

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 608 322**

51 Int. Cl.:

C12Q 1/68

(2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **30.07.2012** **PCT/EP2012/064865**

87 Fecha y número de publicación internacional: **31.01.2013** **WO13014296**

96 Fecha de presentación y número de la solicitud europea: **30.07.2012** **E 12740974 (6)**

97 Fecha y número de publicación de la concesión europea: **28.09.2016** **EP 2737081**

54 Título: **Procedimiento para predecir la respuesta a la quimioterapia en un paciente que padece o está en riesgo de desarrollar cáncer de mama recurrente**

30 Prioridad:

28.07.2011 EP 11175852

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

07.04.2017

73 Titular/es:

SIVIDON DIAGNOSTICS GMBH (100.0%)
Nattermann Allee 1
50829 Köln, DE

72 Inventor/es:

GEHRMANN, MATHIAS;
WEBER, KARSTEN;
KRONENWETT, RALF;
PETRY, CHRISTOPH y
BRASE, JAN

74 Agente/Representante:

CARPINTERO LÓPEZ, Mario

ES 2 608 322 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín Europeo de Patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre Concesión de Patentes Europeas).

DESCRIPCIÓN

Procedimiento para predecir la respuesta a la quimioterapia en un paciente que padece o está en riesgo de desarrollar cáncer de mama recurrente

Campo técnico

- 5 La presente invención se refiere a procedimientos, kits y sistemas para predecir la respuesta de un tumor a la quimioterapia. De forma más específica, la presente invención se refiere a la predicción de la respuesta a agentes quimioterapéuticos, en particular pero sin limitación a un contexto neoadyuvante a base de las mediciones de los niveles de expresión de genes en muestras de tumor de pacientes de cáncer de mama.

Antecedentes de la invención

- 10 El cáncer de mama es el tipo de tumor más común y una de las causas principales de muerte relacionada con el cáncer en las mujeres (Jemal y col., CA Cancer J Clin., 2011). Se estima que una de cada diez mujeres desarrollará su vida cáncer de mama durante. Aunque la incidencia ha aumentado a lo largo de los años, la mortalidad ha disminuido de forma constante debido a los avances en la detección temprana y el desarrollo de estrategias de tratamiento eficaces nuevas.
- 15 Los pacientes con cáncer de mama con frecuencia se tratan tras la cirugía con radioterapia, terapia hormonal o quimioterapia citotóxica (tratamiento adyuvante) para controlar células tumorales residuales y reducir el riesgo de recidiva. La quimioterapia incluye el uso combinado de varios agentes citotóxicos, mientras que las estrategias de tratamiento a base antraciclina y taxano han demostrado ser superiores en comparación con otras terapias de combinación convencionales (Misset y col., J Clin Oncol., 1996, Henderson y col., J Clin Oncol., 2003).
- 20 La quimioterapia sistémica se aplica de forma común para reducir la probabilidad de recidiva en tumores positivos para HER2/neu y en tumores que carecen de la expresión del receptor de estrógenos y del receptor de HER2/neu (triple negativo, basal). La decisión sobre el tratamiento más difícil se refiere a los tumores luminales (positivos para el receptor de estrógenos y negativos para el receptor de HER2/neu), para los que factores clínicos clásicos como la clasificación, el tamaño tumoral y la implicación de los ganglios linfáticos no proporciona una respuesta clara a la
- 25 cuestión de si usa o no la quimioterapia.

Para reducir el número de pacientes que padecen efectos secundarios graves sin un claro beneficio de la terapia sistémica, existe una gran necesidad de nuevos biomarcadores moleculares para predecir la sensibilidad a la quimioterapia y permitir así una estrategia de tratamiento adaptada.

- 30 La quimioterapia también puede aplicarse en el contexto neoadyuvante (prequirúrgica), en la que los pacientes de cáncer de mama reciben terapia sistémica antes de que se extraigan mediante cirugía las células tumorales restantes. La quimioterapia neoadyuvante del cáncer de mama temprano conduce a elevadas tasas de respuestas clínicas del 70-90 %. Sin embargo, en la mayoría de los respondedores clínicos, la evaluación patológica de los restos tumorales revela la presencia de focos residuales de células tumorales. A una completa erradicación de las células cancerosas en la mama y los ganglios linfáticos tras el tratamiento neoadyuvante se le llama respuesta
- 35 patológica completa (RpC) y se observa en solo el 10-25 % de todos los pacientes. La RpC es un marcador sustituto apropiado para la supervivencia sin enfermedad y un fuerte indicador de beneficio de la quimioterapia.

- La estrategia del tratamiento prequirúrgico proporciona la oportunidad de evaluar de forma directa la respuesta de un tumor particular a la terapia aplicada: la reducción de la masa tumoral en respuesta a la terapia puede controlarse de forma directa. Para los pacientes con una baja probabilidad de respuesta deberían considerarse otras estrategias
- 40 terapéuticas. Para identificar los marcadores predictivos más valiosos pueden analizarse los biomarcadores en biopsias con aguja gruesa preterapéuticas. Una estrategia común es aislar ARN de biopsias con aguja gruesa para el análisis de la expresión de genes antes de la terapia neoadyuvante. Más adelante, el éxito terapéutico se puede evaluar de forma directa mediante la reducción tumoral y se puede correlacionar con los datos de expresión de genes.

- 45 Los ensayos multigénicos predictivos, como el DLDA30 (Hess y col., J Clin Oncol., 2006), han demostrado proporcionar información más allá de los parámetros clínicos, como la graduación del tumor y el estado del receptor de hormonas en los pacientes de cáncer de mama tratados con terapia neoadyuvante. Sin embargo, la prueba multigénica predictiva DLDA30 se estableció sin tener en cuenta el estado del receptor de estrógenos. Por lo tanto, la prueba podría reflejar las diferencias fenotípicas entre un respondedor completo y un no respondedor, siendo los
- 50 respondedores predominantemente negativos para ER y positivos para HER2/neu (Tabchy y col., Clin Can Res, 2010).

- De forma adicional, las pruebas multigénicas establecidas para pronóstico se analizaron en el contexto neoadyuvante para evaluar si los ensayos pronósticos también pueden predecir la quimiosensibilidad. Un ejemplo es
- 55 el Índice de Grado Genómico (IGG), una prueba multigénica para definir el grado histológico a base de los perfiles de expresión de genes (Sotiriou y col., JNCI, 2006). Liedtke y colaboradores demostraron que un alto IGG está asociado con quimiosensibilidad aumentada en pacientes de cáncer de mama tratados con terapia neoadyuvante

(Liedtke, J Clin Oncol, 2009).

