genes que se sepa que no exhiben una expresión significativamente diferente en BC en comparación con el tejido mamario no canceroso, y que no se ven afectados significativamente por diversas condiciones de muestra y proceso, por lo que proporcionan para normalizar los efectos extraños. Por ejemplo, los genes de referencia útiles en los métodos descritos en la presente pueden incluir genes utilizados con frecuencia en la técnica para normalizar patrones de expresión génica, que incluyen, de modo no taxativo, gliceraldehído-3-fosfato-deshidrogenasa (GAPDH) y β -actina, o cualquier otro gen de referencia conocido en la técnica. Los expertos en la técnica reconocerán que la normalización se puede lograr de numerosas maneras, y las técnicas descritas anteriormente pretenden ser solo ilustrativas, no exhaustivas.

En algunas modalidades, analizar el nivel de expresión de una o más dianas genéticas en la muestra puede comprender aplicar la muestra a una micromatriz.

En determinadas modalidades, la expresión génica analizada mediante micromatriz se puede normalizar utilizando el método de normalización de matriz de canal único (SCAN). El método SCAN utiliza una técnica de muestra única, en lugar de procesar muestras de micromatrices como grupos, evitando así los sesgos que pueden introducirse en el procesamiento grupal. Este método normaliza cada muestra individualmente modelando los efectos de la composición de la sonda-nucleótido sobre la intensidad de la fluorescencia y eliminando el ruido de fondo específico de la sonda y la matriz. (Piccolo SR, Sun Y, Campbell JD, Lenburg ME, Bild AH, Johnson WE (2012). "A single-sample microarray normalization method to facilitate personalized-medicine workflows." *Genomics*, 100(6), 337-344).

En algunos casos, analizar el nivel de expresión de uno o más genes comprende el uso de un algoritmo. El algoritmo puede usarse para producir un clasificador genómico. Alternativamente, el clasificador puede comprender una región de selección de sonda. En algunos casos, analizar el nivel de expresión para una pluralidad de dianas comprende detectar y/o cuantificar uno o más genes. En algunas modalidades, analizar el nivel de expresión de uno o más genes comprende secuenciar la pluralidad de dianas. En algunas modalidades, ensayar el nivel de expresión para una o más dianas génicas comprende amplificar la pluralidad de dianas. En algunas modalidades, analizar el nivel de expresión para una o más dianas génicas comprende cuantificar las dianas. En algunas modalidades, ensayar el nivel de expresión para una o más dianas comprende llevar a cabo una reacción multiplexada en la pluralidad de dianas.

En algunos casos, analizar el nivel de expresión de una pluralidad de genes comprende detectar y/o cuantificar una pluralidad de analitos diana. En algunas modalidades, analizar el nivel de expresión de una pluralidad de genes comprende secuenciar una pluralidad de ácidos nucleicos diana. En algunas modalidades, analizar el nivel de expresión de una pluralidad de genes biomarcadores comprende amplificar una pluralidad de ácidos nucleicos diana. En algunas modalidades, el ensayo del nivel de expresión de una pluralidad de genes biomarcadores comprende llevar a cabo una reacción multiplexada en una pluralidad de analitos diana.

Los métodos descritos en el presente documento a menudo comprenden analizar el nivel de expresión de una pluralidad de dianas. La pluralidad de dianas puede comprender dianas codificantes y/o dianas no codificantes de un gen codificante de proteínas o un gen no codificante de proteínas. Una estructura de gen codificante de proteínas puede comprender un exón y un intrón. El exón puede comprender además una secuencia codificante (CDS) y una región no traducida (UTR). El gen codificante de proteínas puede transcribirse para producir un pre-ARNm y el pre-ARNm puede procesarse para producir un ARNm maduro. El ARNm maduro puede traducirse para producir una proteína.

Una estructura de gen no codificante de proteínas puede comprender un exón y un intrón. Por lo general, la región del exón de un gen no codificante de proteínas contiene principalmente una UTR. El gen no codificante de proteínas puede transcribirse para producir un pre-ARNm y el pre-ARNm puede procesarse para producir un ARN no codificante (ARNnc).

Una diana codificante puede comprender una secuencia codificante de un exón. Una diana no codificante puede comprender una secuencia UTR de un exón, una secuencia de intrón, una secuencia intergénica, una secuencia promotora, un transcrito no codificante, un antisentido CDS, un antisentido intrónico, un antisentido UTR o un transcrito antisentido no codificante. Un transcrito no codificante puede comprender un ARN no codificante (ARNnc).

Dichos sistemas y métodos no se limitan al análisis de genes específicos. En algunas modalidades,

uno o más genes se seleccionan de AGR2, B4GALT1, CLDN7, EZR, GNG11, JUN, MMP11, PKIB, PRPS1, PSMD10, SH3BP5, SLC16A3, SLC7A11, SPP1, TNNT1 y UBE2E1. Dichos sistemas y modalidades no se limitan al uso de una cantidad o combinación específica del uno o más genes (por ejemplo, una combinación de solo 1 gen, 1-2 genes, 1-3, 1-4, 1-5, 1-6, 1-7, 1-8, 1-9, 1-10, 1-11, 1-12, 1-13, 1-14, 1-15, 1-16, 2-3, 2-4, 2-5, 2-6, 2-7, 2-8, 2-9, 2-10, 2-11, 2-12, 2-13, 2-14, 2-15, 2-16, 3-4, 3-5, 3-6, 3-7, 3-8, 3-9, 3-10, 3-11, 3-12, 3-13, 3-14, 3-15, 3-16, 4-5, 4-6, 4-7, 4-8, 4-9, 4-10, 4-11, 4-12, 4-13, 4-14, 4-15, 4-16, 5-6, 5-7, 5-8, 5-9, 5-10, 5-11, 5-12, 5-13, 5-14, 5-15, 5-16, 6-7, 6-8, 6-9, 6-10, 6-11, 6-12, 6-13, 6-14, 6-15, 6-16, 7-8, 7-9, 7-10, 7-11, 7-12, 7-13, 7-14, 7-15, 7-16, 8-9, 8-10, 8-11, 8-12, 8-13, 8-14, 8-15, 8-16, 9-10, 9-11, 9-12, 9-13, 9-14, 9-15, 9-16, 10-11, 10-12, 10-13, 10-14, 10-15, 10-16, 11-12, 11-13, 11-14, 11-15, 11-16, 12-13, 12-14, 12-15, 12-16, 13-14, 13-15, 13-16, 14-15, 14-16, 15-16, o 16 genes). En algunas modalidades, el método comprende combinaciones de 16 o menos, 15 o menos, 14 o menos, 13 o menos, 12 o menos, 11 o menos, 10 o menos, 9 o menos, 8 o menos, 7 o menos, 6 o menos, 5 o menos, 4 o menos, 3 o menos, o 2 o menos genes. En otras modalidades, el método puede comprender 2 o más, 3 o más, 4 o más, 5 o más, 6 o más, 7 o más, 8 o más, 9 o más, 10 o más, 11 o más, 12 o más, 13 o más, 14 o más, o 15 o más genes. En otra modalidad, el uno o más genes de los métodos divulgados consiste en 16 genes.

En algunos casos, la una o más dianas génicas comprende una diana codificante, una diana no codificante o cualquier combinación de las mismas. En algunos casos, la diana codificante comprende una secuencia exónica. En otros casos, la diana no codificante comprende una secuencia exónica o no exónica. Alternativamente, una diana no codificante comprende una secuencia UTR, una secuencia intrónica, antisentido o un transcrito de ARN no codificante. En algunos casos, una diana no codificante comprende secuencias que se superponen parcialmente con una secuencia UTR o una secuencia intrónica. Una diana no codificante también incluye transcritos no exónicos y/o exónicos. Las secuencias exónicas pueden comprender regiones en un gen que codifica una proteína, tal como un exón, UTR o una porción del mismo. Las secuencias no exónicas pueden comprender regiones en un gen codificante de proteínas, no codificante de proteínas o una parte de este. Por ejemplo, las secuencias no exónicas pueden comprender regiones intrónicas, regiones promotoras, regiones intergénicas, un transcrito no codificante, una región antisentido de exón, una región antisentido intrónica, una región antisentido UTR, una región antisentido de transcrito no codificante o una porción de estas. En otros casos, la pluralidad de dianas comprende un transcrito de ARN no codificante.

Las dianas genéticas pueden comprender una o más dianas seleccionadas de un clasificador descrito en el presente documento. El clasificador puede generarse a partir de uno o más modelos o algoritmos. El uno o más modelos o algoritmos pueden ser Naïve Bayes (NB), Partición recursiva

(Rpart), bosque aleatorio (RF), máquina de vectores de soporte (SVM), k-vecino más cercano (KNN), análisis discriminante de alta dimensión (HDDA), modelo lineal, o una combinación de los mismos. El clasificador puede tener un AUC igual o superior a 0,60. El clasificador puede tener un AUC igual o superior a 0,61. El clasificador puede tener un AUC igual o superior a 0,62. El clasificador puede tener un AUC igual o superior a 0,63. El clasificador puede tener un AUC igual o superior a 0,64. El clasificador puede tener un AUC igual o superior a 0,65. El clasificador puede tener un AUC igual o superior a 0,66. El clasificador puede tener un AUC igual o superior a 0,67. El clasificador puede tener un AUC igual o superior a 0,68. El clasificador puede tener un AUC igual o superior a 0,69. El clasificador puede tener un AUC igual o superior a 0,70. El clasificador puede tener un AUC igual o superior a 0,75. El clasificador puede tener un AUC igual o superior a 0,77. El clasificador puede tener un AUC igual o superior a 0,78. El clasificador puede tener un AUC igual o superior a 0,79. El clasificador puede tener un AUC igual o superior a 0,80. El AUC puede ser clínicamente significativo según su intervalo de confianza (IC) del 95 %. La precisión del clasificador puede ser de al menos aproximadamente el 70 %. La precisión del clasificador puede ser de al menos aproximadamente el 73 %. La precisión del clasificador puede ser de al menos aproximadamente el 75 %. La precisión del clasificador puede ser de al menos aproximadamente el 77 %. La precisión del clasificador puede ser de al menos aproximadamente el 80 %. La precisión del clasificador puede ser de al menos aproximadamente el 83 %. La precisión del clasificador puede ser de al menos aproximadamente el 84 %. La precisión del clasificador puede ser de al menos aproximadamente el 86 %. La precisión del clasificador puede ser de al menos aproximadamente el 88 %. La precisión del clasificador puede ser de al menos aproximadamente el 90 %. El valor p del clasificador puede ser menor o igual a 0,05. El valor p del clasificador puede ser menor o igual a 0,04. El valor p del clasificador puede ser menor o igual a 0,03. El valor p del clasificador puede ser menor o igual a 0,02. El valor p del clasificador puede ser menor o igual a 0,01. El valor p del clasificador puede ser menor o igual a 0,008. El valor p del clasificador puede ser menor o igual a 0,006. El valor p del clasificador puede ser menor o igual a 0,004. El valor p del clasificador puede ser menor o igual a 0,002. El valor p del clasificador puede ser menor o igual a 0,001.

La una o más dianas genéticas pueden comprender una o más dianas seleccionadas de un clasificador de modelo lineal. La pluralidad de dianas puede comprender dos o más dianas seleccionadas de un clasificador de modelo lineal. La pluralidad de dianas puede comprender tres o más dianas seleccionadas de un clasificador de modelo lineal. La pluralidad de dianas puede comprender 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 27 o más dianas seleccionadas de un clasificador de modelo lineal. El clasificador de modelo lineal puede ser un clasificador LM2, y LM3 o LM4. El clasificador de modelo lineal puede ser un clasificador LM16 (por ejemplo, un clasificador de modelo lineal con 16 dianas). Por ejemplo, un clasificador de modelo lineal de la

presente divulgación puede comprender dos o más dianas seleccionadas de la Tabla 5.

La una o más dianas genéticas pueden comprender una o más dianas seleccionadas de un clasificador SVM. La pluralidad de dianas puede comprender 2, 3, 4, 5, 6, 7, 8, 9, 10 o más dianas seleccionadas de un clasificador SVM. La pluralidad de dianas puede comprender 12, 13, 14, 15, 17, 20, 22, 25 o más dianas seleccionadas de un clasificador SVM. El clasificador SVM puede ser un clasificador SVM2. Un clasificador SVM de la presente divulgación puede comprender dos o más dianas seleccionadas de la Tabla 5.

La una o más dianas genéticas pueden comprender una o más dianas seleccionadas de un clasificador KNN. La pluralidad de dianas puede comprender 2, 3, 4, 5, 6, 7, 8, 9, 10 o más dianas seleccionadas de un clasificador KNN. La pluralidad de dianas puede comprender 12, 13, 14, 15, 17, 20, 22, 25 o más dianas seleccionadas de un clasificador KNN. Por ejemplo, un clasificador de KNN de la presente divulgación puede comprender dos o más dianas seleccionadas de la Tabla 5.

La una o más dianas génicas pueden comprender una o más dianas seleccionadas de un clasificador de Bayes sin tratamiento previo (NB). La pluralidad de dianas puede comprender 2, 3, 4, 5, 6, 7, 8, 9, 10 o más dianas seleccionadas de un clasificador de NB. La pluralidad de dianas puede comprender 12, 13, 14, 15, 17, 20, 22, 25 o más dianas seleccionadas de un clasificador NB. Por ejemplo, un clasificador de NB de la presente descripción puede comprender dos o más dianas seleccionadas de la Tabla 5.

La una o más dianas génicas pueden comprender una o más dianas seleccionadas de un clasificador de partición recursiva (Rpart). La pluralidad de dianas puede comprender 2, 3, 4, 5, 6, 7, 8, 9, 10 o más dianas seleccionadas de un clasificador Rpart. La pluralidad de dianas puede comprender 12, 13, 14, 15, 17, 20, 22, 25 o más dianas seleccionadas de un clasificador Rpart. Por ejemplo, un clasificador Rpart de la presente descripción puede comprender dos o más dianas seleccionadas de la Tabla 5.

La una o más dianas genéticas pueden comprender una o más dianas seleccionadas de un clasificador de análisis discriminante de alta dimensión (HDDA). La pluralidad de dianas puede comprender dos o más dianas seleccionadas de un clasificador de análisis discriminado de alta dimensión (HDDA). La pluralidad de dianas puede comprender tres o más dianas seleccionadas de un clasificador de análisis discriminado de alta dimensión (HDDA). La pluralidad de dianas puede comprender 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 17, 20, 22, 25 o más dianas seleccionadas de un clasificador de análisis discriminado de alta dimensión (HDDA). Por ejemplo, un clasificador Rpart de la presente descripción puede comprender dos o más dianas seleccionadas de la Tabla 5.

La presente divulgación proporciona un conjunto de sondas para diagnosticar, controlar y/o predecir un estado o resultado de cáncer de mama en un sujeto que comprende una pluralidad de sondas, en donde (i) las sondas en el conjunto son capaces de detectar un nivel de expresión de al menos

una diana; y (ii) el nivel de expresión determina el estado del cáncer (por ejemplo, riesgo de recurrencia) del sujeto con al menos aproximadamente un 40 % de especificidad.

El conjunto de sondas puede comprender una o más sondas polinucleotídicas. Las sondas de polinucleótidos individuales comprenden una secuencia de nucleótidos derivada de la secuencia de nucleótidos de las secuencias diana o secuencias complementarias de las mismas. La secuencia de nucleótidos de la sonda polinucleotídica se diseña de manera que corresponda o sea complementaria a las secuencias diana. La sonda polinucleotídica puede hibridarse específicamente en condiciones de hibridación estrictas o reducidas con una región de las secuencias diana, con su complemento o con una secuencia de ácido nucleico (tal como un ADNc) derivada de la misma.

La selección de las secuencias de las sondas polinucleotídicas y la determinación de su unicidad se pueden llevar a cabo *in silico* usando técnicas conocidas en la técnica, por ejemplo, basándose en una búsqueda BLASTN de la secuencia polinucleotídica en cuestión frente a bases de datos de secuencias genéticas, tales como la Secuencia del Genoma Humano, UniGene, dbEST o la base de datos no redundante del NCBI. En una modalidad de la divulgación, la sonda polinucleotídica es complementaria a una región de un ARNm diana derivado de una secuencia diana en el conjunto de sondas. También se pueden emplear programas informáticos para seleccionar secuencias de sondas que no pueden hibridarse de forma cruzada o que no pueden hibridarse de forma no específica.