Aunque se ha demostrado que las firmas genéticas predicen la respuesta a la terapia, faltan los estudios de validación a gran escala incluyendo los datos de seguimiento clínico y, hasta ahora, ninguno de ellos se utiliza comúnmente para guiar las decisiones sobre el tratamiento en la rutina clínica.

5 El documento WO2010/076322 A1 desvela un procedimiento para predecir una respuesta a y/o el beneficio de una quimioterapia en un paciente que padece cáncer, que comprende las etapas (i) clasificar un tumor en al menos dos clases, (ii) determinar en una muestra de tumor la expresión de al menos un gen marcador indicativa de una
10 respuesta a la quimioterapia para un tumor en cada una de las clases respectivas, (iii) dependiendo de dicha expresión de genes, predecir dicha respuesta y/o beneficio; en el que dicho al menos un gen marcador comprende un gen seleccionado del grupo que consiste en TMSL8, ABCC1, EGFR, MVP, ACOX2, HER2/NEU, MYH11, TOB1, AKR1C1, ERBB4, NFKB1A, TOP2A, AKR1C3, ESR1, OLFM1, TOP2B, ALCAM, FRAP1, PGR, TP53, BCL2, GADD45A, PRKAB1, TUBA1A, C16orf45, HIF1A, PTPRC, TUBB, CA12, IGKC, RACGAP1, UBE2C, CD14, 1KBKB, S100A7, VEGFA, CD247, KRT5, SEPT8, YBX1, CD3D, MAPK3, SLC2A1, CDKN1A, MAPT, SLC7A8, CHPT1, MLPH, SPON1, CXCL13, MMP1, STAT1, CXCL9, MMP7, STC2, DCN, MUC1, STMN1 y combinaciones de los
15 mismos.

Maia Chanrion y col. informan en Clin Cancer Res 2008; 14(6), 15 de marzo de 2008, p. 1744-1752 sobre una firma de expresión génica que puede predecir la recidiva del cáncer de mama primario tratado con tamoxifeno. El estudio desvelado identifica una firma molecular que especifica un subgrupo de pacientes que no obtienen beneficios del tratamiento con tamoxifeno. Por lo tanto, estos pacientes pueden ser adecuados para terapias endocrinas
20 alternativas y/o para quimioterapia.

El documento WO 2009/158143A1 desvela procedimientos para clasificar y para evaluar el pronóstico de un sujeto que tiene cáncer de mama. Los procedimientos incluyen la predicción del subtipo de cáncer de mama utilizando un algoritmo supervisado entrenado para estratificar sujetos a base del subtipo intrínseco del cáncer de mama. El modelo de predicción es a base del perfil de expresión de genes de los genes intrínsecos enumerados en la Tabla 1.
25 Este modelo de predicción puede utilizarse para predecir de forma precisa el subtipo intrínseco de un sujeto diagnosticado con o que se sospecha que tiene cáncer de mama. Adicionalmente, se proporcionan composiciones y procedimientos para predecir el efecto de o la respuesta a la terapia de un sujeto diagnosticado con o que se sospecha que tiene cáncer de mama. Estos procedimientos son útiles para guiar o determinar las opciones de tratamiento para un sujeto aquejado de cáncer de mama. Los procedimientos de la invención incluyen
30 adicionalmente medios para evaluar los perfiles de expresión de genes, incluyendo micromatrices y ensayos de reacción en cadena de la polimerasa cuantitativos, así como kits que comprenden reactivos para la práctica de los procedimientos de la invención.

El documento WO 2006/119593 desvela procedimientos y sistemas para determinación del pronóstico en muestras de tumor, mediante la medición de la expresión de genes en una muestra de tumor y la aplicación de un índice de
35 grado de expresión de genes (IGG) o una puntuación de recaída (PR) para proporcionar una puntuación de riesgo numérica.

Karen J Taylor y col. informan en Breast Cancer Research 2010, 12:R39 sobre los cambios dinámicos en la expresión de genes *in vivo*, para predecir el pronóstico de pacientes tratados con tamoxifeno con cáncer de mama.

40 El documento WO 2008/006517A2 desvela procedimientos y kits para la predicción de un efecto enérgico de la quimioterapia en un paciente de cáncer. De forma más específica, la invención se refiere a la predicción de la respuesta del tumor a la quimioterapia a base de las mediciones de los niveles de expresión de un pequeño conjunto de genes marcadores. El conjunto de genes marcadores es útil para la identificación de subtipos de cáncer de mama sensibles a la quimioterapia a base de taxano, tal como por ejemplo una quimioterapia a base de taxano-antraciclina-ciclofosfamida (por ejemplo a base de Taxotere (docetaxel)-Adriamicina (doxorubicina)-ciclofosfamida (es decir a base de TAC).
45

En el documento WO 2009/114836 A1 se desvelan conjuntos de genes que son útiles en la evaluación del pronóstico y/o la predicción de la respuesta del cáncer, por ejemplo cáncer colorrectal, a la quimioterapia. También se desvela una prueba de cáncer validada de forma clínica, por ejemplo, una prueba colorrectal, para la evaluación del pronóstico y/o predicción de la respuesta del paciente a la quimioterapia, utilizando el análisis de expresión. Se
50 incluye el uso de material de biopsia incluido en parafina archivado, para el ensayo de todos los marcadores en los conjuntos de genes importantes, y por lo tanto, es compatible con el tipo de material de biopsia más ampliamente disponible.

El documento WO 2011/120984A1 desvela procedimientos, kits y sistemas para el pronóstico del efecto de la enfermedad sobre cáncer de mama, comprendiendo dicho procedimiento: (a) determinar en una muestra de tumor
55 de dicho paciente los niveles de expresión de ARN de al menos 2 de los siguientes 9 genes: UBE2C, BIRC5, RACGAP1, DHCR7, STC2, AZGP1, RBBP8, IL6ST y MGP, (b) combinar de forma matemática los valores del nivel de expresión para los genes de dicho conjunto, valores que se determinaron en la muestra de tumor para proporcionar una puntuación combinada, en los que dicha puntuación combinada es indicativa de un pronóstico de

dicho paciente; y kits y sistemas para realizar dicho procedimiento.

Definiciones

A menos que se defina otra cosa, los términos técnicos y científicos utilizados en el presente documento tienen el mismo significado que entiende comúnmente un experto en la materia a la que pertenece la invención.

- 5 “Predecir la respuesta a la quimioterapia” debería entenderse dentro del significado de la invención como el acto de determinar un efecto probable de la quimioterapia citotóxica en un paciente afectado por cáncer. La predicción de una respuesta se hace preferentemente en referencia a valores de probabilidad de alcanzar un efecto deseado o no deseado de la quimioterapia. Los procedimientos predictivos de la presente invención pueden utilizarse de forma clínica para hacer las decisiones de tratamiento, mediante la elección de las modalidades del tratamiento más apropiadas para cualquier paciente particular.

10 La “respuesta de un tumor a la quimioterapia”, se refiere dentro del significado de la invención a cualquier respuesta del tumor a la quimioterapia citotóxica, preferentemente a un cambio en la masa y/o el volumen tumoral, tras el inicio de la quimioterapia neoadyuvante, y/o la prolongación del tiempo para la metástasis a distancia o el tiempo para la muerte después de la quimioterapia neoadyuvante o adyuvante. La respuesta tumoral se puede evaluar en una situación neoadyuvante en donde el tamaño de un tumor tras la intervención sistémica se puede comparar con el tamaño y las dimensiones iniciales, como se mide mediante TC, PET, mamograma, ultrasonido o palpación, registrado habitualmente como “respuesta clínica” de un paciente. Además, la respuesta puede evaluarse mediante la medición con calibrador o el examen patológico del tumor tras la biopsia o la extirpación quirúrgica. La respuesta puede registrarse de un modo cuantitativo, como el cambio porcentual en el volumen tumoral o de un modo cualitativo, como “sin cambio” (SC), “remisión parcial” (RP), “remisión completa” (RC) u otro criterio cualitativo. La evaluación de la respuesta del tumor se puede hacer de forma temprana tras el comienzo de la terapia neoadyuvante, por ejemplo tras unas pocas horas, días, semanas o preferentemente tras unos pocos meses. Un criterio de valoración típico para la evaluación de la respuesta es tras la finalización de la quimioterapia neoadyuvante o tras la extirpación quirúrgica de las células tumorales residuales y/o del lecho del tumor.

15 20 25 Normalmente, esto es tres meses después del inicio de la terapia neoadyuvante. También, la respuesta puede evaluarse comparando el tiempo hasta la metástasis distal o la muerte de un paciente después de la quimioterapia neoadyuvante o adyuvante, con el tiempo hasta la metástasis a distancia o hasta la muerte de un paciente no tratado con quimioterapia.

30 El término “tumor” como se utiliza en el presente documento, se refiere a todo crecimiento y proliferación de células neoplásicas, ya sean malignas o benignas, y a todas las células y tejidos precancerosos y cancerosos.

El término “cáncer” se refiere a, o describe, el estado fisiológico en mamíferos que normalmente se caracteriza por el crecimiento celular no regulado. El término “cáncer” como se utiliza en el presente documento incluyen carcinomas (por ejemplo, carcinoma *in situ*, carcinoma invasivo, carcinoma metastásico) y estados pre malignos, cambios neomórficos independientes de su origen histológico. El término “cáncer” no se limita a alguna fase, grado, característica histomorfológica, invasividad, agresividad o malignidad de un tejido afectado o agregación celular. En particular se incluyen el cáncer de fase 0, el cáncer de fase I, el cáncer de fase II, el cáncer de fase III, el cáncer de fase IV, el cáncer de grado I, el cáncer de grado II, el cáncer de grado III, el cáncer maligno y los carcinomas primarios.

40 La expresión “quimioterapia citotóxica” se refiere a diversas modalidades de tratamiento que afectan a la proliferación celular y/o la supervivencia. El tratamiento puede incluir la administración de agentes alquilantes, antimetabolitos, antraciclinas, alcaloides vegetales, inhibidores de la topoisomerasa y otros agentes antitumorales, incluyendo anticuerpos monoclonales e inhibidores de quinasas. En particular, el tratamiento citotóxico puede relacionarse a un tratamiento con taxano. Los taxanos son alcaloides vegetales que bloquean la división celular mediante el impedimento de la función de los microtúbulos. El taxano prototipo es el producto natural paclitaxel, conocido de forma original como Taxol y obtenido al principio de la corteza del Tejo del Pacífico. El docetaxel es un análogo semi sintético del paclitaxel. Los taxanos potencian la estabilidad de los microtúbulos, impidiendo la separación de los cromosomas durante la anafase.

50 El término “terapia” se refiere a la administración secuencial o simultánea en el tiempo de agentes anti tumorales, y/o de agentes anti vasculares y/o anti estroma, y/o estimulantes o supresores del sistema inmunitario, y/o proliferativos de hematocitos y/o de terapia de radiación, y/o de hipertermia, y/o de la hipotermia para la terapia del cáncer. La administración de estos se puede realizar de un modo adyuvante y/o neoadyuvante. La composición de tal “protocolo” puede variar en la dosis de cada uno de los agentes individuales, el periodo de tiempo de aplicación y la frecuencia de la administración dentro de una ventana terapéutica definida. Actualmente están en investigación diversas combinaciones de diversos fármacos y/o procedimientos físicos, y diversos programas. Una “quimioterapia que contiene taxano/antraciclina” es una modalidad de terapia que comprende la administración de taxano y/o de antraciclina y de derivados terapéuticamente eficaces de los mismos.

La expresión “quimioterapia neoadyuvante” se refiere a un régimen de terapia prequirúrgico que consiste en un panel de agentes hormonales, quimioterápicos y/o de anticuerpos, que se desea que retraiga el tumor primario,

haciendo de este modo a la terapia local (cirugía o radioterapia) menos destructiva o más eficaz, permitiendo la cirugía conservadora de la mama y la evaluación de la respuesta de sensibilidad del tumor hacia los agentes específicos *in vivo*.

La expresión “implicación de los ganglios linfáticos” significa un paciente al que se le ha diagnosticado de forma previa metástasis en ganglios linfáticos. Abarcará a los ganglios linfáticos de drenaje, a los ganglios linfáticos cercanos y a la metástasis en los ganglios linfáticos distantes. Este diagnóstico previo no formará en sí parte del procedimiento de la invención. Más bien, es una precondition para la selección de pacientes cuyas muestras puedan utilizarse para una realización de la presente invención. Se puede haber llegado a este diagnóstico previo mediante cualquier procedimiento adecuado conocido en la técnica, incluyendo, pero sin limitación, la extirpación de ganglios linfáticos y el análisis patológico, el análisis de biopsias, análisis *in vitro* de biomarcadores indicativos de metástasis, procedimientos de formación de imágenes (por ejemplo tomografía computarizada, rayos X, formación de imágenes por resonancia magnética, ultrasonido) y hallazgos intraquirúrgicos.