En algunos casos, la hibridación de micromatrices de ARN, extraído de muestras de tejido de cáncer de mama y amplificado, puede producir un conjunto de datos que luego se resume y normaliza mediante la técnica fRMA. Después de la eliminación (o filtración) de los PSR de hibridación cruzada y de los PSR que contienen menos de 4 sondas, los PSR restantes se pueden utilizar en análisis posteriores. Después de fRMA y filtración, los datos se pueden descomponer en sus componentes principales y se utiliza un modelo de análisis de varianza para determinar en qué medida un efecto de lote permanece presente en los primeros 10 componentes principales.

Estos PSR restantes pueden luego someterse a filtración mediante una prueba T entre muestras CR (recurrencia clínica) y no CR. Utilizando un valor de corte de p de 0,01, las características restantes (por ejemplo, PSR) se pueden refinar aún más. La selección de características se puede realizar mediante regresión logística regularizada utilizando la penalización de red elástica. La regresión regularizada se puede iniciar más de 1000 veces utilizando todos los datos de entrenamiento; con cada iteración del arranque, se pueden tabular las características que tienen un coeficiente distinto de cero después de una validación cruzada triple. En algunos casos, para la construcción del modelo se utilizaron características que se seleccionaron en al menos el 25 % del total de ejecuciones.

Las sondas de polinucleótidos de la presente divulgación pueden variar en longitud desde aproximadamente 15 nucleótidos hasta la longitud completa de la diana codificante o la diana no

codificante. En una modalidad de la divulgación, las sondas de polinucleótidos tienen al menos aproximadamente 15 nucleótidos de longitud. En otra modalidad, las sondas de polinucleótidos tienen al menos aproximadamente 20 nucleótidos de longitud. En otra modalidad, las sondas polinucleotídicas tienen al menos unos 25 nucleótidos de longitud. En otra modalidad, las sondas de polinucleótidos tienen una longitud de entre aproximadamente 15 nucleótidos y aproximadamente 500 nucleótidos. En otras modalidades, las sondas de polinucleótidos tienen entre aproximadamente 15 nucleótidos y aproximadamente 450 nucleótidos, aproximadamente 15 nucleótidos y aproximadamente 400 nucleótidos, aproximadamente 15 nucleótidos y aproximadamente 350 nucleótidos, aproximadamente 15 nucleótidos y aproximadamente 300 nucleótidos, aproximadamente 15 nucleótidos y aproximadamente 250 nucleótidos, aproximadamente 15 nucleótidos y unos 200 nucleótidos de longitud. En algunas modalidades, las sondas tienen al menos 15 nucleótidos de longitud. En algunas modalidades, las sondas tienen al menos 15 nucleótidos de longitud. En algunas modalidades, las sondas son de al menos 20 nucleótidos, al menos 25 nucleótidos, al menos 50 nucleótidos, al menos 75 nucleótidos, al menos 100 nucleótidos, al menos 125 nucleótidos, al menos 150 nucleótidos, al menos 200 nucleótidos, al menos 225 nucleótidos, al menos 250 nucleótidos, al menos 275 nucleótidos, al menos 300 nucleótidos, al menos 325 nucleótidos, al menos 350 nucleótidos, al menos 375 nucleótidos de longitud.

Las sondas polinucleotídicas de un conjunto de sondas pueden comprender ARN, ADN, ARN o miméticos de ADN, o combinaciones de los mismos, y pueden ser monocatenarias o bicatenarias. Por lo tanto, las sondas polinucleotídicas pueden estar compuestas de nucleobases de origen natural, azúcares y enlaces internucleosídicos covalentes (estructura), así como sondas polinucleotídicas que tienen porciones de origen no natural que funcionan de manera similar. Dichas sondas de polinucleótidos modificadas o sustituidas pueden proporcionar propiedades deseables tales como, por ejemplo, una mayor afinidad por un gen diana y una mayor estabilidad. El conjunto de sondas puede comprender una diana codificante y/o una diana no codificante.

En algunas modalidades, el conjunto de sondas comprende una pluralidad de secuencias diana que se hibridan con al menos aproximadamente 5 dianas codificantes y/o dianas no codificantes. Alternativamente, el conjunto de sondas comprende una pluralidad de secuencias diana que se hibridan con al menos aproximadamente 10 dianas codificantes y/o dianas no codificantes. En algunas modalidades, el conjunto de sondas comprende una pluralidad de secuencias diana que se hibridan con al menos aproximadamente 15 dianas codificantes y/o dianas no codificantes. En algunas modalidades, el conjunto de sondas comprende una pluralidad de secuencias diana que se hibridan con al menos aproximadamente 20 dianas codificantes y/o dianas no codificantes. En algunas modalidades, el conjunto de sondas comprende una pluralidad de secuencias diana que se hibridan con al menos alrededor de 30 dianas codificantes y/o dianas no codificantes.

El sistema de la presente descripción proporciona además cebadores y pares de cebadores capaces de amplificar secuencias diana definidas por el conjunto de sondas, o fragmentos o subsecuencias o complementos de las mismas. Las secuencias de nucleótidos del conjunto de sondas pueden proporcionarse en medios legibles por computadora para aplicaciones *in silico* y como base para el diseño de cebadores apropiados para la amplificación de una o más secuencias diana del conjunto de sondas.

Se pueden diseñar cebadores basados en las secuencias de nucleótidos de secuencias diana para su uso en la amplificación de las secuencias diana. Para su uso en reacciones de amplificación como la PCR, se pueden utilizar un par de cebadores. La composición exacta de las secuencias de los cebadores no es crítica para la divulgación, pero para la mayoría de las aplicaciones los cebadores pueden hibridarse con secuencias específicas del conjunto de sondas en condiciones estrictas, particularmente en condiciones de alta rigurosidad, como se conoce en la técnica. Los pares de cebadores normalmente se eligen de manera que generen un producto de amplificación de al menos aproximadamente 50 nucleótidos, más habitualmente al menos aproximadamente 100 nucleótidos.

Los algoritmos para la selección de secuencias de cebadores son generalmente conocidos y están disponibles en paquetes de software comerciales. Estos cebadores se pueden utilizar en ensayos estándar cuantitativos o cualitativos basados en PCR para evaluar los niveles de expresión de transcripción de los ARN definidos por el conjunto de sondas. Alternativamente, estos cebadores se pueden usar en combinación con sondas, como balizas moleculares en amplificaciones que utilizan PCR en tiempo real.

En una modalidad, los cebadores o pares de cebadores, cuando se usan en una reacción de amplificación, amplifican específicamente al menos una porción de una secuencia de ácido nucleico de una diana (o subgrupos de la misma como se establece en el presente documento), una forma de ARN de la misma o un complemento para cualquiera de los dos.

Opcionalmente, se puede unir o incorporar un marcador a una sonda o polinucleótido cebador para permitir la detección y/o cuantificación de un polinucleótido diana que representa la secuencia diana de interés. El polinucleótido diana puede ser el propio ARN de la secuencia diana expresada, una copia de ADNc del mismo o un producto de amplificación derivado del mismo, y puede ser la cadena positiva o negativa, siempre que pueda detectarse específicamente en el ensayo que se utiliza. De manera similar, se puede marcar un anticuerpo.

En ciertos formatos multiplex, los marcadores utilizados para detectar diferentes dianas pueden distinguirse. El marcador se puede unir directamente (por ejemplo, mediante un enlace covalente) o indirectamente, por ejemplo, mediante una molécula puente o una serie de moléculas (por ejemplo, una molécula o complejo que puede unirse a un componente de ensayo, o mediante miembros de un par de unión que puede incorporarse a los componentes del ensayo, por ejemplo, biotina-avidina

o estreptavidina). Muchos marcadores están disponibles comercialmente en formas activadas que se pueden usar fácilmente para dicha conjugación (por ejemplo mediante acilación de amina), o se pueden unir marcadores mediante esquemas de conjugación conocidos o determinables, muchos de los cuales se conocen en la técnica.

5 Los marcadores útiles en la divulgación descrita en el presente documento incluyen cualquier sustancia que pueda detectarse cuando se une o se incorpora a la biomolécula de interés. Puede usarse cualquier método de detección eficaz, incluido óptico, espectroscópico, eléctrico, piezoeléctrico, magnético, dispersión Raman, resonancia de plasmón superficial, colorimétrico, calorimétrico, etc. Un marcador se selecciona típicamente entre un cromóforo, un lumíforo, un

10 fluoróforo, un miembro de un sistema de extinción, un cromógeno, un hapteno, un antígeno, una partícula magnética, un material que presenta una óptica no lineal, un nanocrystal semiconductor, una nanopartícula metálica, una enzima, un anticuerpo o porción de unión o equivalente del mismo, un aptámero y un miembro de un par de unión, y combinaciones de los mismos. Se pueden usar esquemas de extinción, en donde se pueden usar un inactivador y un fluoróforo como miembros de

15 un par de extinción en una sonda, de modo que se produzca un cambio en los parámetros ópticos al unirse a la diana, introducir o extinguir la señal del fluoróforo. Un ejemplo de tal sistema es una baliza molecular. En la técnica se conocen sistemas desactivadores/fluoróforos adecuados. El marcador puede estar unido a través de una variedad de enlaces intermedios. Por ejemplo, un polinucleótido puede comprender una especie que se une a biotina, y un marcador ópticamente

20 detectable puede conjugarse con biotina y luego unirse al polinucleótido marcado. De manera similar, un sensor de polinucleótidos puede comprender una especie inmunológica tal como un anticuerpo o fragmento, y se puede agregar un anticuerpo secundario que contenga un marcador ópticamente detectable.

25 Los cromóforos útiles en los métodos descritos en el presente documento incluyen cualquier sustancia que pueda absorber energía y emitir luz. Para ensayos multiplexados, se puede utilizar una pluralidad de cromóforos de señalización diferentes con espectros de emisión detectablemente diferentes. El cromóforo puede ser un lumóforo o un fluoróforo. Los fluoróforos típicos incluyen tintes fluorescentes, nanocristales semiconductores, quelatos de lantánidos, tintes específicos de polinucleótidos y proteínas fluorescentes verdes.

30 En algunas modalidades, los polinucleótidos de la divulgación comprenden al menos 20 bases consecutivas de la secuencia de ácido nucleico de una diana o un complemento de la misma. Los polinucleótidos pueden comprender al menos 21, 22, 23, 24, 25, 27, 30, 32, 35, 40, 45, 50 o más bases consecutivas de la secuencia de ácidos nucleicos de una diana.

35 Los polinucleótidos pueden proporcionarse en una variedad de formatos, incluidos sólidos, en solución o en una matriz. Los polinucleótidos pueden comprender opcionalmente uno o más

marcadores, que pueden incorporarse química y/o enzimáticamente al polinucleótido.

En algunas modalidades, uno o más polinucleótidos proporcionados en el presente documento se pueden proporcionar sobre un sustrato. El sustrato puede comprender una amplia gama de materiales, ya sean biológicos, no biológicos, orgánicos, inorgánicos o una combinación de cualquiera de estos. Por ejemplo, el sustrato puede ser una película de Langmuir Blodgett polimerizada, vidrio funcionalizado, Si, Ge, GaAs, GaP, SiO₂, SiN₄, silicio modificado o cualquiera de una amplia variedad de geles o polímeros tales como (poli)tetrafluoroetileno, (poli)difluoruro de vinilideno, poliestireno, poliestireno reticulado, poliacrílico, ácido poliláctico, ácido poliglicólico, poli(lactida coglicolida), polianhídridos, poli(metacrilato de metilo), poli(etilen-co-acetato de vinilo), polisiloxanos, sílice polimérica, látex, polímeros de dextrano, epoxis, policarbonatos o combinaciones de los mismos. Se pueden utilizar polímeros conductores y materiales fotoconductores.

El sustrato puede tomar la forma de una matriz, un fotodiodo, un sensor optoelectrónico tal como un chip semiconductor optoelectrónico o un semiconductor optoelectrónico de película delgada, o un biochip. Las ubicaciones de las sondas sobre el sustrato pueden ser direccionables; esto se puede hacer en formatos muy densos y las ubicaciones pueden ser microdireccionables o nanodireccionables.

En determinadas modalidades, se recoge una muestra (por ejemplo, muestra biológica) que contiene células de cáncer de mama de un sujeto que necesita tratamiento para el cáncer para evaluar si el sujeto tiene un riesgo bajo o alto de recurrencia del cáncer en función de un nivel de expresión o perfil de expresión. y probabilidad de beneficiarse de la radioterapia adyuvante. Las muestras de diagnóstico para usar con los sistemas y en los métodos de la presente divulgación comprenden ácidos nucleicos adecuados para proporcionar información de expresión de ARN. En principio, la muestra biológica a partir de la cual se obtiene el ARN expresado y se analiza para determinar la expresión del gen diana puede ser cualquier material que se sospeche que comprende tejido o células mamarias cancerosas. La muestra de diagnóstico puede ser una muestra biológica utilizada directamente en un método de la divulgación. Alternativamente, la muestra de diagnóstico puede ser una muestra preparada a partir de una muestra biológica.

En una modalidad, la muestra o porción de la muestra que comprende o se sospecha que comprende tejido o células cancerosas puede ser cualquier fuente de material biológico, incluidas células, tejido o fluido, incluidos fluidos corporales. Los ejemplos no limitantes de la fuente de la muestra incluyen un aspirado, una biopsia con aguja, un sedimento de citología, una preparación de tejido en masa o una sección del mismo obtenida, por ejemplo, mediante cirugía o autopsia, líquido linfático, sangre, plasma, suero, tumores y órganos. En algunas modalidades, la muestra proviene de una biopsia de tumor de mama.

Las muestras pueden ser muestras de archivo, que tengan un resultado médico conocido y documentado, o pueden ser muestras de sujetos actuales cuyo resultado médico final aún no se conoce.

En algunas modalidades, la muestra se puede diseccionar antes del análisis molecular. La muestra se puede preparar mediante macrodissección de una muestra tumoral en masa o de una porción de la misma, o se puede tratar mediante microdissección, por ejemplo mediante microdissección por captura láser (LCM).

La muestra puede proporcionarse inicialmente en una variedad de estados, como tejido fresco, tejido fresco congelado, aspirados con aguja fina, y puede estar fijada o no fijada. Con frecuencia, los laboratorios médicos preparan de forma rutinaria muestras médicas en un estado fijo, lo que facilita el almacenamiento de tejidos. Se pueden usar diversos fijadores para fijar tejido y estabilizar la morfología de las células, y se pueden usar solos o en combinación con otros agentes. Los fijadores ilustrativos incluyen agentes reticulantes, alcoholes, acetona, solución de Bouin, solución de Zenker, solución de Hely, solución de ácido ósmico y solución de Carnoy.

Los fijadores reticulantes pueden comprender cualquier agente adecuado para formar dos o más enlaces covalentes, por ejemplo un aldehído. Las fuentes de aldehídos que normalmente se utilizan para la fijación incluyen formaldehído, paraformaldehído, glutaraldehído o formalina. En algunas modalidades, el agente reticulante comprende formaldehído, que puede incluirse en su forma nativa o en forma de paraformaldehído o formalina. Un experto en la técnica apreciaría que para muestras en las que se han utilizado fijadores reticulantes pueden ser necesarios pasos preparatorios especiales que incluyen, por ejemplo, pasos de calentamiento y digestión con proteinasa-k; véanse los métodos.

Se pueden usar uno o más alcoholes para fijar el tejido, solos o en combinación con otros fijadores. Los alcoholes ilustrativos usados para la fijación incluyen metanol, etanol e isopropanol.

La fijación con formalina se utiliza frecuentemente en laboratorios médicos. La formalina comprende tanto un alcohol, típicamente metanol, como formaldehído, los cuales pueden actuar para fijar una muestra biológica.

Ya sea fijada o no, la muestra biológica puede opcionalmente incluirse en un medio de inclusión. Medios de inclusión ilustrativos utilizados en histología que incluyen parafina, Tissue-Tek® V.I.P., Paramat, Paramat Extra, Paraplast, Paraplast X-tra, Paraplast Plus, cera de inclusión de parafina Peel Away, cera de poliéster, polietilenglicol Carbowax, Polyfin, medio de congelación de tejidos TFMFM, Cryo-Gef y compuesto OCT (Ciencias de microscopía electrónica, Hatfield, PA). Antes del análisis molecular, el material de inclusión se puede eliminar mediante cualquier técnica adecuada, como se conoce en la técnica. Por ejemplo, cuando la muestra está embebida en cera, el material de inclusión puede eliminarse mediante extracción con disolventes orgánicos, por ejemplo xilenos.

Hay kits disponibles comercialmente para eliminar medios de inclusión de tejidos. Las muestras o secciones de las mismas pueden someterse a pasos de procesamiento adicionales según sea necesario, por ejemplo pasos de hidratación o deshidratación en serie.

En algunas modalidades, la muestra es una muestra biológica fijada e incluida en cera. Con frecuencia, las muestras de laboratorios médicos se proporcionan como muestras fijadas e incluidas en cera, más comúnmente como tejidos fijados en formalina e incluidos en parafina (FFPE).

Cualquiera que sea la fuente de la muestra biológica, el polinucleótido diana que finalmente se analiza se puede preparar sintéticamente (en el caso de secuencias de control), pero normalmente se purifica a partir de la fuente biológica y se somete a una o más etapas preparativas. El ARN puede purificarse para eliminar o disminuir uno o más componentes no deseados de la muestra biológica o para concentrarlo. Por el contrario, cuando el ARN está demasiado concentrado para un ensayo concreto, puede diluirse.

El ARN se puede extraer y purificar de muestras biológicas utilizando cualquier técnica adecuada. Se conocen varias técnicas en la técnica y varias están disponibles comercialmente (por ejemplo, kit de extracción de ácido nucleico FormaPure, Agencourt Biosciences, Beverly MA, High Pure FFPE RNA Micro Kit, Roche Applied Science, Indianápolis, IN). El ARN se puede extraer de secciones de tejido congeladas usando TRIzol (Invitrogen, Carlsbad, CA) y purificarse usando el kit RNeasy Protect (Qiagen, Valencia, CA). El ARN se puede purificar aún más mediante el tratamiento con ADNasa I (Ambion, Austin, TX) para eliminar cualquier ADN contaminante. Las concentraciones de ARN se pueden realizar utilizando un espectrofotómetro Nanodrop ND-1000 (Nanodrop Technologies, Rockland, DE). El ARN se puede purificar aún más para eliminar los contaminantes que interfieren con la síntesis de ADNc mediante precipitación en frío con acetato de sodio. La integridad del ARN se puede evaluar ejecutando electroferogramas y el número de integridad del ARN (RIN, una medida correlativa que indica la integridad del ARNm) se puede determinar utilizando el RNA 6000 PicoAssay para el Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA).

También se proporcionan kits para realizar los métodos deseados, y comprenden un recipiente o alojamiento para contener los componentes del kit, uno o más recipientes que contienen uno o más ácidos nucleicos y, opcionalmente, uno o más recipientes que contienen uno o más. más reactivos. Los reactivos incluyen los descritos en la sección de composición de materia anterior y los reactivos útiles para realizar los métodos descritos, que incluyen reactivos de amplificación, y pueden incluir una o más sondas, cebadores o pares de cebadores, enzimas (que incluyen polimerasas y ligasas), colorantes intercalantes, sondas marcadas y marcadores que se pueden incorporar en los productos de amplificación.

En algunas modalidades, el kit comprende cebadores o pares de cebadores específicos para aquellos subconjuntos y combinaciones de secuencias diana descritas en el presente documento.

Los cebadores o pares de cebadores son adecuados para amplificar selectivamente las secuencias diana. El kit puede comprender al menos dos, tres, cuatro o cinco cebadores o pares de cebadores adecuados para amplificar selectivamente una o más dianas. El kit puede comprender al menos 5, 10, 15, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110 o más cebadores o pares de cebadores adecuados para amplificar selectivamente una o más dianas.

En algunas modalidades, los cebadores o pares de cebadores del kit, cuando se usan en una reacción de amplificación, amplifican específicamente una diana no codificante, diana codificante, diana exónica o no exónica descrita en la presente, una secuencia de ácido nucleico correspondiente a una diana seleccionada de la Tabla 5, una forma de ARN de esta o un complemento de cualquiera de estas. El kit puede incluir una pluralidad de dichos cebadores o pares de cebadores que pueden amplificar específicamente una pluralidad correspondiente de diferentes amplificaciones de una diana no codificante, diana codificante, transcrito exónico o no exónico descrito en la presente, una secuencia de ácido nucleico correspondiente a una diana seleccionada de la Tabla 5, formas de ARN de esta o complementos de esta. Se pueden proporcionar en forma de kit al menos dos, tres, cuatro o cinco cebadores o pares de cebadores adecuados para amplificar selectivamente una o más dianas. En algunas modalidades, el kit comprende de cinco a cincuenta cebadores o pares de cebadores adecuados para amplificar una o más dianas.

Los reactivos pueden estar independientemente en forma líquida o sólida. Los reactivos pueden proporcionarse en mezclas. Opcionalmente, en el kit se pueden proporcionar muestras de control y/o ácidos nucleicos. Las muestras de control pueden incluir tejido y/o ácidos nucleicos obtenidos o representativos de muestras tumorales de sujetos que no muestran evidencia de enfermedad, así como tejido y/o ácidos nucleicos obtenidos o representativos de muestras tumorales de sujetos que desarrollan cáncer sistémico.

Los ácidos nucleicos se pueden proporcionar en un formato de matriz y, por tanto, se puede incluir una matriz o micromatriz en el kit. Opcionalmente, el kit puede estar certificado por una agencia gubernamental para su uso en el pronóstico del resultado de la enfermedad de sujetos con cáncer y/o para designar una modalidad de tratamiento.

Las instrucciones para usar el kit para realizar uno o más métodos de la divulgación se pueden proporcionar con el contenedor y se pueden proporcionar en cualquier medio fijo. Las instrucciones pueden estar ubicadas dentro o fuera del contenedor o alojamiento, y/o pueden estar impresas en el interior o exterior de cualquier superficie del mismo. Un kit puede estar en forma múltiple para detectar y/o cuantificar simultáneamente uno o más polinucleótidos diana diferentes que representan los genes diana expresados.

Después de la recolección de la muestra y la extracción del ácido nucleico, la porción de ácido nucleico de la muestra que comprende ARN que se usa o se puede usar para preparar el

polinucleótido diana de interés se puede someter a una o más reacciones preparativas. Estas reacciones preparativas pueden incluir transcripción *in vitro* (IVT), marcaje, fragmentación, amplificación y otras reacciones. El ARNm puede tratarse primero con transcriptasa inversa y un cebador para crear ADNc antes de la detección, cuantificación y/o amplificación; esto se puede hacer

in vitro con ARNm purificado o *in situ*, por ejemplo, en células o tejidos fijados a un portaobjetos. Por "amplificación" se entiende cualquier proceso de producción de al menos una copia de un ácido nucleico, como un ARN expresado, y en muchos casos produce múltiples copias. Un producto de amplificación puede ser ARN o ADN y puede incluir una cadena complementaria a la secuencia diana expresada. Los productos de amplificación de ADN se pueden producir inicialmente mediante traducción inversa y luego, opcionalmente, a partir de reacciones de amplificación adicionales. El producto de amplificación puede incluir toda o una parte de una secuencia diana y, opcionalmente, puede estar marcado. Son adecuados para su uso una variedad de métodos de amplificación, incluidos métodos basados en polimerasa y métodos basados en ligación. Las técnicas de amplificación ilustrativas incluyen el método de reacción en cadena de la polimerasa (PCR), la reacción en cadena de la lipasa (LCR), métodos basados en ribozimas, replicación de secuencia autosostenida (3SR), amplificación basada en secuencia de ácido nucleico (NASBA), el uso de Q Beta replicasa, transcripción inversa, traducción de mellas y similares.

Se pueden usar reacciones de amplificación asimétricas para amplificar preferentemente una hebra que representa la secuencia diana que se usa para la detección como polinucleótido diana. En algunos casos, la presencia y/o cantidad del propio producto de amplificación puede usarse para determinar el nivel de expresión de una secuencia diana determinada. En otros casos, el producto de amplificación se puede usar para hibridar con una matriz u otro sustrato que comprende polinucleótidos sensores que se usan para detectar y/o cuantificar la expresión de la secuencia diana.

El primer ciclo de amplificación en métodos basados en polimerasa normalmente forma un producto de extensión del cebador complementario a la cadena plantilla. Si la plantilla es ARN monocatenario, se utiliza una polimerasa con actividad transcriptasa inversa en la primera amplificación para transcribir inversamente el ARN a ADN, y se pueden realizar ciclos de amplificación adicionales para copiar los productos de extensión del cebador. Por supuesto, los cebadores para una PCR deben diseñarse para hibridar con regiones en su plantilla correspondiente que puedan producir un segmento amplificable; por lo tanto, cada cebador debe hibridarse de modo que su nucleótido 3' se aparee con un nucleótido en su cadena plantilla complementaria que se encuentra en 3' del nucleótido 3' del cebador utilizado para replicar esa cadena plantilla complementaria en la PCR.

El polinucleótido diana se puede amplificar poniendo en contacto una o más hebras del polinucleótido diana con un cebador y una polimerasa que tenga actividad adecuada para extender

el cebador y copiar el polinucleótido diana para producir un polinucleótido complementario de longitud completa o una porción más pequeña del mismo. Puede usarse cualquier enzima que tenga una actividad polimerasa que pueda copiar el polinucleótido diana, incluyendo ADN polimerasas, ARN polimerasas, transcriptasas inversas, enzimas que tengan más de un tipo de polimerasa o actividad enzimática. La enzima puede ser termolábil o termoestable. También se pueden utilizar mezclas de enzimas. Los ejemplos de enzimas incluyen: ADN polimerasas tales como ADN polimerasa I ("Pol I"), el fragmento Klenow de Pol I, T4, T7, Sequenase® T7, Sequenase® Versión 2.0 T7, *Tub*, *Taq*, *Tth*, *Pfic*, *Pfu*, *Tsp*, *Tfl*, *Tli* y *Pyrococcus sp* ADN GB-D polimerasas; ARN polimerasas tales como polimerasas *E. coli*, SP6, T3 y ARN T7; y transcriptasas inversas tales como AMV, M-MuLV, MMLV, ARNasa H MMLV (SuperScript®), SuperScript® II, ThermoScript®, VIH-1 y transcriptasas inversas RAV2. Todas estas enzimas están disponibles comercialmente. Las polimerasas ilustrativas con múltiples especificidades incluyen RAV2 y *Tli* (exo-)polimerasas. Las polimerasas termoestables ilustrativas incluyen *Tub*, *Taq*, *Tth*, *Pfic*, *Pfu*, *Tsp*, *Tf1*, *Tli* y *Pyrococcus sp*. ADN polimerasas GB-D.

Se eligen condiciones de reacción adecuadas para permitir la amplificación del polinucleótido diana, incluyendo pH, amortiguador, fuerza iónica, presencia y concentración de una o más sales, presencia y concentración de reactivos y cofactores tales como nucleótidos y magnesio y/u otros iones metálicos (p. ej., manganeso), codisolventes opcionales, temperatura, perfil de ciclo térmico para esquemas de amplificación que comprenden una reacción en cadena de la polimerasa, y puede depender en parte de la polimerasa que se utilice, así como de la naturaleza de la muestra. Los codisolventes incluyen formamida (típicamente de aproximadamente 2 a aproximadamente 10 %), glicerol (típicamente de aproximadamente 5 a aproximadamente 10 %) y DMSO (típicamente de aproximadamente 0,9 a aproximadamente 10 %). Se pueden utilizar técnicas en el esquema de amplificación para minimizar la producción de falsos positivos o artefactos producidos durante la amplificación. Estas incluyen la PCR de "anotación", técnicas de arranque en caliente, uso de cebadores anidados o diseño de cebadores de PCR para que formen estructuras de bucle en el caso de formación de cebador-dímero y, por lo tanto, no se amplifiquen. Se pueden utilizar técnicas para acelerar la PCR, por ejemplo la PCR centrífuga, que permite una mayor convección dentro de la muestra y que comprende etapas de calentamiento por infrarrojos para un calentamiento y enfriamiento rápidos de la muestra. Se pueden realizar uno o más ciclos de amplificación. Puede usarse un exceso de un cebador para producir un exceso de un producto de extensión del cebador durante la PCR. En algunas modalidades, el producto de extensión del cebador producido en exceso es el producto de amplificación que se va a detectar. Se puede usar una pluralidad de cebadores diferentes para amplificar diferentes polinucleótidos diana o diferentes regiones de un polinucleótido diana particular dentro de la muestra.

Se puede realizar una reacción de amplificación en condiciones que permitan que un polinucleótido sensor opcionalmente marcado se hibride con el producto de amplificación durante al menos parte de un ciclo de amplificación. Cuando el ensayo se realiza de esta manera, puede tener lugar la detección en tiempo real de este evento de hibridación controlando la emisión de luz o la fluorescencia durante la amplificación, como se conoce en la técnica.

Cuando el producto de amplificación se va a utilizar para la hibridación con una matriz o micromatriz, están disponibles varios productos de amplificación adecuados disponibles comercialmente. Estos incluyen kits de amplificación disponibles de NuGEN, Inc. (San Carlos, CA), incluidos el sistema WT-OvationTm, el sistema WT-OvationTm v2, el sistema WT-OvationTm Pico, el módulo de exón FFPE WT-OvationTm, el módulo de exón FFPE WT-OvationTm RiboAmp. y kits de amplificación de ARN RiboAmp^{Plus} (MDS Analytical Technologies (anteriormente Arcturus) (Mountain View, CA), Genisphere, Inc. (Hatfield, PA), incluidos los kits de amplificación de ARN RampUp Plus y SenseAmp, solos o en combinación. Los ácidos nucleicos amplificados pueden someterse a una o más reacciones de purificación después de la amplificación y el marcaje, por ejemplo usando microesferas magnéticas (por ejemplo, microesferas magnéticas RNAClean, Agencourt Biosciences).

Se pueden analizar múltiples biomarcadores de ARN utilizando plataformas de RT-PCR multiplex cuantitativa en tiempo real y otras tecnologías de multiplexación como el sistema de análisis genético GenomeLab GeXP (Beckman Coulter, Foster City, CA), SmartCycler® 9600 o GeneXpert® Systems (Cepheid, Sunnyvale, CA), sistema de PCR en tiempo real rápido ABI 7900 HT (Applied Biosystems, Foster City, CA), sistema LightCycler® 480 (Roche Molecular Systems, Pleasanton, CA), sistema xMAP 100 (Luminex, Austin, TX) Sistema de análisis del genoma Solexa (Illumina, Hayward, CA), OpenArray Real Time qPCR (BioTrove, Woburn, MA) y BeadXpress System (Illumina, Hayward, CA).

En principio, en la divulgación se puede utilizar cualquier método para detectar y/o cuantificar la expresión de los genes diana codificados. Los genes diana expresados pueden detectarse y/o cuantificarse directamente, o pueden copiarse y/o amplificarse para permitir la detección de copias amplificadas de los genes diana expresados.

Los métodos para detectar y/o cuantificar un gen diana pueden incluir transferencia Northern, secuenciación, hibridación en matriz o micromatriz, mediante escisión enzimática de estructuras específicas (por ejemplo, un ensayo Clariom S, ThermoFisher Scientific, un ensayo Invader®, Third Wave Technologies, por ejemplo como descritos en las patentes de EE. UU. No. 5,846,717, 6,090,543; 6,001,567; 5,985,557; y 5,994,069) y métodos de amplificación, por ejemplo RT-PCR, incluyendo un ensayo TaqMan® (PE Biosystems, Foster City, California, por ejemplo como se describe en las patentes de EE. UU. No. 5,962,233 y 5,538,848), y pueden ser cuantitativos o

semicuantitativos, y pueden variar dependiendo del origen, cantidad y condición de la muestra biológica disponible. También se pueden utilizar combinaciones de estos métodos. Por ejemplo, los ácidos nucleicos pueden amplificarse, marcarse y someterse a análisis de micromatrices. Los métodos para detectar y/o cuantificar un gen diana pueden incluir análisis de expresión a nivel genético de genes anotados usando hibridación de micromatrices (por ejemplo, ensayo GeneChip Human Exon 1.0 ST o ensayo Clariom S, ThermoFisher Scientific).