La expresión “respuesta patológica completa” (RpC), como se utiliza en el presente documento, se refiere a una desaparición completa o a la ausencia de células tumorales invasivas en la mama y/o en los ganglios linfáticos, según se evalúa mediante un examen histopatológico de la muestra de estudio quirúrgica después de la quimioterapia neoadyuvante.

El término “marcador” o “biomarcador” se refiere a una molécula biológica, por ejemplo, un ácido nucleico, péptido, proteína, hormona, etc., cuya presencia o concentración se puede detectar y correlacionar con una afección conocida, tal como una enfermedad.

La expresión “marcador predictivo” se refiere a un marcador que puede utilizarse para predecir la respuesta clínica de un paciente hacia un tratamiento dado.

El término “pronóstico”, como se utiliza en el presente documento, se refiere a una evaluación individual de la malignidad de un tumor, o de la respuesta esperada si no hay terapia farmacológica. En contraste al mismo, el término “predicción” se refiere a una evaluación individual de la malignidad de un tumor, o a la respuesta esperada si la terapia contiene un fármaco, en comparación con la malignidad o respuesta sin este fármaco.

El término “inmunohistoquímica” o IHQ se refiere al procedimiento de localizar proteínas en células de un corte de tejido aprovechando el principio de la unión de los anticuerpos a los antígenos de forma específica en tejidos biológicos. La tinción inmunohistoquímica se utiliza ampliamente en el diagnóstico y el tratamiento del cáncer. Los marcadores moleculares específicos son característicos de los tipos de cáncer particulares. Además, la IHQ se utiliza ampliamente en la investigación básica para entender la distribución y localización de biomarcadores en distintas partes de un tejido.

El término “muestra”, como se utiliza en el presente documento, se refiere a una muestra obtenida a partir de un paciente. La muestra puede ser de cualquier tejido o fluido biológico. Tales muestras incluyen, pero sin limitación, esputo, sangre, suero, plasma, células sanguíneas (por ejemplo, leucocitos), tejido, muestras de biopsia con aguja gruesa o con aguja fina, fluidos corporales que contienen células, ácidos nucleicos que flotan libres, orina, fluido peritoneal y fluido pleural, o células procedentes de los mismos. Las muestras biológicas también pueden incluir cortes de tejidos tales como cortes congelados o fijados tomados para fines histológicos o células microdisecionadas o partes extracelulares de los mismos. Una muestra biológica a analizar es material de tejido procedente de una lesión neoplásica tomada mediante aspiración o punción, escisión o mediante cualquier otro procedimiento quirúrgico que conduzca al material celular de biopsia o extirpado. Tal muestra biológica puede comprender células obtenidas a partir de un paciente. Las células pueden encontrarse en un “frotis” de células recolectado, por ejemplo, mediante una aspiración del pezón, de un lavado ductal, de una biopsia de aguja fina o a partir de la descarga del pezón provocada o espontánea. En otra realización la muestra es un fluido corporal. Tales fluidos incluyen, por ejemplo, pero sin limitación a estos fluidos, fluidos sanguíneos, suero, plasma, linfa, fluidos ascíticos, fluidos ginecológicos u orina.

Una “muestra de tumor” es una muestra que contiene material tumoral, por ejemplo material de tejido procedente de una lesión neoplásica tomado mediante aspiración o punción, escisión o mediante cualquier otro procedimiento quirúrgico que conduzca al material celular de biopsia o extirpado, incluyendo material conservado, tal como material recién congelado, material fijado en formalina, material incluido en parafina, y similares. Tal muestra biológica puede comprender células obtenidas a partir de un paciente. Las células pueden encontrarse en un “frotis” recolectado, por ejemplo, mediante una aspiración del pezón, un lavado ductal, una biopsia de aguja fina o a partir de la descarga del pezón provocada o espontánea. En otra realización la muestra es un fluido corporal. Tales fluidos incluyen, pero sin limitación a estos fluidos, por ejemplo fluidos sanguíneos, suero, plasma, linfa, fluidos ascíticos, fluidos ginecológicos u orina.

La expresión “combinación de niveles de expresión de forma matemática”, dentro del significado de la invención se entenderá como la obtención de un valor numérico procedente de un determinado nivel de expresión de un gen y la aplicación de un algoritmo a uno o más de tales valores numéricos, para obtener un valor numérico combinado o puntuación combinada.

Una “puntuación” dentro del significado de la invención se entenderá como un valor numérico, que está relacionado con el efecto de la enfermedad del paciente y/o la respuesta de un tumor a la quimioterapia. El valor numérico se obtiene mediante la combinación de los niveles de expresión de los genes marcadores utilizando coeficientes pre-especificados en un algoritmo matemático. Los niveles de expresión pueden emplearse como valores CT o delta CT obtenidos mediante RT-PCR cinética, como valores de intensidad de fluorescencia absoluta o relativa, obtenidos a través de micromatrices o mediante cualquier otro procedimiento útil para cuantificar niveles de ARN absolutos o relativos. La combinación de estos niveles de expresión se puede llevar a cabo, por ejemplo, multiplicando cada nivel de expresión por un coeficiente definido y especificado, y sumando tales productos para producir una puntuación. La puntuación también puede obtenerse de los niveles de expresión junto con otra información, por ejemplo, los datos clínicos como el tamaño tumoral, el estado de los ganglios linfáticos o la clasificación tumoral, y tales variables también pueden codificarse como números en una ecuación. La puntuación puede utilizarse en una escala continua para predecir la respuesta de un tumor a la quimioterapia y/o el efecto de una enfermedad del paciente. Los valores límite se pueden aplicar para distinguir subgrupos clínicos importantes. Los valores límite para tales puntuaciones se pueden determinar del mismo modo que los valores límite para los marcadores de diagnóstico convencionales y son bien conocidos para los expertos en la materia. Un modo útil para determinar tal valor límite es construir una curva operativa del receptor (curva ROC) a base de todos los valores límites concebibles y determinar el punto individual en la curva ROC con la mayor proximidad a la esquina superior izquierda (0/1) en la representación ROC. Obviamente, la mayoría de los valores límite temporales se determinarán mediante procedimientos menos formalizados, eligiendo la combinación de sensibilidad y especificidad determinada mediante tal valor límite, proporcionando la información médica más beneficiosa para el problema investigado.

La expresión “un procedimiento basado en PCR” como se utiliza en el presente documento se refiere a los procedimientos que comprenden una reacción en cadena de la polimerasa (PCR). Esta es una estrategia para amplificar de forma exponencial ácidos nucleicos, como ADN o ARN, a través de replicación enzimática, sin utilizar un organismo vivo. Dado que la PCR es una técnica *in vitro*, se puede realizar sin restricciones en la forma del ADN, y puede modificarse de forma extensa para realizar una amplia serie de manipulaciones genéticas. Cuando se trata de la determinación de los niveles de expresión, puede utilizarse por ejemplo un procedimiento basado en PCR para detectar la presencia de un dado ARNm mediante (1) transcripción inversa del grupo de ARNm completo (el así llamado transcriptoma) en ADNc, con la ayuda de una enzima transcriptasa inversa, y (2) detectar la presencia de un dado ADNc con ayuda de los cebadores respectivos. Esta estrategia se conoce comúnmente como PCR por transcriptasa inversa (rtPCR). Además, los procedimientos basados en PCR comprenden por ejemplo PCR en tiempo real y, particularmente adecuada para el análisis de los niveles de expresión, la PCR cinética o cuantitativa (qPCR).

Una “micromatriz”, en el presente documento también denominada como un “biochip” o “chip biológico”, es una matriz de regiones que tengan una densidad de regiones discretas de al menos aproximadamente 100/cm², y preferentemente de al menos aproximadamente 1000/cm². Las regiones en una micromatriz tienen dimensiones típicas, por ejemplo, diámetros, en el intervalo de entre aproximadamente 10-250 µm, y están separadas de otras regiones en la matriz mediante aproximadamente la misma distancia.

La expresión “procedimiento basado en hibridación”, como se utiliza en el presente documento, se refiere a procedimientos que imparten un procedimiento de combinación de ácidos nucleicos o análogos de nucleótido monocatenarios complementarios, en una molécula bicatenaria única. Los nucleótidos o análogos de nucleótido se unirán a sus complementos en condiciones normales, y así dos cadenas perfectamente complementarias se unirán entre sí fácilmente. En la bioanalítica, las sondas monocatenarias, a menudo marcadas, son para encontrar secuencias diana complementarias. Si tales secuencias existen en la muestra, las sondas hibridarán con dichas secuencias, lo que después pueden detectarse debido al marcador. Otros procedimientos basados en hibridación comprenden procedimientos de micromatriz y/o de biochip. En ellos, las sondas están inmovilizadas sobre una fase sólida, la que después se expone a una muestra. Si están presentes ácidos nucleicos complementarios en la muestra, estos hibridarán con las sondas y así pueden detectarse. Estas estrategias también son conocidas como “procedimientos basados en matrices”. Aún otro procedimiento basado en hibridación es la PCR, la cual se describe anteriormente. Cuando se trata de la determinación de los niveles de expresión, los procedimientos basados en hibridación pueden por ejemplo utilizarse para determinar la cantidad de ARNm para un dado gen.

La expresión “gen marcador”, como se utiliza en el presente documento, se refiere a un gen expresado de forma diferencial cuyo patrón de expresión puede utilizarse como parte de un procedimiento predictivo, pronóstico o diagnóstico en la evaluación de la neoplasia maligna o el cáncer, o que, como alternativa, puede utilizarse en procedimientos para identificar compuestos útiles para el tratamiento o la prevención de la neoplasia maligna y en particular del cáncer de cabeza y cuello, de colon o de mama. Un gen marcador también puede tener las características de un gen diana.

Un “algoritmo” es un procedimiento que realiza alguna secuencia de operaciones para producir información.

La expresión “medición de un nivel de proteína”, como se utiliza en el presente documento, se refiere a los procedimientos que permiten la determinación cuantitativa y/o cualitativa de una o más proteínas en una muestra. Estos procedimientos incluyen, entre otros, la purificación de proteínas, incluyendo ultracentrifugación, precipitación y cromatografía, así como el análisis y la determinación de proteínas, incluyendo inmunohistoquímica,

inmunofluorescencia, ELISA (inmunoensayo ligado a enzimas), RIA (radioinmunoensayo) o el uso de micromatrices de proteínas, la exploración por doble híbrido, los procedimientos de transferencia incluyendo transferencia de western, electroforesis en geles mono y bidimensionales, isoelectroenfoque, así como procedimientos que están basados en espectrometría de masas, como MALDI-TOF y similares.

5 La expresión "PCR cinética" o "PCR cuantitativa" (qPCR) se refiere a cualquier tipo de procedimiento de PCR que permite la cuantificación del molde en una muestra. La PCR en tiempo real cuantitativa comprende distintas técnicas de realización o de detección de producto, como por ejemplo la técnica TaqMan o la técnica LightCycler. La técnica TaqMan, por ejemplo, utiliza una sonda fluorogénica marcada de forma dual. La PCR en tiempo real TaqMan mide la
10 acumulación de un producto a través del fluoróforo durante las fases exponenciales de la PCR, más bien que en el punto final como en la PCR convencional. El aumento exponencial de producto se utiliza para determinar el ciclo umbral, CT (sigla del inglés *cycle threshold*), es decir, el número de ciclos de PCR en el cual se detecta un aumento exponencial de la fluorescencia significativo, y que se correlaciona de forma directa con el número de copias de ADN molde presente en la reacción. La preparación de la reacción es muy similar a la PCR convencional, pero se lleva a cabo en un ciclador térmico en tiempo real que permite la medición de moléculas fluorescentes en los tubos PCR. A
15 diferencia de la PCR regular, en la PCR en tiempo real TaqMan se añade una sonda a la reacción, es decir, un oligonucleótido monocatenario complementario a un segmento de 20-60 nucleótidos dentro del molde de ADN y emplazado entre los dos cebadores. Un indicador fluorescente o fluoróforo (por ejemplo 6-carboxifluoresceína, sigla: FAM, o tetraclorofluoresceína, sigla: TET) y un extintor (por ejemplo, tetrametilrodamina, sigla: TAMRA, o un dihidrociclopirroloindol tripéptido "ligante del surco menor", sigla: MGB) están acoplados de forma covalente a los
20 extremos 5' y 3' de la sonda, respectivamente [2]. La estrecha proximidad entre el fluoróforo y el extintor acoplados a la sonda inhibe la fluorescencia procedente del fluoróforo. Durante la PCR, a medida que se inicia la síntesis de ADN la actividad exonucleasa 5' a 3' de la polimerasa Taq degrada la proporción de la sonda que se ha apareado con el molde (de ahí su nombre: polimerasa Taq + PacMan). La degradación de la sonda libera al fluoróforo de ella y rompe la estrecha proximidad con el extintor, atenuando así el efecto extintor y permitiendo la fluorescencia del
25 fluoróforo. Por lo tanto, la fluorescencia detectada en el ciclador térmico de PCR en tiempo real es directamente proporcional al fluoróforo liberado y la cantidad de molde de ADN presente en la PCR.