En algunos casos, los genes diana pueden detectarse mediante secuenciación. Los métodos de secuenciación pueden comprender secuenciación del genoma completo o secuenciación del exoma. También se pueden utilizar métodos de secuenciación como Maxim-Gilbert, terminación de cadena o sistemas de alto rendimiento. Las técnicas de secuenciación adecuadas adicionales incluyen reacciones de secuenciación dideoxi clásicas (método de Sanger) usando terminadores o cebadores marcados y separación en gel en placa o capilar, secuenciación mediante síntesis usando nucleótidos marcados terminados de forma reversible, pirosecuenciación, secuenciación 454, hibridación específica de alelo a una biblioteca de sondas oligonucleotídicas marcadas, secuenciación mediante síntesis usando hibridación específica de alelo a una biblioteca de clones marcados que va seguida de ligación, control en tiempo real de la incorporación de nucleótidos marcados durante una etapa de polimerización y secuenciación SOLiD.

Los métodos adicionales para detectar y/o cuantificar un gen diana incluyen secuenciación de una sola molécula (por ejemplo, Helicos, PacBio), secuenciación por síntesis (por ejemplo, Illumina, Ion Torrent), secuenciación por ligación (por ejemplo, ABI SOLiD), secuenciación por hibridación (por ejemplo, Complete Genomics), hibridación in situ, tecnologías de matriz de perlas (por ejemplo, Luminex xMAP, Illumina BeadChips), tecnología de ADN ramificado (por ejemplo, Panomics, Genisphere). Los métodos de secuenciación pueden utilizar métodos fluorescentes (p. ej., Illumina) o electrónicos (p. ej., Ion Torrent, Oxford Nanopore) para detectar nucleótidos.

La transcripción inversa se puede realizar mediante cualquier método conocido en la técnica. Por ejemplo, la transcripción inversa se puede realizar usando el kit Omniscript (Qiagen, Valencia, CA), el kit Superscript III (Invitrogen, Carlsbad, CA), para RT-PCR. Se puede realizar un cebado específico de la diana para aumentar la sensibilidad de detección de genes diana y generar ADNc específico de la diana.

La RT-PCR TaqMan® se puede realizar utilizando instrumentos Applied Biosystems Prism (ABI) 7900 HT en un volumen de 5 μ l con ADNc específico del gen diana equivalente a 1 ng de ARN total.

Se añaden concentraciones de cebadores y sondas para el análisis TaqMan para amplificar amplicones fluorescentes utilizando condiciones de ciclo de PCR como 95 °C durante 10 minutos para un ciclo, 95 °C durante 20 segundos y 60 °C durante 45 segundos durante 40 ciclos. Se puede

analizar una muestra de referencia para garantizar la estabilidad del reactivo y del proceso. Se deben analizar controles negativos (p. ej., sin plantilla) para controlar cualquier contaminación exógena de ácido nucleico.

La presente divulgación contempla que un conjunto de sondas o sondas derivadas del mismo pueden proporcionarse en un formato de matriz. En el contexto de la presente comunicación, una "matriz" es una colección de sondas polinucleótidas organizada espacial o lógicamente. Puede usarse una matriz que comprende sondas específicas para una diana codificante, una diana no codificante o una combinación de las mismas. Alternativamente, se puede utilizar una matriz que comprende sondas específicas para dos o más transcritos de una diana, o un producto derivado de la misma. Deseablemente, una matriz puede ser específica para 5, 10, 15, 20, 25, 30 o más transcritos de un gen diana. La expresión de estos genes puede detectarse sola o en combinación con otros transcritos. En algunas modalidades, se utiliza una matriz que comprende una amplia gama de sondas sensoras para productos de expresión específicos de mama, junto con secuencias de control apropiadas. En algunos casos, la matriz puede comprender la matriz Human Exon 1.0 ST (HuEx 1.0 ST, Thermo Fisher Scientific, Santa Clara, CA.).

Normalmente, las sondas polinucleotídicas están unidas a un sustrato sólido y están ordenadas de manera que se conozca la ubicación (en el sustrato) y la identidad de cada una. Las sondas polinucleotídicas se pueden unir a uno de una variedad de sustratos sólidos capaces de resistir los reactivos y las condiciones necesarias para el uso de la matriz. Los ejemplos incluyen, pero no se limitan a, polímeros, tales como (poli)tetrafluoroetileno, (poli)difluoruro de vinilideno, poliestireno, policarbonato, polipropileno y poliestireno; cerámico; silicio; dióxido de silicio; silicio modificado; sílice (fundida), cuarzo o vidrio; vidrio funcionalizado; papel, tal como papel de filtro; celulosa diazotizada; filtro de nitrocelulosa; membrana de nailon; y almohadilla de gel de poliacrilamida. Los sustratos que son transparentes a la luz son útiles para matrices que pueden usarse en un ensayo que implica detección óptica.

Los ejemplos de formatos de matriz incluyen matrices de membrana o filtro (por ejemplo, matrices de nitrocelulosa, matrices de nylon), matrices de placas (por ejemplo, matrices de placas de múltiples pocillos, tales como matrices de placas de microtitulación de 24, 96, 256, 384, 864 o 1536 pocillos), matrices de clavijas y matrices de perlas (por ejemplo, en una "suspensión" líquida). Las matrices en sustratos tales como portaobjetos de vidrio o cerámica a menudo se denominan matrices de chips o "chips". Tales conjuntos son bien conocidos en la técnica. En una modalidad de la presente divulgación, la matriz de pronóstico de cáncer es un chip.

En algunas modalidades, se pueden usar uno o más métodos de reconocimiento de patrones para analizar el nivel de expresión de genes diana. El método de reconocimiento de patrones puede comprender una combinación lineal de niveles de expresión o una combinación no lineal de niveles

de expresión. En algunas modalidades, las mediciones de expresión para transcritos de ARN o combinaciones de niveles de transcritos de ARN se formulan en modelos o algoritmos lineales o no lineales (por ejemplo, una "expresión distintiva") y se convierten en una puntuación de probabilidad. Esta puntuación de probabilidad indica la probabilidad de que una muestra biológica provenga de un sujeto que puede no presentar evidencia de enfermedad, que puede presentar cáncer sistémico o que puede presentar recurrencia bioquímica o recurrencia locorregional. La puntuación de probabilidad se puede utilizar para distinguir estos estados patológicos. Los modelos y/o algoritmos pueden proporcionarse en formato legible por máquina y pueden usarse para correlacionar niveles de expresión o un perfil de expresión con un estado de enfermedad y/o para designar una modalidad de tratamiento para un sujeto o clase de sujetos.

El análisis del nivel de expresión para una pluralidad de genes diana puede comprender el uso de un algoritmo o clasificador. Los datos de la matriz se pueden gestionar, clasificar y analizar utilizando técnicas conocidas en la técnica. El ensayo del nivel de expresión para una pluralidad de dianas genéticas puede comprender modelado de conjuntos de sondas y preprocesamiento de datos. El modelado de conjuntos de sondas y el preprocesamiento de datos se pueden derivar utilizando el algoritmo Robust Multi-Array (RMA) o sus variantes GC-RMA, *f*RMA, el algoritmo Probe Logarithmic Intensity Error (PLIER) o su variante iterPLIER. Se pueden aplicar filtros de varianza o intensidad para preprocesar datos utilizando el algoritmo RMA, por ejemplo, eliminando genes diana con una desviación estándar de <10 o una intensidad media de <100 unidades de intensidad de un rango de datos normalizado, respectivamente.

Alternativamente, analizar el nivel de expresión de una pluralidad de dianas genéticas puede comprender el uso de un algoritmo de aprendizaje automático. El algoritmo de aprendizaje automático puede comprender un algoritmo de aprendizaje supervisado. Los ejemplos de algoritmos de aprendizaje supervisado pueden incluir estimadores de dependencia única promedio (AODE), red neuronal artificial (por ejemplo, retropropagación), estadísticas bayesianas (por ejemplo, clasificador Naive Bayes, red bayesiana, base de conocimiento bayesiano), razonamiento basado en casos, árboles de decisión, programación lógica inductiva, regresión de procesos gaussianos, método de manejo de datos de grupo (GMDH), autómatas de aprendizaje, cuantificación vectorial de aprendizaje, longitud mínima del mensaje (árboles de decisión, gráficos de decisión, etc.), aprendizaje vago, aprendizaje basado en instancias de vecinos más cercanos, modelado analógico, aprendizaje probablemente aproximadamente correcto (PAC), reglas de Ripple down, una metodología de adquisición de conocimiento, algoritmos simbólicos de aprendizaje automático, algoritmos subsimbólicos de aprendizaje automático, máquinas de vectores de soporte, bosques aleatorios, conjuntos de clasificadores, agregación Bootstrap (embolsado) e impulso. El aprendizaje supervisado puede comprender clasificación ordinal, como análisis de regresión y redes difusas de

información (IFN). Alternativamente, los métodos de aprendizaje supervisado pueden comprender clasificación estadística, como AODE, clasificadores lineales (por ejemplo, discriminante lineal de Fisher, regresión logística, clasificador Naive Bayes, máquina de vectores Perceptron y Support, modelos lineales generales regularizados Lasso y Elastic-Net), clasificadores cuadráticos, vecino k más cercano, Boosting, árboles de decisión (por ejemplo, C4,5, bosques aleatorios), redes bayesianas y modelos ocultos de Markov.

Los algoritmos de aprendizaje automático también pueden comprender un algoritmo de aprendizaje no supervisado. Los ejemplos de algoritmos de aprendizaje no supervisados pueden incluir redes neuronales artificiales, agrupamiento de datos, algoritmo de maximización de expectativas, mapa de autoorganización, red de funciones de base radial, cuantificación vectorial, mapa topográfico generativo, método de cuello de botella de información e IBSEAD. El aprendizaje no supervisado también puede comprender algoritmos de aprendizaje de reglas de asociación, como el algoritmo Apriori, el algoritmo Eclat y el algoritmo de crecimiento FP. También se pueden utilizar agrupaciones jerárquicas, como agrupaciones de vínculo único y agrupaciones conceptuales. Alternativamente, el aprendizaje no supervisado puede comprender agrupamiento particional como el algoritmo *K-means* y el agrupamiento difuso.

En algunos casos, los algoritmos de aprendizaje automático comprenden un algoritmo de aprendizaje por refuerzo. Los ejemplos de algoritmos de aprendizaje de refuerzo incluyen, entre otros, aprendizaje de diferencia temporal, Q-learning y Learning Automata. Alternativamente, el algoritmo de aprendizaje automático puede comprender un preprocesamiento de datos.

En algunas modalidades, los algoritmos de aprendizaje automático pueden incluir, de modo no taxativo, estimadores de dependencia única promedio (AODE), discriminante lineal de Fisher, regresión logística, perceptrón, perceptrón multicapa, redes neuronales artificiales, máquinas de vectores de soporte, clasificadores cuadráticos, refuerzo, árboles de decisión, C4,5, redes bayesianas, modelos ocultos de Markov, análisis discriminante de alta dimensión y modelos de mezcla gaussiana. El algoritmo de aprendizaje automático puede comprender máquinas de vectores de soporte, clasificador Naïve Bayes, k-vecino más cercano, análisis discriminante de alta dimensión o modelos de mezcla gaussiana. En algunos casos, el algoritmo de aprendizaje automático comprende bosques aleatorios.

La subtipificación molecular es un método para clasificar el cáncer de mama en una de múltiples categorías o subtipos genéticamente distintos. Cada subtipo responde de manera diferente a diferentes tipos de tratamientos, y la presencia de un subtipo particular predice, por ejemplo, radiorresistencia o quimiorresistencia, mayor o menor riesgo de recurrencia, o buen o mal pronóstico para un individuo. El análisis de expresión diferencial de una pluralidad de dianas genéticas enumeradas en la Tabla 5 permite la identificación de sujetos con bajo riesgo de recurrencia del

cáncer que, por ejemplo, pueden beneficiarse menos de la radioterapia adyuvante. En algunos casos, los métodos de subtipificación molecular de la presente divulgación se usan en combinación con otros biomarcadores, como el grado del tumor y los niveles hormonales, para analizar el cáncer de mama. Por ejemplo, un sujeto con cáncer de mama con receptor de estrógeno positivo (ER+), receptor del factor de crecimiento epidérmico humano 2 negativo (HER2-), cáncer de mama con ganglios negativos, que es posmenopáusica, puede tener más probabilidades de tener un menor riesgo de recurrencia. (p. ej., recurrencia locorregional).

La recurrencia del cáncer es la reaparición del cáncer después de un período en donde no se detectan células cancerosas en el cuerpo. Después de la cirugía para el cáncer de mama operable, la enfermedad puede recurrir localmente, regionalmente y/o en sitios metastásicos distantes. Una recurrencia local es la reaparición del cáncer en la pared torácica ipsilateral. Por el contrario, una recurrencia regional denota un tumor que afecta a los ganglios linfáticos regionales, generalmente axilares o supraclaviculares ipsilaterales, y con menos frecuencia a los ganglios infraclaviculares y/o mamarios internos. Algunas pacientes con cáncer de mama tendrán recurrencia local o locorregional después de una cirugía de conservación de la mama y radioterapia dentro de los diez años posteriores al primer diagnóstico de cáncer de mama. Si se les extirpó la mama durante el tratamiento inicial, estas mujeres tendrán una recurrencia local en la axila o en la pared torácica dentro de diez años. En algunas modalidades, el sujeto tratado en los métodos de la presente divulgación tiene un cáncer de mama con ganglios negativos.

Diagnosticar, predecir o controlar un estado o resultado de un cáncer puede comprender tratar un cáncer o prevenir la progresión del cáncer. Además, diagnosticar, predecir o controlar un estado o resultado de un cáncer puede comprender identificar o predecir que un sujeto tiene un riesgo alto o bajo de recurrencia (por ejemplo, recurrencia locorregional). En algunos casos, diagnosticar, predecir o controlar puede comprender determinar un régimen terapéutico. La determinación de un régimen terapéutico puede comprender la administración de una terapia anticancerígena. Alternativamente, determinar un régimen terapéutico puede comprender modificar, recomendar, continuar o suspender un régimen anticancerígeno. Por ejemplo, un sujeto que se determina que tiene un riesgo bajo de recurrencia de cáncer de mama según el perfil de expresión, como se describe en el presente documento, puede evitar la radioterapia adyuvante. En algunos casos, un sujeto que se determina que tiene un alto riesgo de recurrencia de cáncer de mama según el perfil de expresión, como se describe en el presente documento, puede tratarse con mastectomía, refuerzo de radiación o terapia sistémica adyuvante. En algunos casos, si los patrones de expresión de la muestra son consistentes con el patrón de expresión de una enfermedad o resultado de una enfermedad conocida, los patrones de expresión se pueden usar para designar una o más modalidades de tratamiento (por ejemplo, regímenes terapéuticos, régimen anticancerígeno). Un

régimen anticancerígeno puede comprender una o más terapias anticancerígenas. Los ejemplos de terapias contra el cáncer incluyen terapia hormonal/endocrina, cirugía, quimioterapia, radioterapia, inmunoterapia/terapia biológica y terapia fotodinámica.

La terapia hormonal o la terapia endocrina pueden implicar la administración de hormonas, como hormonas esteroides o antagonistas hormonales, para modular los niveles de ciertas hormonas con el fin de detener el crecimiento o inducir la apoptosis de células cancerosas que responden a hormonas. Por ejemplo, el cáncer de mama puede tratarse con un modulador selectivo del receptor de estrógeno (SERM) tal como, entre otros, tamoxifeno, raloxifeno y toremifeno. De forma alternativa o adicional, se pueden usar inhibidores de la síntesis hormonal tales como inhibidores de la aromatasa, incluidos, entre otros, letrozol, anastrozol, exemestano y aminoglutetimida para tratar el cáncer de mama. En algunos casos, la suplementación hormonal con progestinas como, entre otras, acetato de megestrol y acetato de medroxiprogesterona, se puede utilizar para el tratamiento del cáncer de mama avanzado que responde a las hormonas. En particular, el cáncer de mama ER+ se puede tratar con un antagonista del receptor de estrógeno (p. ej., tamoxifeno) o un fármaco que bloquea la producción de estrógeno, como un inhibidor de la aromatasa (p. ej., anastrozol o letrozol). La terapia hormonal también puede incluir la extirpación quirúrgica de órganos endocrinos (p. ej., orquiectomía u ooforectomía).