"Cebador" y "sondas", dentro del significado de la invención, tendrá el significado ordinario de estos términos que son bien conocidos para el experto en la materia de biología molecular. En una realización preferente de la invención, "cebador" y "sondas" se entenderán como moléculas de polinucleótido que tienen una secuencia idéntica,
30 complementaria, homóloga u homóloga a los complementos de las regiones de un polinucleótido diana que se va a detectar o cuantificar. En aún otra realización, están también comprendidos para el uso como cebadores y/o sondas los análogos de nucleótido y/o morfolinos. "Sondas marcadas de forma individual" dentro del significado de la invención, se entenderá que son sondas moleculares que comprenden un polinucleótido, oligonucleótido o análogo de nucleótido y un marcador, útiles en la detección o cuantificación de la sonda. Los marcadores preferentes son
35 moléculas fluorescentes, moléculas luminiscentes, moléculas radioactivas, moléculas enzimáticas y/o moléculas extintoras.

Objeto de la invención

Es un objeto de la presente invención proporcionar un procedimiento mejorado para la predicción de una respuesta de un tumor en un paciente que padece o está en riesgo de desarrollar una enfermedad neoplásica - en particular
40 cáncer de mama - para al menos un modo dado de tratamiento.

Es otro objeto de la presente invención evitar la quimioterapia citotóxica adyuvante y/o neoadyuvante innecesaria en pacientes que padecen una enfermedad neoplásica, en especial cáncer de mama.

Es otro objeto de la presente invención ofrecer para las muestras de tejido clínicas fijadas de rutina, un sistema de ensayo diagnóstico más sólido y específico que la inmunohistoquímica convencional que ayude mejor al médico a
45 seleccionar modalidades de tratamiento individualizadas.

En una realización más preferente el procedimiento desvelado puede utilizarse para seleccionar una terapia adecuada para una enfermedad neoplásica, en particular para cánceres de mama.

Es otro objeto de la presente invención detectar nuevas dianas para fármacos dirigidos recientemente disponibles, o para determinados fármacos todavía por desarrollar.

Sumario de la invención

Debe indicarse que, como se utiliza en la memoria descriptiva y en las reivindicaciones adjuntas, las formas singulares "un", "una" y "el/la" incluyen referentes singulares y/o plurales, a menos que el contexto indique claramente otra cosa. Además debe entenderse que, en el caso que se proporcionen intervalos de parámetros que estén delimitados por valores numéricos, se considera que los intervalos incluyen estos valores de límite.

55 Los problemas anteriores se resuelven mediante procedimientos y medios proporcionados por la invención.

El estado del receptor de estrógeno en general se determina utilizando inmunohistoquímica. El estado de HER2/NEU (ERBB2) en general se determina utilizando inmunohistoquímica e hibridación *in situ* fluorescente. Sin embargo, el estado del receptor de estrógenos y el estado de HER2/NEU (ERBB2) pueden, para los fines de la invención, determinarse mediante cualquier procedimiento adecuado, por ejemplo, inmunohistoquímica, hibridación *in situ* fluorescente (FISH) o análisis de expresión de genes.

Se desvela un procedimiento para predecir una respuesta a y/o el beneficio de la quimioterapia, incluyendo la quimioterapia neoadyuvante, en un paciente que padece o está en riesgo de desarrollar enfermedad neoplásica recurrente, en particular cáncer de mama. Dicho procedimiento comprende las etapas de:

(a) determinar en una muestra de tumor procedente de dicho paciente los niveles de expresión de genes de al menos 3 de los siguientes 9 genes: UBE2C, BIRCS, RACGAP1, DHCR7, STC2, AZGP1, RBBP8, IL6ST y MGP

(b) combinar de forma matemática los valores del nivel de expresión de los genes de dicho conjunto, cuyos valores se determinaron en la muestra de tumor para producir una puntuación combinada, en los que dicha puntuación combinada está prediciendo dicha respuesta y/o el beneficio de la quimioterapia.

El documento WO 2011/120984A1 utiliza los nueve genes, sin embargo, para predecir un efecto del cáncer de mama en un tumor de un paciente de cáncer de mama positivo para receptor de estrógenos y negativo para HER2, que no está relacionado con el procedimiento de la presente invención que está prediciendo una respuesta para/o el beneficio de la quimioterapia. Los genes de la presente invención se utilizan para un objeto distinto.

El procedimiento de la invención para predecir una respuesta a y/o el beneficio de la quimioterapia con taxano o antraciclina, incluyendo quimioterapia neoadyuvante, en un paciente que padece o está en riesgo de desarrollar cáncer de mama recurrente, comprende las etapas de:

a) determinar en una muestra de tumor de mama procedente de dicho paciente los niveles de expresión de ARN de los siguientes 8 genes: UBE2C, RACGAP1, DHCR7, STC2, AZGP1, RBBP8, IL6ST y MGP, indicativos de una respuesta a la quimioterapia para un tumor o

b) determinar en una muestra de tumor de mama procedente de dicho paciente los niveles de expresión de ARN de los siguientes 8 genes: UBE2C, BIRC5, DHCR7, STC2, AZGP1, RBBP8, IL6ST y MGP; indicativos de una respuesta a la quimioterapia para un tumor

(c) combinar de forma matemática los valores del nivel de expresión para los genes de dicho conjunto, cuyos valores se determinaron en la muestra de tumor para producir una puntuación combinada, en la que dicha puntuación combinada está prediciendo dicha respuesta y/o beneficio de la quimioterapia.

Una realización adicional desvela un procedimiento que comprende:

(a) determinar en una muestra de tumor procedente de dicho paciente los niveles de expresión de ARN de los siguientes 8 genes: UBE2C, BIRCS, DHCR7, STC2, AZGP1, RBBP8, IL6ST y MGP; indicativos de una respuesta a la quimioterapia para un tumor, **aunque BIRCS** puede reemplazarse por UBE2C o TOP2A o RACGAP1 o AURKA o NEK2 o E2F8 o PCNA o CYBRD1 o DCN o ADRA2A o SQLE o CXCL12 o EPHX2 o ASPH o PRSS16 o EGFR o CCND1 o TRIM29 o DHCR7 o PIP o TFAP2B o WNT5A o APOD o PTPRT, a condición de que tras un reemplazo se seleccionen 8 genes distintos; y

aunque UBE2C puede reemplazarse por BIRC5 o RACGAP1 o TOP2A o AURKA o NEK2 o E2F8 o PCNA o CYBRD1 o ADRA2A o DCN o SQLE o CCND1 o ASPH o CXCL12 o PIP o PRSS16 o EGFR o DHCR7 o EPHX2 o TRIM29, a condición de que tras un reemplazo se seleccionen los 8 genes distintos; y