La oncología quirúrgica utiliza métodos quirúrgicos para diagnosticar, estadificar y tratar el cáncer y para aliviar ciertos síntomas relacionados con el cáncer. La cirugía se puede utilizar para extirpar el tumor (p. ej., escisiones, resecciones, cirugía citorreductora), reconstruir una parte del cuerpo (p. ej., cirugía restauradora) y/o aliviar síntomas como el dolor (p. ej., cirugía paliativa). La cirugía también puede incluir criocirugía. La criocirugía (también llamada crioterapia) puede utilizar frío extremo producido por nitrógeno líquido (o gas argón) para destruir tejido anormal. La criocirugía se puede utilizar para tratar tumores externos, como los de la piel. Para los tumores externos, el nitrógeno líquido se puede aplicar directamente a las células cancerosas con un hisopo de algodón o un dispositivo rociador. La criocirugía también se puede utilizar para tratar tumores dentro del cuerpo (tumores internos y tumores en los huesos). Para los tumores internos, se puede hacer circular nitrógeno líquido o gas argón a través de un instrumento hueco llamado crioonda, que se coloca en contacto con el tumor. Se puede utilizar una ecografía o una resonancia magnética para guiar la crioonda y controlar la congelación de las células, limitando así el daño al tejido sano cercano. Se puede formar una bola de cristales de hielo alrededor de la sonda, congelando las células cercanas. A veces se utiliza más de una sonda para administrar nitrógeno líquido a varias partes del tumor. Las sondas se pueden colocar en el tumor durante la cirugía o a través de la piel (por vía percutánea). Después de la criocirugía, el tejido congelado se descongela y el cuerpo puede absorberlo de forma natural (en el caso de tumores internos) o puede disolverse y formar una costra (en el caso de

tumores externos).

También se pueden utilizar agentes quimioterapéuticos para el tratamiento del cáncer. Los ejemplos de agentes quimioterapéuticos incluyen agentes alquilantes, antimetabolitos, alcaloides y terpenoides vegetales, alcaloides de la vinca, podofilotoxina, taxanos, inhibidores de la topoisomerasa y antibióticos citotóxicos. Cisplatino, carboplatino y oxaliplatino son ejemplos de agentes alquilantes. Otros agentes alquilantes incluyen mecloretamina, ciclofosfamida, clorambucilo e ifosfamida. Los agentes alquilantes pueden alterar la función celular al formar enlaces covalentes con los grupos amino, carboxilo, sulfhidrilo y fosfato en moléculas biológicamente importantes. Alternativamente, los agentes alquilantes pueden modificar químicamente el ADN de una célula.

Los antimetabolitos son otro ejemplo de agentes quimioterapéuticos. Los antimetabolitos pueden disfrazarse de purinas o pirimidinas y pueden impedir que las purinas y pirimidinas se incorporen al ADN durante la fase "S" (del ciclo celular), deteniendo así el desarrollo y la división normales. Los antimetabolitos también pueden afectar la síntesis de ARN. Ejemplos de metabolitos incluyen azatioprina y mercaptopurina.

Los alcaloides pueden derivarse de plantas y bloquear la división celular también pueden usarse para el tratamiento del cáncer. Los alquiloides pueden impedir la función de los microtúbulos. Ejemplos de alcaloides son los alcaloides de la vinca y los taxanos. Los alcaloides de la vinca pueden unirse a sitios específicos de la tubulina e inhibir el ensamblaje de la tubulina en microtúbulos (fase M del ciclo celular). Los alcaloides de la vinca pueden derivarse del bígaro de Madagascar, *Catharanthus roseus* (anteriormente conocido como *Vinca rosea*). Los ejemplos de alcaloides de vinca incluyen, entre otros, vincristina, vinblastina, vinorelbina o vindesina. Los taxanos son diterpenos producidos por las plantas del género *Taxus* (tejos). Los taxanos pueden derivarse de fuentes naturales o sintetizarse artificialmente. Los taxanos incluyen paclitaxel (Taxol) y docetaxel (Taxotere). Los taxanos pueden alterar la función de los microtúbulos. Los microtúbulos son esenciales para la división celular y los taxanos pueden estabilizar la tubulina unida al PIB en los microtúbulos, inhibiendo así el proceso de división celular. Por tanto, en esencia, los taxanos pueden ser inhibidores de la mitosis. Los taxanos también pueden ser radiosensibilizantes y a menudo contienen numerosos centros quirales.

Los agentes quimioterapéuticos alternativos incluyen podofilotoxina. La podofilotoxina es un compuesto de origen vegetal que puede ayudar con la digestión y puede usarse para producir fármacos citostáticos como el etopósido y el tenipósido. Pueden impedir que la célula entre en la fase G1 (el inicio de la replicación del ADN) y la replicación del ADN (la fase S).

Las topoisomerasas son enzimas esenciales que mantienen la topología del ADN. La inhibición de las topoisomerasas tipo I o II puede interferir tanto con la transcripción como con la replicación del ADN al alterar el superenrollamiento adecuado del ADN. Algunos agentes quimioterapéuticos

pueden inhibir las topoisomerasas. Por ejemplo, algunos inhibidores de la topoisomerasa tipo I incluyen *camptotecinas*: irinotecán y topotecán. Los ejemplos de inhibidores de tipo II incluyen amsacrina, etopósido, fosfato de etopósido y tenipósido. Los inhibidores de la cinasa se pueden usar para tratar el cáncer de mama.

5 Otro ejemplo de agentes quimioterapéuticos son los antibióticos citotóxicos. Los antibióticos citotóxicos son un grupo de antibióticos que se utilizan para el tratamiento del cáncer porque pueden interferir con la replicación del ADN y/o la síntesis de proteínas. Los antibióticos citotóxicos incluyen, entre otros, actinomicina, antraciclinas, doxorubicina, daunorrubicina, valrubicina, idarrubicina, epirubicina, bleomicina, plicamicina y mitomicina.

10 En algunos casos, el tratamiento contra el cáncer puede comprender radioterapia. La radiación puede provenir de una máquina fuera del cuerpo (radioterapia de haz externo) o de material radiactivo colocado en el cuerpo cerca de las células cancerosas (radioterapia interna, más comúnmente llamada braquiterapia). La radioterapia sistémica utiliza una sustancia radiactiva, administrada por vía oral o en una vena, que viaja en la sangre a los tejidos de todo el cuerpo.

15 La radioterapia de haz externo se puede administrar en forma de haces de fotones (ya sea rayos X o rayos gamma). Un fotón es la unidad básica de la luz y otras formas de radiación electromagnética. Un ejemplo de radioterapia de haz externo se llama radioterapia conformada tridimensional (3D-CRT). 3D-CRT puede utilizar software de computadora y máquinas de tratamiento avanzadas para administrar radiación a áreas diana con formas muy precisas. Actualmente se están probando y
20 utilizando muchos otros métodos de radioterapia de haz externo en el tratamiento del cáncer. Estos métodos incluyen, entre otros, radioterapia de intensidad modulada (IMRT), radioterapia guiada por imágenes (IGRT), radiocirugía estereotáctica (SRS), radioterapia corporal estereotáctica (SBRT) y terapia de protones.

La radioterapia de intensidad modulada (IMRT) es un ejemplo de radiación de haz externo y puede
25 utilizar cientos de pequeños dispositivos formadores de haces de radiación, llamados colimadores, para administrar una dosis única de radiación. Los colimadores pueden ser estacionarios o moverse durante el tratamiento, lo que permite que la intensidad de los haces de radiación cambie durante las sesiones de tratamiento. Este tipo de modulación de dosis permite que diferentes áreas de un tumor o tejidos cercanos reciban diferentes dosis de radiación. La IMRT se planifica a la inversa (lo
30 que se denomina planificación de tratamiento inversa). En la planificación del tratamiento inverso, las dosis de radiación a diferentes áreas del tumor y el tejido circundante se planifican de antemano, y luego un programa informático de alta potencia calcula el número requerido de haces y ángulos del tratamiento de radiación. Por el contrario, durante la planificación del tratamiento tradicional (adelante), el número y los ángulos de los haces de radiación se eligen de antemano y las
35 computadoras calculan cuánta dosis se puede administrar desde cada uno de los haces planificados.

El objetivo de la IMRT es aumentar la dosis de radiación en las áreas que la necesitan y reducir la exposición a la radiación en áreas sensibles específicas del tejido normal circundante.

Otro ejemplo de radiación de haz externo es la radioterapia guiada por imágenes (IGRT). En IGRT, se pueden realizar exploraciones por imágenes repetidas (CT, MRI o PET) durante el tratamiento.

5 Estas exploraciones por imágenes pueden ser procesadas por computadoras para identificar cambios en el tamaño y la ubicación de un tumor debido al tratamiento y para permitir que la posición del sujeto o la dosis de radiación planificada se ajuste durante el tratamiento según sea necesario. Las imágenes repetidas pueden aumentar la precisión del tratamiento con radiación y pueden permitir reducciones en el volumen planificado de tejido a tratar, disminuyendo así la dosis total de radiación dirigida al tejido normal.

10 La tomoterapia es un tipo de IMRT guiada por imágenes. Una máquina de tomoterapia es un híbrido entre un escáner de imágenes por tomografía computarizada y una máquina de radioterapia de haz externo. La parte de la máquina de tomoterapia que administra radiación tanto para la obtención de imágenes como para el tratamiento puede girar completamente alrededor del sujeto de la misma manera que un escáner de tomografía computarizada normal. Las máquinas de tomoterapia pueden capturar imágenes de tomografía computarizada del tumor del sujeto inmediatamente antes de las sesiones de tratamiento, para permitir un abordaje muy preciso del tumor y preservar el tejido normal. La radiocirugía estereotáctica (SRS) puede administrar una o más dosis altas de radiación a un tumor pequeño. SRS utiliza una localización de tumores y un posicionamiento del sujeto guiados por imágenes extremadamente precisos. Por lo tanto, se puede administrar una dosis alta de radiación sin dañar excesivamente el tejido normal. La SRS se puede utilizar para tratar tumores pequeños con bordes bien definidos. Se usa más comúnmente en el tratamiento de tumores cerebrales o espinales y metástasis cerebrales de otros tipos de cáncer. Para el tratamiento de algunas metástasis cerebrales, los sujetos pueden recibir radioterapia en todo el cerebro (llamada radioterapia de todo el cerebro) además de SRS. La SRS requiere el uso de un marco para la cabeza u otro dispositivo para inmovilizar al sujeto durante el tratamiento y garantizar que la dosis alta de radiación se administre con precisión.

25 La radioterapia corporal estereotáctica (SBRT) administra radioterapia en menos sesiones, utilizando campos de radiación más pequeños y dosis más altas que la 3D-CRT en la mayoría de los casos. La SBRT puede tratar tumores que se encuentran fuera del cerebro y la médula espinal. Debido a que es más probable que estos tumores se muevan con el movimiento normal del cuerpo y, por lo tanto, no pueden atacarse con tanta precisión como los tumores del cerebro o la columna vertebral, la SBRT generalmente se administra en más de una dosis. La SBRT se puede utilizar para tratar tumores pequeños y aislados, incluidos los cánceres de pulmón y de hígado. Los sistemas SBRT pueden ser conocidos por sus marcas, como CyberKnife®.

En la terapia de protones, la radioterapia de haz externo puede administrarse mediante protones. Los protones son un tipo de partícula cargada. Los haces de protones se diferencian de los haces de fotones principalmente en la forma en que depositan energía en el tejido vivo. Mientras que los fotones depositan energía en pequeños paquetes a lo largo de su recorrido a través del tejido, los protones depositan gran parte de su energía al final de su recorrido (llamado pico de Bragg) y depositan menos energía a lo largo del camino. El uso de protones puede reducir la exposición del tejido normal a la radiación, permitiendo posiblemente la administración de dosis más altas de radiación a un tumor.

Se pueden utilizar otros haces de partículas cargadas, como los haces de electrones, para irradiar tumores superficiales, como el cáncer de piel o tumores cerca de la superficie del cuerpo, pero no pueden viajar muy lejos a través del tejido.

La radioterapia interna (braquiterapia) es radiación administrada desde fuentes de radiación (materiales radiactivos) que se colocan dentro o sobre el cuerpo. En el tratamiento del cáncer se utilizan varias técnicas de braquiterapia. La braquiterapia intersticial puede utilizar una fuente de radiación que se coloca dentro del tejido tumoral, como dentro de un tumor de mama. La braquiterapia intracavitaria puede utilizar una fuente que se coloca dentro de una cavidad quirúrgica o una cavidad corporal, como la cavidad torácica, cerca de un tumor. La braquiterapia epiescleral, que puede usarse para tratar el melanoma dentro del ojo, puede usar una fuente adherida al ojo. En la braquiterapia, los isótopos radiactivos se pueden sellar en pequeños gránulos o "semillas". Estas semillas pueden colocarse en sujetos utilizando dispositivos de administración, como agujas, catéteres o algún otro tipo de portador. A medida que los isótopos se desintegran naturalmente, emiten radiación que puede dañar las células cancerosas cercanas. La braquiterapia puede administrar dosis más altas de radiación a algunos cánceres que la radioterapia de haz externo y, al mismo tiempo, causar menos daño al tejido normal.

La braquiterapia se puede administrar como un tratamiento de dosis baja o de dosis alta. En el tratamiento de dosis baja, las células cancerosas reciben radiación continua de dosis baja de la fuente durante un período de varios días. En el tratamiento de alta tasa de dosis, una máquina robótica conectada a tubos de administración colocados dentro del cuerpo puede guiar una o más fuentes radiactivas hacia o cerca de un tumor y luego eliminar las fuentes al final de cada sesión de tratamiento. El tratamiento con dosis altas se puede administrar en una o más sesiones de tratamiento. Un ejemplo de tratamiento de alta tasa de dosis es el sistema MammoSite®. La braquiterapia se puede utilizar para tratar sujetos con cáncer de mama que se han sometido a una cirugía de conservación de la mama.

La colocación de fuentes de braquiterapia puede ser temporal o permanente. Para la braquiterapia permanente, las fuentes pueden sellarse quirúrgicamente dentro del cuerpo y dejarse allí, incluso

después de que se haya emitido toda la radiación. En algunos casos, el material restante (en donde se sellaron los isótopos radiactivos) no causa ninguna molestia ni daño al sujeto. La braquiterapia permanente es un tipo de braquiterapia de baja dosis. Para la braquiterapia temporal, se utilizan tubos (catéteres) u otros portadores para administrar las fuentes de radiación, y tanto los portadores como las fuentes de radiación se retiran después del tratamiento. La braquiterapia temporal puede ser tanto un tratamiento de dosis baja o alta. La braquiterapia se puede utilizar sola o además de la radioterapia de haz externo para proporcionar un "impulso" de radiación a un tumor sin afectar el tejido normal circundante.

En la radioterapia sistémica, un sujeto puede tragar o recibir una inyección de una sustancia radiactiva, como yodo radiactivo o una sustancia radiactiva unida a un anticuerpo monoclonal. El yodo radiactivo (^{131}I) es un tipo de radioterapia sistémica que se usa comúnmente para ayudar a tratar el cáncer, como el cáncer de tiroides. Las células tiroideas absorben naturalmente el yodo radiactivo. Para la radioterapia sistémica para algunos otros tipos de cáncer, un anticuerpo monoclonal puede ayudar a dirigir la sustancia radiactiva al lugar correcto. El anticuerpo unido a la sustancia radiactiva viaja a través de la sangre, localizando y destruyendo las células tumorales. Por ejemplo, el medicamento ibritumomab tiuxetan (Zevalin®) se puede usar para el tratamiento de ciertos tipos de linfoma no Hodgkin (LNH) de células B. La parte de anticuerpos de este medicamento reconoce y se une a una proteína que se encuentra en la superficie de los linfocitos B. El régimen de medicamentos combinados de tositumomab y yodo ^{131}I tositumomab (Bexxar®) se puede usar para el tratamiento de ciertos tipos de cáncer, como el LNH. En este régimen, primero se pueden administrar a los sujetos anticuerpos no radiactivos contra tositumomab, seguidos del tratamiento con anticuerpos contra tositumomab que tienen ^{131}I adherido. Tositumomab puede reconocer y unirse a la misma proteína de los linfocitos B que ibritumomab. La forma no radiactiva del anticuerpo puede ayudar a proteger a los linfocitos B normales del daño causado por la radiación del ^{131}I .

Algunos medicamentos de radioterapia sistémica alivian el dolor causado por el cáncer que se ha diseminado al hueso (metástasis óseas). Este es un tipo de radioterapia paliativa. Los fármacos radiactivos samario-153-lexidronam (Quadramet®) y cloruro de estroncio-89 (Metastron®) son ejemplos de radiofármacos que pueden usarse para tratar el dolor causado por metástasis óseas.

La terapia biológica (a veces llamada inmunoterapia, bioterapia, terapia biológica o terapia modificadora de la respuesta biológica (BRM)) utiliza el sistema inmunológico del cuerpo, ya sea directa o indirectamente, para combatir el cáncer o disminuir los efectos secundarios que pueden causar algunos tratamientos contra el cáncer. Las terapias biológicas incluyen interferones, interleucinas, factores estimulantes de colonias, anticuerpos monoclonales, vacunas, terapia génica y agentes inmunomoduladores inespecíficos.

Los interferones (IFN) son tipos de citoquinas que se encuentran naturalmente en el cuerpo. El

interferón alfa, el interferón beta y el interferón gamma son ejemplos de interferones que pueden usarse en el tratamiento del cáncer.

Al igual que los interferones, las interleucinas (IL) son citoquinas que se producen naturalmente en el cuerpo y se pueden producir en el laboratorio. Se han identificado muchas interleucinas para el tratamiento del cáncer. Por ejemplo, la interleucina-2 (IL-2 o aldesleucina), la interleucina 7 y la interleucina 12 pueden usarse como tratamiento contra el cáncer. La IL-2 puede estimular el crecimiento y la actividad de muchas células inmunitarias, como los linfocitos, que pueden destruir las células cancerosas. Las interleucinas se pueden utilizar para tratar varios tipos de cáncer, incluidos la leucemia, el linfoma y el cáncer de cerebro, colorrectal, de ovario, de mama, de riñón y de próstata.

Los factores estimulantes de colonias (CSF) (a veces llamados factores de crecimiento hematopoyético) también se pueden usar para el tratamiento del cáncer. Algunos ejemplos de CSF incluyen, entre otros, G-CSF (filgrastim) y GM-CSF (sargramostim). Los CSF pueden promover la división de las células madre de la médula ósea y su desarrollo en glóbulos blancos, plaquetas y glóbulos rojos. La médula ósea es fundamental para el sistema inmunológico del cuerpo porque es la fuente de todas las células sanguíneas. Debido a que los medicamentos contra el cáncer pueden dañar la capacidad del cuerpo para producir glóbulos blancos, glóbulos rojos y plaquetas, la estimulación del sistema inmunológico mediante los CSF puede beneficiar a los sujetos que se someten a otros tratamientos contra el cáncer, por lo que los CSF pueden combinarse con otras terapias contra el cáncer. como la quimioterapia. Los CSF se pueden utilizar para tratar una gran variedad de cánceres, incluidos linfoma, leucemia, mieloma múltiple, melanoma y cánceres de cerebro, pulmón, esófago, mama, útero, ovario, próstata, riñón, colon y recto.

Otro tipo de terapia biológica incluye anticuerpos monoclonales (MOAB o MoAB). Estos anticuerpos pueden ser producidos por un solo tipo de célula y pueden ser específicos de un antígeno particular. Para crear MOAB, se pueden inyectar células cancerosas humanas en ratones. En respuesta, el sistema inmunológico del ratón puede producir anticuerpos contra estas células cancerosas. Las células plasmáticas de ratón que producen anticuerpos pueden aislarse y fusionarse con células cultivadas en laboratorio para crear células "híbridas" llamadas hibridomas. Los hibridomas pueden producir indefinidamente grandes cantidades de estos anticuerpos puros o MOAB. Los MOAB se pueden utilizar en el tratamiento del cáncer de varias formas. Por ejemplo, los MOAB que reaccionan con tipos específicos de cáncer pueden mejorar la respuesta inmune de un sujeto al cáncer. Los MOAB se pueden programar para actuar contra los factores de crecimiento celular, interfiriendo así con el crecimiento de las células cancerosas.

Los MOAB pueden estar relacionados con otras terapias contra el cáncer, como quimioterapia, radioisótopos (sustancias radiactivas), otras terapias biológicas u otras toxinas. Cuando los

anticuerpos se adhieren a las células cancerosas, administran estas terapias anticancerígenas directamente al tumor, ayudando a destruirlo. Los MOAB que transportan radioisótopos también pueden resultar útiles en el diagnóstico de ciertos cánceres, como el colorrectal, el de ovario, el de próstata y el de mama.

5 Rituxan® (rituximab) y Herceptin® (trastuzumab) son ejemplos de MOAB que pueden usarse como terapia biológica. Rituxan se puede utilizar para el tratamiento del linfoma no Hodgkin. Herceptin se puede utilizar para tratar el cáncer de mama metastásico en sujetos con tumores que producen cantidades excesivas de una proteína llamada HER2. Alternativamente, los MOAB se pueden usar para tratar linfoma, leucemia, melanoma y cánceres de cerebro, mama, pulmón, riñón, colon, recto, 10 ovario, próstata y otras áreas.

Las vacunas contra el cáncer son otra forma de terapia biológica. Las vacunas contra el cáncer pueden diseñarse para estimular el sistema inmunológico del sujeto a reconocer las células cancerosas. Las vacunas contra el cáncer pueden diseñarse para tratar cánceres existentes (vacunas terapéuticas) o para prevenir el desarrollo del cáncer (vacunas profilácticas). Se pueden 15 inyectar vacunas terapéuticas en una persona después de que se le diagnostica cáncer. Estas vacunas pueden detener el crecimiento de tumores existentes, prevenir la recurrencia del cáncer o eliminar las células cancerosas que no fueron eliminadas por tratamientos anteriores. Las vacunas contra el cáncer administradas cuando el tumor es pequeño pueden erradicar el cáncer. Por otro lado, las vacunas profilácticas se administran a personas sanas antes de que se desarrolle el cáncer. 20 Estas vacunas están diseñadas para estimular el sistema inmunológico para que ataque los virus que pueden causar cáncer. Al atacar estos virus que causan cáncer, se puede prevenir el desarrollo de ciertos cánceres. Por ejemplo, cervarix y gardasil son vacunas para tratar el virus del papiloma humano y pueden prevenir el cáncer de cuello uterino. Las vacunas terapéuticas se pueden utilizar para tratar el melanoma, el linfoma, la leucemia y los cánceres de cerebro, mama, pulmón, riñón, 25 ovario, próstata, páncreas, colon y recto. Las vacunas contra el cáncer se pueden utilizar en combinación con otras terapias contra el cáncer.

La terapia génica es otro ejemplo de terapia biológica. La terapia génica puede implicar la introducción de material genético en las células de una persona para combatir enfermedades. Los métodos de terapia génica pueden mejorar la respuesta inmune de un sujeto al cáncer. Por ejemplo, 30 se puede insertar un gen en una célula inmunitaria para mejorar su capacidad de reconocer y atacar células cancerosas. En otro enfoque, se pueden inyectar a las células cancerosas genes que hacen que las células cancerosas produzcan citoquinas y estimulen el sistema inmunológico.

En algunos casos, la terapia biológica incluye agentes inmunomoduladores no específicos. Los agentes inmunomoduladores inespecíficos son sustancias que estimulan o aumentan 35 indirectamente el sistema inmunológico. A menudo, estos agentes se dirigen a células clave del

sistema inmunológico y pueden provocar respuestas secundarias, como una mayor producción de citocinas e inmunoglobulinas. Dos agentes inmunomoduladores no específicos utilizados en el tratamiento del cáncer son el bacilo de Calmette-Guérin (BCG) y el levamisol. BCG puede usarse en el tratamiento del cáncer de vejiga superficial después de una cirugía. BCG puede actuar estimulando una respuesta inflamatoria y posiblemente inmune. Se puede instilar una solución de BCG en la vejiga. El levamisol a veces se usa junto con la quimioterapia con fluorouracilo (5-FU) en el tratamiento del cáncer de colon en estadio III (Dukes C) después de una cirugía. El levamisol puede actuar para restaurar la función inmune deprimida.

La terapia fotodinámica (TFD) es un tratamiento contra el cáncer que puede utilizar un medicamento, llamado fotosensibilizador o agente fotosensibilizante, y un tipo particular de luz. Cuando los fotosensibilizadores se exponen a una longitud de onda de luz específica, pueden producir una forma de oxígeno que mata las células cercanas. Un fotosensibilizador puede activarse mediante luz de una longitud de onda específica. Esta longitud de onda determina hasta dónde puede viajar la luz dentro del cuerpo. Por tanto, se pueden utilizar fotosensibilizadores y longitudes de onda de luz para tratar diferentes áreas del cuerpo con TFD.

En el primer paso de la TFD para el tratamiento del cáncer, se puede inyectar un agente fotosensibilizante en el torrente sanguíneo. El agente puede ser absorbido por células de todo el cuerpo, pero puede permanecer en las células cancerosas por más tiempo que en las células normales. Aproximadamente de 24 a 72 horas después de la inyección, cuando la mayor parte del agente ha abandonado las células normales pero permanece en las células cancerosas, el tumor puede exponerse a la luz. El fotosensibilizador del tumor puede absorber la luz y produce una forma activa de oxígeno que destruye las células cancerosas cercanas. Además de matar directamente las células cancerosas, la TFD puede reducir o destruir los tumores de otras dos formas. El fotosensibilizador puede dañar los vasos sanguíneos del tumor, impidiendo así que el cáncer reciba los nutrientes necesarios. La TFD también puede activar el sistema inmunológico para atacar las células tumorales.

La luz utilizada para la TFD puede provenir de un láser u otras fuentes. La luz láser se puede dirigir a través de cables de fibra óptica (fibras delgadas que transmiten luz) para enviar luz a áreas internas del cuerpo. Por ejemplo, se puede insertar un cable de fibra óptica a través de un endoscopio (un tubo delgado e iluminado que se usa para observar los tejidos dentro del cuerpo) hasta los pulmones o el esófago para tratar el cáncer en estos órganos. Otras fuentes de luz incluyen diodos emisores de luz (LED), que pueden usarse para tumores superficiales, como el cáncer de piel. La TFD generalmente se realiza como un procedimiento externo. La TFD también puede repetirse y usarse con otras terapias, como cirugía, radiación o quimioterapia.

La fotoféresis extracorpórea (ECP) es un tipo de TFD en donde se puede utilizar una máquina para

recolectar las células sanguíneas del sujeto. Las células sanguíneas del sujeto pueden tratarse fuera del cuerpo con un agente fotosensibilizante, exponerse a la luz y luego devolverse al sujeto. La ECP se puede utilizar para ayudar a disminuir la gravedad de los síntomas cutáneos del linfoma cutáneo de células T que no ha respondido a otras terapias. La ECP se puede utilizar para tratar otros cánceres de la sangre y también puede ayudar a reducir el rechazo después de los trasplantes.

Además, se pueden usar agentes fotosensibilizantes, como porfímero sódico o Photofrin®, en la TFD para tratar o aliviar los síntomas del cáncer de esófago y del cáncer de pulmón de células no pequeñas. El porfímero sódico puede aliviar los síntomas del cáncer de esófago cuando el cáncer obstruye el esófago o cuando el cáncer no se puede tratar satisfactoriamente solo con terapia con láser. El porfímero sódico se puede utilizar para tratar el cáncer de pulmón de células no pequeñas en sujetos para quienes los tratamientos habituales no son apropiados y para aliviar los síntomas en sujetos con cáncer de pulmón de células no pequeñas que obstruye las vías respiratorias. El porfímero sódico también se puede utilizar para el tratamiento de lesiones precancerosas en sujetos con esófago de Barrett, una afección que puede provocar cáncer de esófago.

La terapia con láser puede utilizar luz de alta intensidad para tratar el cáncer y otras enfermedades. Los láseres se pueden utilizar para reducir o destruir tumores o crecimientos precancerosos. Los láseres se usan más comúnmente para tratar cánceres superficiales (cánceres en la superficie del cuerpo o en el revestimiento de los órganos internos), como el cáncer de piel de células basales y las etapas muy tempranas de algunos cánceres, como el de cuello uterino, pene, vagina, vulva y cáncer de pulmón de células no pequeñas.

Los láseres también se pueden usar para aliviar ciertos síntomas del cáncer, como sangrado u obstrucción. Por ejemplo, se pueden utilizar láseres para reducir o destruir un tumor que bloquea la tráquea o el esófago de un sujeto. Los láseres también se pueden utilizar para extirpar pólipos o tumores del colon que bloquean el colon o el estómago.

La terapia con láser a menudo se administra a través de un endoscopio flexible (un tubo delgado e iluminado que se usa para observar los tejidos dentro del cuerpo). El endoscopio está equipado con fibras ópticas (fibras delgadas que transmiten luz). Se inserta a través de una abertura en el cuerpo, como la boca, la nariz, el ano o la vagina. Luego, la luz láser se dirige con precisión para cortar o destruir un tumor.

La termoterapia intersticial inducida por láser (LITT), o fotocoagulación con láser intersticial, también utiliza láseres para tratar algunos cánceres. LITT es similar a un tratamiento contra el cáncer llamado hipertermia, que utiliza calor para reducir los tumores dañando o matando las células cancerosas. Durante LITT, se inserta una fibra óptica en un tumor. La luz láser en la punta de la fibra eleva la temperatura de las células tumorales y las daña o destruye. A veces, LITT se utiliza para reducir el tamaño de los tumores en el hígado.

La terapia con láser se puede utilizar sola, pero la mayoría de las veces se combina con otros tratamientos, como cirugía, quimioterapia o radioterapia. Además, los láseres pueden sellar las terminaciones nerviosas para reducir el dolor después de la cirugía y sellar los vasos linfáticos para reducir la hinchazón y limitar la propagación de las células tumorales.

5 Los láseres utilizados para tratar el cáncer pueden incluir láseres de dióxido de carbono (CO₂), láseres de argón y láseres de neodimio:itrio-aluminio-granate (Nd:YAG). Cada uno de estos puede reducir o destruir tumores y puede usarse con endoscopios. Los láseres de CO₂ y argón pueden cortar la superficie de la piel sin penetrar en capas más profundas. Por tanto, pueden utilizarse para eliminar cánceres superficiales, como el cáncer de piel. Por el contrario, el láser Nd:YAG se aplica
10 más comúnmente a través de un endoscopio para tratar órganos internos, como el útero, el esófago y el colon. La luz láser Nd:YAG también puede viajar a través de fibras ópticas a áreas específicas del cuerpo durante LITT. Los láseres de argón se utilizan a menudo para activar los fármacos utilizados en la TFD.

Para los sujetos con enfermedad sistémica, se pueden designar modalidades de tratamiento
15 adicionales como quimioterapia adyuvante (p. ej., docetaxel, mitoxantrona y prednisona) y radioterapia sistémica (p. ej., samario o estroncio). Es probable que estos sujetos sean tratados inmediatamente con radioterapia para eliminar la presunta enfermedad micrometastásica, que no puede detectarse clínicamente pero puede revelarse mediante la expresión distintiva del gen diana. Estos sujetos también pueden ser controlados más de cerca para detectar signos de progresión de
20 la enfermedad.

Para los sujetos que no tienen enfermedad sistémica, se puede administrar radioterapia adyuvante localizada (p. ej., localizada en el tejido mamario) o terapia endocrina o quimioterapia. Para los sujetos sin evidencia de enfermedad (NED), se puede usar el perfil de expresión y/o el cálculo de una puntuación de riesgo, como se describe en el presente documento, para determinar el riesgo de
25 recurrencia del cáncer de mama. Para los sujetos identificados con un riesgo bajo de recurrencia de cáncer de mama y con un alto beneficio de la radioterapia utilizando los métodos descritos en este documento, se recomendaría la radioterapia adyuvante.