aunque DHCR7 puede reemplazarse por AURKA, BIRC5, UBE2C o por cualquier otro gen que pueda reemplazar a BIRC5 o UBE2C, a condición de que tras un reemplazo se seleccionen 8 genes distintos; y

aunque STC2 puede reemplazarse por INPP4B o IL6ST o SEC14L2 o MAPT o CHPT1 o ABAT o SCUBE2 o ESR1 o RBBP8 o PGR o PTPRT o HSPA2 o PTGER3, a condición de que tras un reemplazo se seleccionen 8 genes distintos; y

aunque AZGP1 puede reemplazarse por PIP o EPHX2 o PLAT o SEC14L2 o SCUBE2 o PGR, a condición de que tras un reemplazo se seleccionen 8 genes distintos; y

aunque RBBP8 puede reemplazarse por CELSR2 o PGR o STC2 o ABAT o IL6ST a condición de que tras un reemplazo se seleccionen 8 genes distintos; y

aunque IL6ST puede reemplazarse por INPP4B o STC2 o MAPT o SCUBE2 o ABAT o PGR o SEC14L2 o ESR1 o GJA1 o MGP o EPHX2 o RBBP8 o PTPRT o PLAT, a condición de que tras un reemplazo se seleccionen 8 genes distintos; y

aunque MGP puede reemplazarse por APOD o IL6ST o EGFR, a condición de que tras un reemplazo se seleccionen 8 genes distintos;

(b) combinar de forma matemática los valores del nivel de expresión para los genes de dicho conjunto, cuyos valores se determinaron en la muestra de tumor para producir una puntuación combinada, en la que dicha puntuación combinada está prediciendo dicha respuesta y/o beneficio de la quimioterapia.

El procedimiento de la invención es particularmente adecuado para predecir una respuesta a la quimioterapia citotóxica que contiene taxano/antraciclina, preferentemente en tumores negativos para Her2/neu, positivos para receptor de estrógenos (luminales), preferentemente en el modo neoadyuvante.

- 5 De acuerdo con un aspecto de la invención, se proporciona un procedimiento como se describe anteriormente, en el que dicho nivel de expresión se determina como un nivel de ARNm. De acuerdo con un aspecto de la divulgación, se proporciona un procedimiento como se describe anteriormente, en el que dicho nivel de expresión se determina como un nivel de expresión de genes.

De acuerdo con un aspecto de la invención, se proporciona un procedimiento como se describe anteriormente, en el que dicho nivel de expresión se determina mediante al menos uno de

- 10 un procedimiento basado en PCR,
un procedimiento basado en micromatriz,
un procedimiento basado en hibridación, y
una estrategia de secuenciación y/o de secuenciación de última generación

- 15 De acuerdo con un aspecto de la invención, se proporciona un procedimiento como se describe anteriormente, en el que dicha determinación de niveles de expresión es en una muestra de tumor fijada con formalina incluida en parafina o en una muestra de tumor congelada recientemente.

De acuerdo con un aspecto de la invención, se proporciona un procedimiento como se describe anteriormente, en el que el nivel de expresión de dicho al menos un gen marcador se determina como un patrón de expresión con respecto a al menos un gen de referencia o a un valor de expresión promedio calculado.

- 20 De acuerdo con un aspecto de la invención, se proporciona un procedimiento como se describe anteriormente, en el que dicha etapa de combinar de forma matemática comprende una etapa de aplicar un algoritmo a valores representativos de un nivel de expresión de un dado gen.

- 25 De acuerdo con un aspecto de la invención, se proporciona un procedimiento como se describe anteriormente, en el que dicho algoritmo es una combinación lineal de dichos valores, representativo de un nivel de expresión de un dado gen.

De acuerdo con un aspecto de la invención, se proporciona un procedimiento como se describe anteriormente, en el que un valor representativo de un nivel de expresión de un dado gen se multiplica por un coeficiente.

- 30 De acuerdo con un aspecto de la invención, se proporciona un procedimiento como se describe anteriormente, en el que se determinan uno, dos o más umbrales para dicha puntuación combinada y se discrimina en riesgo elevado y bajo, riesgo elevado, intermedio y bajo, o más grupos de riesgo, aplicando el umbral de la puntuación combinada.

- 35 De acuerdo con un aspecto de la invención, se proporciona un procedimiento como se describe anteriormente, en el que una puntuación combinada elevada es indicativa del beneficio de una terapia más agresiva, por ejemplo quimioterapia citotóxica. Los expertos en la materia entenderán que una "puntuación elevada" en este aspecto se refiere a un valor de referencia o un valor límite. El experto en la materia entenderá adicionalmente que dependiendo del algoritmo particular utilizado para obtener la puntuación combinada, también puede ser indicativa del beneficio de una terapia más agresiva una puntuación "baja", por debajo de un límite o un valor de referencia, por ejemplo quimioterapia citotóxica.

- 40 De acuerdo con un aspecto de la invención, se proporciona un procedimiento como se describe anteriormente, en el que la información con respecto al estado de los ganglios del paciente se procesa en la etapa de combinar de forma matemática los valores de nivel de expresión para los genes, para producir una puntuación combinada.

- 45 De acuerdo con un aspecto de la invención, se proporciona un procedimiento como se describe anteriormente, en el que dicha información con respecto al estado de los ganglios es un valor numérico ≤ 0 si dicho estado de los ganglios es negativo, y dicha información es un valor numérico > 0 si dicho estado de los ganglios es positivo o desconocido. En realizaciones ejemplares de la invención, a un estado de los ganglios negativo se le asigna el valor 0, a un estado de los ganglios desconocido se le asigna el valor 0,5 y a un estado de los ganglios positivo se le asigna el valor 1. Se pueden elegir otros valores para reflejar un peso distinto del estado de los ganglios dentro de un algoritmo.

- 50 De acuerdo con un aspecto de la invención, se proporciona un procedimiento como se describe anteriormente, en el que dicha información con respecto al tamaño tumoral del paciente se procesa en la etapa de combinar de forma matemática los valores del nivel de expresión para los genes, para producir una puntuación combinada.

De acuerdo con un aspecto de la invención, se proporciona un procedimiento como se describe anteriormente, en el que dicha información con respecto al estado de los ganglios y el tamaño tumoral del paciente se procesa en la etapa de combinar de forma matemática los valores del nivel de expresión para los genes, para producir una puntuación combinada.