Los genes diana se pueden agrupar de modo que la información obtenida sobre el conjunto de genes diana en el grupo pueda usarse para realizar o ayudar a realizar un juicio clínicamente
30 relevante tal como un diagnóstico, pronóstico o elección de tratamiento.

También se proporciona un informe del sujeto que comprende una representación de los niveles de expresión medidos de una pluralidad de genes diana en una muestra biológica del sujeto, en donde la representación comprende niveles de expresión de genes diana correspondientes a uno
35 cualquiera, dos, tres, cuatro, cinco, seis, ocho, diez, veinte, treinta o más de los genes diana, los subconjuntos descritos en el presente documento o una combinación de los mismos. En algunas

modalidades, la representación de los niveles de expresión medidos puede tomar la forma de una combinación lineal o no lineal de niveles de expresión de los genes diana de interés. El informe del tema puede proporcionarse en un formato legible por máquina (por ejemplo, una computadora) y/o en una copia impresa (papel). El informe también puede incluir mediciones estándar de los niveles de expresión de dicha pluralidad de genes diana de uno o más conjuntos de sujetos con estado y/o resultado de la enfermedad conocido. El informe se puede utilizar para informar al sujeto y/o al médico tratante de los niveles de expresión de los genes diana expresados, el diagnóstico médico probable y/o las implicaciones, y opcionalmente puede recomendar una modalidad de tratamiento para el sujeto.

También se proporcionan representaciones de los perfiles de expresión génica útiles para tratar, diagnosticar, pronosticar y evaluar de otro modo enfermedades. En algunas modalidades, estas representaciones de perfil se reducen a un medio que puede ser leído automáticamente por una máquina tal como medios legibles por computadora (magnéticos, ópticos y similares). Los artículos también pueden incluir instrucciones para evaluar los perfiles de expresión génica en dichos medios. Por ejemplo, los artículos pueden comprender una forma de almacenamiento legible que tiene instrucciones informáticas para comparar perfiles de expresión génica de las colecciones de genes descritas anteriormente. Los artículos también pueden tener perfiles de expresión génica registrados digitalmente en ellos para que puedan compararse con datos de expresión génica de muestras de sujetos. Alternativamente, los perfiles se pueden grabar en diferentes formatos de representación. Una grabación gráfica es uno de esos formatos. Los algoritmos de agrupación pueden ayudar en la visualización de dichos datos.

EXPERIMENTAL

El siguiente ejemplo se proporciona para demostrar e ilustrar mejor ciertas modalidades preferidas y aspectos de la presente invención y no deben interpretarse como taxativos del alcance de la misma. El uso de pronombres como "nosotros", "nuestro" y "yo" se refieren a la entidad inventiva.

Ejemplo 1: Desarrollo y validación de un clasificador genómico para el pronóstico de recurrencia local de cáncer de mama y predicción de respuesta a radioterapia.

Métodos

Cohorte SweBCG91-RT

SweBCG91-RT fue un ensayo aleatorio realizado en Suecia y se ha descrito previamente (véase, Malmström, P. et al. European Journal of Cancer (Oxford, Inglaterra: 1990) 39, 1690-1697, doi:10.1016/s0959-8049(03)00324-1 (2003); Killander, F. et al. European journal of cancer (Oxford, Inglaterra: 1990) 67, 57-65, (2016); Sjöström, M. et al. Journal of Clinical Oncology 37, 3340-3349

(2019)). Brevemente, el ensayo aleatorizó a 1178 pacientes con cáncer de mama en estadio I-IIA con ganglios negativos para recibir RT adyuvante de toda la mama o ninguna RT después de la BCS. Todos los pacientes tuvieron márgenes quirúrgicos negativos. La terapia endocrina y la quimioterapia se proporcionaron de acuerdo con las directrices locales de la época y solo se administraron al 8 % de los pacientes en el ensayo original. La subtipificación mediante tinción inmunohistoquímica de ER, PgR, HER2 y Ki67 se realizó como se describió anteriormente (consulte, Sjöström, M. et al. Journal of Clinical Oncology 35, 3222-3229 (2017)). 597 muestras de pacientes con tumores ER+, HER2- tenían datos de expresión génica disponibles e información LRR completa. Este subconjunto de pacientes se dividió a su vez en una cohorte de entrenamiento (N=243) y una cohorte de validación (N=354) (información adicional a continuación y Figura 1).

Cohorte Princess Margaret

La cohorte Princess Margaret fue un ensayo aleatorio realizado en Canadá y se ha descrito previamente (consulte, Fyles, AW et al. The New England Journal of Medicine 351, 963-970 (2004); Liu, FF et al. Journal of clinical oncology: Official journal of the American Society of Clinical Oncology 33, 2035-2040 (2015)). Brevemente, el ensayo asignó al azar a 769 mujeres de 50 años o más con cáncer de mama con ganglios T1 o T2 negativos a RT adyuvante de toda la mama o no a RT después de BCS. Todos los pacientes del ensayo Princess Margaret fueron tratados con tamoxifeno. La subtipificación mediante tinción inmunohistoquímica de ER, PgR, HER2 y Ki67 se realizó como se describió anteriormente (consulte, Liu, FF et al. Journal of clinical oncology: Official journal of the American Society of Clinical Oncology 33, 2035-2040 (2015)). Los datos de expresión génica estaban disponibles a partir de 132 muestras de pacientes con tumores ER+, HER2- e información LRR completa (Figura 2).

Datos de expresión génica.

Los datos de expresión génica (Gene Expression Omnibus, GSE119295) se obtuvieron de tumores primarios de 764 pacientes incluidos en el ensayo SweBCG91-RT (consulte, Sjöström, M. et al. Journal of Clinical Oncology 37, 3340-3349 (2019)). De manera similar a la cohorte SweBCG91-RT, los datos de expresión génica para la cohorte Princess Margaret se generaron a partir de GeneChip Human Exon Arrays 1.0 ST (Thermo Fisher Scientific, South San Francisco) en un laboratorio certificado por CLIA/CAP (Decipher Biosciences, San Diego, CA).

Desarrollo de conjuntos de capacitación y validación

En este análisis, solo se incluyeron pacientes con tumores ER+, HER2-. Esto fue para garantizar que la característica resultante pudiera distinguir el riesgo entre pacientes clínicamente de bajo

riesgo que serían candidatos para la omisión de RT. Además, como la terapia sistémica puede influir en la LRR y solo una pequeña porción de los pacientes con SweBCG91-RT fueron tratados sistémicamente, los pacientes tratados con quimioterapia y/o terapia endocrina se eliminaron del análisis de la cohorte de SweBCG91-RT para mejorar la interpretación de los resultados. Después de estas consideraciones, se incluyeron en este estudio 597 pacientes del ensayo SweBCG91-RT. Para mejorar la relación señal-ruido para la identificación de un evento LRR en la cohorte de entrenamiento compuesta por datos SweBCG91-RT, solo pacientes con un evento locorregional registrado como primer evento antes de 10 años o pacientes sin ningún evento registrado con un tiempo de seguimiento Se incluyeron mayores de 10 años. El 60 % de estos pacientes fueron retenidos para la cohorte de entrenamiento y todos los pacientes restantes se utilizaron para la validación (Figura 1). No hubo diferencias significativas en la distribución de las variables clínico-patológicas entre los grupos Sin RT y RT de las cohortes de entrenamiento y validación de SweBCG91-RT (Tablas 1 y 2), con la excepción de la edad en la cohorte de entrenamiento, en donde hubo una mayor proporción de mujeres de 40-59 años en el grupo de RT en comparación con el grupo sin RT. La tasa más baja de recurrencia de tumores de mama ipsilateral en el grupo de RT en comparación con Sin RT se mantuvo tanto en la cohorte de entrenamiento como en la de validación (Tabla 3). La totalidad de la cohorte Princess Margaret con datos genómicos y LRR disponibles de pacientes con tumores ER+ y HER2- se reservó para validación y ninguno se utilizó con fines de capacitación.

Análisis de enriquecimiento de conjuntos de genes

Para identificar conjuntos de genes relacionados con LRR, se realizó un Análisis de Enriquecimiento de Conjuntos de Genes (GSEA) dentro de los 131 pacientes de la cohorte de entrenamiento no tratados con RT (consulte, Subramanian, A. et al. Proc Natl Acad Sci USA 102, 15545-15550, (2005); Mootha, VK et al. Nat Genet 34, 267-273 (2003)). Primero, para cada transcripción genética anotada, se calculó un modelo de Cox para LRR. A continuación, se introdujo una lista de genes preclasificados en la GSEA, en donde para cada gen se creó una estadística con la siguiente ecuación: $-\log(\text{valor } p) * (\text{si } HR > 1, 1, \text{ sino } -1)$, en donde el valor p y el cociente de riesgo (HR) son del modelo de Cox al LRR (véase la Tabla 8 para los valores de estadística / coeficiente determinados para cada gen). Por lo tanto, los genes individuales altamente pronósticos para una tasa LRR más alta en pacientes tendrán valores positivos grandes, y los genes individuales altamente pronósticos para una tasa LRR más baja en pacientes tendrán valores negativos grandes. La lista clasificada se ingresó en GSEA (versión 4.0.3) y se especificaron las colecciones Hallmarks (H), C2 y C5 (versión 7.2 de la base de datos Molecular Signatures, MSigDB) (consulte, Liberzon, A. et al. Cell Systems 1, 417-425 (2015)). Debido a las diferencias inherentes entre las colecciones

de conjuntos de genes, las listas H, C2 y C5 se analizaron por separado para buscar tendencias entre las colecciones. Para los genes analizados, se acumuló una puntuación de enriquecimiento total en función de los valores de coeficiente determinados para cada gen (véase la Tabla 8). Los conjuntos de genes con puntuaciones de enriquecimiento positivo se interpretaron como asociados con un peor pronóstico (aumento de LRR), y los conjuntos de genes con puntuaciones de enriquecimiento negativo se interpretaron como asociados con un mejor pronóstico (disminución de LRR).

Desarrollo del modelo

Los genes se identificaron como candidatos potenciales para el modelo de las listas de genes GSEA con una tasa de descubrimiento falso valor $q < 0,05$ y entre los diez primeros de las listas de genes positivos y negativos de las colecciones H, C2 y C5 (Tabla 4). Estos genes nominados, junto con los genes que codifican ER, PgR, HER2 y Ki67, se cruzaron con genes que cumplían con los siguientes requisitos: genes con una desviación estándar en el cuartil superior de los valores de desviación estándar y genes con un valor de $p < 0,05$ cuando se evalúa en un modelo de Cox hasta el criterio de valoración LRR en pacientes no tratados con RT en el conjunto de entrenamiento (Figura 3). Los 82 genes resultantes se introdujeron en un modelo de regresión neta elástica (R glmnet, $\alpha = 0,5$) para el entrenamiento en la cohorte de entrenamiento, entre pacientes no tratados con RT, utilizando LRR como criterio de valoración. El modelo de regresión neta elástica dio como resultado un conjunto final de 16 genes para el modelo. Las puntuaciones se dicotomizaron entre alta y baja aplicando un límite preespecificado utilizando el valor del percentil 25º de las puntuaciones en el conjunto de datos de entrenamiento SweBCG91-RT. Luego, este umbral se aplicó a los conjuntos de validación SweBCG91-RT y Princess Margaret.

Métodos estadísticos

Los análisis estadísticos se realizaron utilizando R, versión 4.0.2 (<https://cran.r-project.org/bin/windows/base/old/4.0.2/>). El criterio de valoración principal fue la incidencia acumulada de LRR utilizando el tiempo transcurrido hasta la LRR como primer evento. Las incidencias acumuladas se calcularon utilizando un enfoque de riesgo competitivo (paquete R cmprsk) y los HR y las pruebas para diferencias significativas entre los grupos de riesgo dicotomizados se calcularon utilizando la regresión de riesgos proporcionales de Cox por causa específica. Como se describió anteriormente, los riesgos no son proporcionales a lo largo del tiempo para varias variables en la cohorte SweBCG91RT cuando se verifican gráficamente y con la prueba de Schoenfeld, y los HR deben interpretarse como la media durante todo el período de tiempo (consulte, Liberzon, A. et al. Cell Systems 1, 417-425 (2015)). Las metástasis a distancia y la muerte

sin recurrencia se consideraron eventos competitivos. Se consideró que los pacientes con metástasis a distancia sincrónica y LRR, definida como LRR registrada al mismo tiempo o dentro de los 3 meses de la metástasis, tenían una LRR. Los índices de riesgo y las estimaciones puntuales se informan con intervalos de confianza (IC) del 95 % entre paréntesis. Un valor de p inferior a 0,05 se consideró estadísticamente significativo. Para las tablas demográficas, las diferencias en las distribuciones de grupo entre los grupos tratados y no tratados se probaron mediante la prueba de Chi cuadrado con corrección de continuidad de Yates para tablas de 2x2. Los análisis de Cox univariantes y multivariantes y la prueba de interacción se realizaron utilizando puntuaciones de características continuas, por lo que no dependen de ningún umbral. La prueba de interacción entre la puntuación y el RT incluyó a todos los pacientes de las cohortes de validación SweBCG91-RT y Princess Margaret y se estratificó por cohorte.

Resultados

Conceptos biológicos asociados a la recurrencia locorregional

En pacientes no tratados con RT en la cohorte de entrenamiento SweBCG91-RT, utilizamos GSEA para identificar conceptos biológicos asociados con LRR en el cáncer de mama. Las principales listas de genes enriquecidos positivamente en las colecciones Hallmark, C2 y C5 de la base de datos de Molecular Signatures incluían listas que involucran el ciclo celular y la proliferación (Tabla 4), mientras que las principales listas de genes enriquecidos negativamente incluían listas relacionadas con el sistema inmunológico.

Las listas de genes principales se cruzaron con los genes más informativos y de pronóstico individual para LRR en los pacientes del conjunto de entrenamiento que no recibieron RT. Se entrenó un modelo lineal generalizado con regularización neta elástica utilizando el criterio de valoración LRR, y la característica de 16 genes resultante incluye genes que incorporan los conceptos que se muestran ampliamente en las listas GSEA (Tabla 5). A esta característica la denominamos Perfil para la omisión de radiación adyuvante local (POLAR). La Tabla 8 muestra los valores de estadística/coeficiente determinados para cada uno de los genes utilizados para generar una puntuación de enriquecimiento para evaluar el pronóstico de LRR en pacientes.

Se utilizó un algoritmo lineal para generar una puntuación de riesgo y clasificar a los pacientes como de bajo o alto riesgo utilizando POLAR. Específicamente, los niveles de expresión para cada uno de los 16 genes en la característica (Tabla 5) se determinaron y multiplicaron por sus respectivos valores de coeficiente (Tabla 8) para producir un valor de expresión ponderado para cada gen analizado. Luego, los valores de expresión ponderados se sumaron para generar una puntuación de riesgo final. Luego se aplicó un umbral predeterminado de 0,5622 a las puntuaciones de riesgo generadas para asignar a los pacientes a categorías de alto o bajo riesgo. Por ejemplo, los pacientes

con una puntuación de riesgo superior a 0,5622 se clasificaron como de alto riesgo y los pacientes con una puntuación de riesgo inferior a 0,5622 se clasificaron como de bajo riesgo.

POLAR es pronóstico para LRR en pacientes no tratados con RT en la cohorte de validación SweBCG91-RT

En la cohorte de validación SweBCG91-RT, POLAR fue pronóstico para LRR en pacientes no tratados con radioterapia en un modelo de Cox univariable con LRR como criterio de valoración (HR=1,7[1,3,2,2], $p<0,001$). La característica siguió siendo pronóstica en un modelo multivariable que incluye edad, grado, tamaño del tumor y luminal A frente a luminal B (HR = 1,7 [1,2, 2,3], $p<0,001$) (Tabla 6).

Efecto diferencial de la RT en pacientes con riesgo POLAR bajo y alto en las cohortes de validación SweBCG91-RT y Princess Margaret

La incidencia acumulada de LRR para pacientes con SweBCG91-RT clasificados como de bajo riesgo en la cohorte de validación por POLAR fue inferior al 10 % a los 10 años, independientemente de si recibieron radioterapia o no y no se pudo demostrar un beneficio de la RT (años incidencia acumulada LRR a 10 años Sin RT: 6 % [2 %-16 %], RT: 5 % [1 %-13 %], HR=1,1 [0,39-3,4], $p=0,81$, Figura 4A). Sin embargo, estos resultados deben interpretarse con cautela dado el pequeño número de pacientes (N=108). Para los pacientes con SweBCG91-RT clasificados como de alto riesgo por la característica, la RT resultó en un riesgo significativamente menor de LRR en comparación con aquellos que no recibieron RT (incidencia acumulada locorregional a 10 años para No RT: 19 % [13 %-27 %], RT: 8 % [4 %-14 %], HR=0,43 [0,24-0,78], $p=0,0055$, Figura 4B).

Luego evaluamos POLAR en 132 pacientes del ensayo Princess Margaret (Figura 5). Los detalles demográficos y descriptivos de esta cohorte se proporcionan en la Tabla 7. Para las mujeres tratadas con tamoxifeno solo después de BCS, las pacientes clasificadas como de bajo riesgo por POLAR tuvieron una incidencia acumulada de LRR a 10 años del 7 % [0 %-27 %] sin RT y el 13 % [2 %-34 %] con RT. No se pudo demostrar un beneficio de la RT en estas mujeres de bajo riesgo (HR=1,5 [0,14-16], $p=0,74$), aunque el pequeño número de pacientes (N=34) impide una conclusión definitiva (Figura 5A). Las mujeres clasificadas por POLAR como de alto riesgo tuvieron una incidencia acumulada de LRR a 10 años del 22 % [10 %-36 %] sin RT y del 8 % [2 %-20 %] con RT, y un beneficio significativo de la RT (HR =0,25 [0,07-0,92], $p=0,038$). Realizamos un análisis exploratorio probando la interacción entre RT y la puntuación POLAR continua en el total de N = 486 pacientes de las muestras evaluables SweBCG91-RT y Princess Margaret y el valor p de interacción fue 0,066. Estos resultados mostraron que los métodos y clasificadores de la presente divulgación son pronósticos para la recurrencia locorregional e identifican a pacientes con cáncer de mama que no

se beneficiarían de la radioterapia adyuvante (tratamiento con radiación). Los resultados también mostraron que los métodos y clasificadores de la divulgación pueden identificar pacientes con cáncer de mama que tienen un riesgo bajo de recurrencia locorregional. Estos resultados demostraron además que los métodos y clasificadores del presente son útiles para tratar el cáncer de mama.

El clasificador genómico de la presente divulgación identificó pacientes con bajo riesgo de LRR sin radioterapia adyuvante después de la cirugía de conservación de la mama, identificándolos así como candidatos para la omisión de radioterapia después de la cirugía de conservación de la mama. Además, el clasificador genómico de la presente divulgación mostró que los pacientes con cáncer de mama con menor riesgo de LRR no tienen ningún beneficio significativo de la RT.

Tabla 1. Datos demográficos de la cohorte de entrenamiento de SweBCG91-RT

N (%)	Sin RT N=131	RT N=112	Total N=243	valor p
Edad en el momento de la cirugía (años)				
≤39	5 (3,8)	1 (0,9)	6 (2,5)	0,03
40-49	26 (19,8)	26 (23,2)	52 (21,4)	
50-59	34 (26)	47 (42)	81 (33,3)	
60-69	57 (43,5)	34 (30,4)	91 (37,4)	
≥70	9 (6,9)	4 (3,6)	13 (5,3)	
Estado menopáusico				
premenopáusica	33 (25,2)	24 (21,4)	57 (23,5)	0,63
posmenopáusica	95 (72,5)	84 (75)	179 (73,7)	
N/C	3 (2,3)	4 (3,6)	7 (2,9)	
grado histológico				
Grado 1	14 (10,7)	24 (21,4)	38 (15,6)	0,07
Grado 2	99 (75,6)	73 (65,2)	172 (70,8)	
Grado 3	18 (13,7)	14 (12,5)	32 (13,2)	
N/C	0 (0)	1 (0,9)	1 (0,4)	
Tamaño del tumor (mm)				
≤10	55 (42)	43 (38,4)	98 (40,3)	0,75

11 – 20	71 (54,2)	64 (57,1)	135 (55,6)	
21-30	4 (3,1)	3 (2,7)	7 (2,9)	
N/C	1 (0,8)	2 (1,8)	3 (1,2)	
Subtipo por IHQ				
Luminal A	93 (71)	73 (65,2)	166 (68,3)	0,41
Luminal B (HER2-)	38 (29)	39 (34,8)	77 (31,7)	
Terapia endocrina adyuvante				
No	131 (100)	112 (100)	243 (100)	
Quimioterapia adyuvante				
No	131 (100)	112 (100)	243 (100)	

Tabla 2. Datos demográficos de la cohorte de validación SweBCG91-RT

N (%)	Sin RT N=178	RT N=176	Total N=354	valor p
Edad en el momento de la cirugía (años)				
≤39	4 (2,2)	3 (1,7)	7 (2)	0,69
40-49	34 (19,1)	26 (14,8)	60 (16,9)	
50-59	54 (30,3)	49 (27,8)	103 (29,1)	
60-69	56 (31,5)	64 (36,4)	120 (33,9)	
≥70	30 (16,9)	34 (19,3)	64 (18,1)	
Estado menopáusico				
premenopáusica	36 (20,2)	29 (16,5)	65 (18,4)	0,37
posmenopáusica	135 (75,8)	145 (82,4)	280 (79,1)	
N/C	7 (3,9)	2 (1,1)	9 (2,5)	
grado histológico				
Grado 1	32 (18)	28 (15,9)	60 (16,9)	0,36
Grado 2	108 (60,7)	118 (67)	226 (63,8)	
Grado 3	34 (19,1)	25 (14,2)	59 (16,7)	
N/C	4 (2,2)	5 (2,8)	9 (2,5)	

Tamaño del tumor (mm)				
≤10	70 (39,3)	69 (39,2)	139 (39,3)	0,78
11-20	97 (54,5)	99 (56,2)	196 (55,4)	
21-30	11 (6,2)	8 (4,5)	19 (5,4)	
Subtipo por IHQ				
Luminal A	115 (64,6)	113 (64,2)	228 (64,4)	1,00
Luminal B (HER2-)	63 (35,4)	63 (35,8)	126 (35,6)	
Terapia endocrina adyuvante				
No	178 (100)	176 (100)	354 (100)	
Quimioterapia adyuvante				
No	178 (100)	176 (100)	354 (100)	

Tabla 3. Distribución de eventos en cohortes de entrenamiento y validación de SweBCG91-RT

Cohorte de entrenamiento (N=243)				
N (%)	Sin RT (N=131)	RT (N=112)	Total	valor p
Recurrencia del tumor de mama ipsilateral				
No	91 (69,5)	97 (86,6)	188 (77,4)	0,0025
Sí	40 (30,5)	15 (13,4)	55 (22,6)	
Recurrencia regional				
No	125 (95,4)	109 (97,3)	234 (96,3)	0,66
Sí	6 (4,6)	3 (2,7)	9 (3,7)	
Metástasis distante				
No	120 (91,6)	107 (95,5)	227 (93,4)	0,33
Sí	11 (8,4)	5 (4,5)	16 (6,6)	

5	Muerte específica por cáncer de mama				
	No	116 (88,5)	102 (91,1)	218 (89,7)	0,66
	Sí	15 (11,5)	10 (8,9)	25 (10,3)	
	Muerte por cualquier causa				
	No	89 (67,9)	84 (75)	173 (71,2)	0,28
	Sí	42 (32,1)	28 (25)	70 (28,8)	
10	Cohorte de validación (N=354)				
		Sin RT (N=178)	RT (N=176)	Total (N=354)	valor p
15	Recurrencia del tumor de mama ipsilateral				
	No	135 (75,8)	153 (86,9)	288 (81,4)	0,01
	Sí	43 (24,2)	23 (13,1)	66 (18,6)	
20	Recurrencia regional				
	No	170 (95,5)	172 (97,7)	342 (96,6)	0,39
	Sí	8 (4,5)	4 (2,3)	12 (3,4)	
	Metástasis distante				
	No	151 (84,8)	146 (83)	297 (83,9)	0,74
25	Sí	27 (15,2)	30 (17)	57 (16,1)	
	Muerte específica por cáncer de mama				
	No	141 (79,2)	136 (77,3)	277 (78,2)	0,75
	Sí	37 (20,8)	40 (22,7)	77 (21,8)	
30	Muerte por cualquier causa				
	No	79 (44,4)	77 (43,8)	156 (44,1)	0,99
	Sí	99 (55,6)	99 (56,2)	198 (55,9)	

Tabla 4. Listas de genes identificados por GSEA como altamente enriquecidos positiva y negativamente.

	Positivo (asociado con peor pronóstico)	Negativo (asociado con mejor pronóstico)
5	H	HALLMARK_TNFA_SIGNALING_VI A_NFKB HALLMARK_KRAS_SIGNALING_UP HALLMARK_ALLOGRAFT_REJECTI ON
10		
15		
20		
25		
30	C2	SMID_BREAST_CANCER_NORMAL _LIKE_UP WINTER_HYPOXIA_DN PHONG_TNF_TARGETS_UP ZHENG_FOXP3_TARGETS_IN_T_L YMPHOCYTE_DN KEGG_GRAFT_VERSUS_HOST_DI SEASE BOQUEST_STEM_CELL_CULTURE D_VS_FRESH_DN BERTUCCI_INVASIVE_CARCINOM A_DUCTAL_VS_LOBULAR_DN NAKAYAMA_SOFT_TISSUE_TUMO RS_PCA2_DN REACTOME_CHEMOKINE_RECEP TORS_BIND_CHEMOKINES
35		

5	CHIANG_LIVER_CANCER_SUBCLASS_PROLIFERATION_UP	KEGG_AUTOIMMUNE_THYROID_DISEASE
10	C5	GO_ANAPHASE_PROMOTING_COMPLEX_DEPENDENT_CATABOLIC_PROCESS
15		GO_DNA_REPLICATION_INDEPENDENT_NUCLEOSOME_ORGANIZATION
20		GO_CENTROMERE_COMPLEX_ASSEMBLY
25		GO_CHROMATIN_REMODELING_AT_CENTROMERE
30		GO_B_CELL_RECEPTOR_SIGNALING_PATHWAY
		GO_POSITIVE_REGULATION_OF_INTERLEUKIN_2_PRODUCTION
		GO_INTERLEUKIN_1_BETA_PRODUCTION
		GO_REGULATION_OF_ANTIGEN_RECEPTOR_MEDIATED_SIGNALING_PATHWAY
		GO_NEGATIVE_REGULATION_OF_SMOOTH_MUSCLE_CELL_PROLIFERATION
		GO_INTERLEUKIN_1_PRODUCTION
		GO_REGULATION_OF_B_CELL_RECEPTOR_SIGNALING_PATHWAY
		GO_RESPONSE_TO_CHEMOKINE
		GO_LYMPHOCYTE_CHEMOTAXIS
		GO_DENDRITIC_CELL_CHEMOTAXIS

Tabla 5. 16 genes en el modelo.

AGR2	Gradiente anterior 2, proteína disulfuro isomerasa
B4GALT1	Beta-1,4-Galactosiltransferasa 1
CLDN7	Claudin 7

EZR	Ezrin
GNG11	Subunidad de proteína G gamma 11
JUN	Protooncogén Jun
MMP11	Matriz metalopeptidasa 11
PKIB	Inhibidor beta de la proteína quinasa dependiente de CAMP
PRPS1	Fosforribosilpirofosfato sintetasa 1
PSMD10	Subunidad del proteasoma 26S, sin ATPasa 10
SH3BP5	Proteína de unión al dominio SH3 5
SLC16A3	Familia de portadores de solutos 16 Miembro 3
SLC7A11	Familia de portadores de solutos 7 Miembro 11
SPP1	Fosfoproteína 1 secretada
TNNT1	Troponina T1, tipo esquelético lento
UBE2E1	Ubiquitina conjugando la enzima E2 E1

Tabla 6.

Análisis univariable		HR [IC95 %]	valor p
POLAR		1,7 [1,3, 2,2]	<0,001
Análisis multivariable		HR [IC95 %]	valor p
POLAR		1,7 [1,2, 2,3]	<0,001
Edad		0,98 [0,95, 1,0]	0,1
Grado histológico	1	Referencia	
	2	1,1 [0,43, 2,9]	0,82
	3	0,74 [0,23, 2,4]	0,61
Tamaño del tumor	≤20mm	Referencia	
	>20 mm, ≤50 mm	0,41 [0,055, 3,1]	0,39
Subtipo	Luminal A	Referencia	

	Luminal A	1,5 [0,75, 2,8]	0,27
--	-----------	-----------------	------

Tabla 7.

N (%)	Sin RT N=62	RT N=70	Total N=132	valor p
Edad en el momento de la cirugía (años)				
50-59	16 (25,8)	14 (20)	30 (22,7)	0,76
60-69	23 (37,1)	24 (34,3)	47 (35,6)	
70-79	19 (30,6)	27 (38,6)	46 (34,8)	
≥80	4 (6,5)	5 (7,1)	9 (6,8)	
Grado histológico				
I	9 (14,5)	9 (12,9)	18 (13,6)	0,43
II	34 (54,8)	45 (64,3)	79 (59,8)	
III	17 (27,4)	13 (18,6)	30 (22,7)	
N/C	2 (3,2)	3 (4,3)	5 (3,8)	
Tamaño del tumor (mm)				
≤10	15 (24,2)	9 (12,9)	24 (18,2)	0,41
11-20	36 (58,1)	46 (65,7)	82 (62,1)	
21-30	10 (16,1)	14 (20)	24 (18,2)	
≥31	1 (1,6)	1 (1,4)	2 (1,5)	
Estado de ER				
Positivo	62 (100)	70 (100)	132 (100)	
Estado de PR				
Negativo	7 (11,3)	10 (14,3)	17 (12,9)	0,77
Positivo	49 (79,0)	52 (74,3)	101 (76,5)	
N/C	6 (9,7)	8 (11,4)	14 (10,6)	
Estado de HER2				
Negativo	62 (100)	70 (100)	132 (100)	

Tabla 8.

Abreviatura genética	Nombre del gen	Valor del coeficiente	Número de identificación del gen Entrez
AGR2	Gradiente anterior 2, proteína disulfuro isomerasa	0,200353	10551
B4GALT1	Beta-1,4-Galactosiltransferasa 1	-0,33626	2683
CLDN7	Claudin 7	0,115625	1366
EZR	Ezrin	0,003256	7430
GNG11	Subunidad de proteína G gamma 11	-0,41974	2791
JUN	Protooncogén Jun	-0,01851	3725
MMP11	Matriz metalopeptidasa 11	0,444806	4320
PKIB	Inhibidor beta de la proteína quinasa dependiente de CAMP	0,170331	5570
PRPS1	Fosforribosilpirofosfato sintetasa 1	0,170411	5631
PSMD10	Subunidad del proteasoma 26S, sin ATPasa 10	0,042903	5716
SH3BP5	Proteína de unión al dominio SH3 5	-0,17135	9467
SLC16A3	Familia de portadores de solutos 16 Miembro 3	0,214877	9123
SLC7A11	Familia de portadores de solutos 7 Miembro 11	0,292485	23657
SPP1	Fosfoproteína 1 secretada	0,183821	6696
TNNT1	Troponina T1, tipo esquelético lento	0,396711	7138
UBE2E1	Ubiquitina conjugando la enzima E2 E1	0,033286	7324

EQUIVALENTES

La invención puede materializarse en otras formas específicas sin apartarse del espíritu o de las características esenciales de la misma. Las modalidades anteriores son, por tanto, para ser consideradas en todos los aspectos ilustrativas en lugar de taxativas de la invención descrita en este documento. El alcance de la invención está indicado por las reivindicaciones adjuntas más que por

la descripción anterior, y todos los cambios que se encuentran comprendidos dentro del significado y el rango de equivalencia de las reivindicaciones, están destinados a ser incluidos en las mismas.

INCORPORACIÓN POR REFERENCIA

5 La descripción completa de cada uno de los documentos de patentes y artículos científicos que se hace referencia en el presente documento se incorpora por referencia para todos los propósitos.

10

15

20

25

30

35