La invención se refiere adicionalmente al uso de un kit para realizar el procedimiento como se describe anteriormente, comprendiendo dicho kit un conjunto de oligonucleótidos que tiene una capacidad de unirse de forma específica a secuencias, o para secuencias de fragmentos de los genes en una combinación de genes, en los que

(i) dicha combinación comprende al menos los 8 genes UBE2C, BIRCS, DHCR7, STC2, AZGP1, RBBP8, IL6ST y MGP; o

(ii) dicha combinación comprende al menos los 8 genes UBE2C, RACGAP1, DHCR7, STC2, AZGP1, RBBP8, IL6ST y MGP.

Adicionalmente, se desvela un producto de programa de ordenador que tiene la capacidad de procesar valores representativos de un nivel de expresión de una combinación de genes, combinando de forma matemática dichos valores para producir una puntuación combinada, en el que dicha puntuación combinada está prediciendo dicha respuesta y/o el beneficio de la quimioterapia en dicho paciente.

Dicho producto de programa de ordenador puede almacenarse en un soporte de datos o implementarse en un sistema diagnóstico que tenga la capacidad de producir valores representativos de un nivel de expresión de un dado gen, tal como un sistema de PCR en tiempo real.

Si el producto de programa de ordenador está almacenado en un soporte de datos o se ejecuta en un ordenador, los operarios pueden introducir los valores de expresión obtenidos para el nivel de expresión de los genes respectivos. Después, el producto del programa de ordenador puede aplicar un algoritmo para producir una puntuación combinada indicativa del beneficio de la quimioterapia citotóxica para un dado paciente.

Los procedimientos de la presente invención tienen la ventaja de proporcionar una predicción fiable de la respuesta y/o el beneficio de la quimioterapia a base del uso solo de un pequeño número de genes. Se ha encontrado que los procedimientos de la presente invención son especialmente adecuados para analizar la respuesta y/o el beneficio de la quimioterapia de pacientes con tumores clasificados como positivos para ESR1 y negativos para ERBB2.

Descripción detallada de la invención

En las subreivindicaciones y en la siguiente descripción de las figuras y ejemplos respectivos, se desvelan detalles, aspectos, características y ventajas adicionales del objeto de la invención, los cuales, de un modo ejemplar, muestran realizaciones preferentes de la presente invención.

A partir del depósito de datos combinado de expresión de genes (GEO) se recuperaron cuatro conjuntos de datos de expresión de genes disponibles para el público (Affymetrix HG-U133A). Todos los pacientes de cáncer de mama analizados se trataron con quimioterapia neoadyuvante basada en antraciclina o taxano/antraciclina. Los archivos de celdas de micromatriz se normalizaron por MAS5, con un procedimiento de escalado global y una intensidad diana de 500. Para la evaluación de la respuesta al tratamiento se utilizó como criterio de valoración principal la respuesta patológica completa (RpC). El análisis se realizó en todos los pacientes de cáncer de mama negativos para HER2/neu y en el subconjunto de pacientes de cáncer de mama positivos para ER, negativos para HER2, de acuerdo con niveles límite pre especificados (conjunto de sondas ERBB2 216836 < 6000 = negativo para HER2/neu, conjunto de sondas ERBB2 216836 < 6000 y conjunto de sondas ESR1 > 1000 = positivo para ER/negativo para HER2/neu).

La puntuación T5 se examinó en 374 pacientes de cáncer de mama negativos para HER2 tratados con terapia neoadyuvante (figura 1). Entre los 374 pacientes, se clasificaron 63 tumores (el 16,8 %) como riesgo bajo de T5, mientras 311 tumores (el 83,2 %) fueron de riesgo bajo de TS. Solo uno de los tumores de riesgo bajo por T5 alcanzó una RpC tras la terapia neoadyuvante, mientras que 84 de los 85 acontecimientos de RpC se clasificaron como de riesgo alto por T5. La sensibilidad de la puntuación T5 fue del 99 % y el valor predictivo negativo del 98 %, con un área bajo la curva operativa característica del receptor de 0,69 (figura 1).

La Figura 1 muestra:

a) La distribución de la puntuación T5 en 374 pacientes de cáncer de mama negativos para HER2/neu (85 acontecimientos RpC frente a 289 muestras con enfermedad residual); prueba de Mann-Whitney de dos colas

b) Utilizando la puntuación de T5 5 como límite pre-especificado, la sensibilidad fue del 99 %, la especificidad de 21 %, el valor predictivo negativo del 98 % y el valor predictivo positivo del 27 %, área bajo la curva operativa característica del receptor de 0,69.

La puntuación T5 se examinó en 221 pacientes de cáncer de mama positivos para ER, negativos para HER2, tratados con terapia neoadyuvante (figura 2). Entre los 221 pacientes, 61 tumores (el 27,6 %) se clasificaron como de riesgo bajo por T5, mientras que 160 tumores (el 72,4 %) fueron de riesgo elevado por T5. Solo uno de los tumores con riesgo bajo por T5 alcanzó una RpC tras la terapia neoadyuvante, mientras que 24 de los 25 acontecimientos de RpC se clasificaron como de riesgo elevado por T5. La sensibilidad de la puntuación T5 fue del 96 % y el valor predictivo negativo del 98 %, con área bajo la curva operativa característica del receptor de 0,73 (figura 2).

La Figura 2 muestra:

c) La distribución de la puntuación T5 en 221 pacientes de cáncer de mama positivo para receptor de estrógenos y negativo para HER2/neu (25 acontecimientos de RpC frente a 196 muestras con enfermedad residual); prueba de Mann-Whitney de dos colas

- 5 d) Utilizando la puntuación TS 5 como límite pre-especificado, la sensibilidad fue del 96 %, la especificidad del 30 %, el valor predictivo negativo del 98 % y el valor predictivo positivo del 15 %, con área bajo la curva operativa característica del receptor de 0,73.

10 En el presente documento se desvelan combinaciones exclusivas de genes de marcadores que pueden combinarse en algoritmos para la nueva prueba predictiva presentada en el presente documento. Técnicamente, el procedimiento de la invención puede practicarse utilizando dos tecnologías: 1.) Aislamiento de ARN total a partir de tejido de tumor recién obtenido o fijado y 2.) RT-PCR cuantitativa de los ácidos nucleicos aislado. Como alternativa, se contempla medir los niveles de expresión utilizando tecnologías alternativas, por ejemplo mediante micromatriz, en particular matrices U-133 de affymetrix o mediante la medición al nivel de proteínas.

15 Los procedimientos de la invención son a base de la determinación cuantitativa de las especies de ARN aisladas a partir del tumor, para obtener los valores de expresión y el posterior análisis bioinformático de dichos valores de expresión determinados. Las especies de ARN se pueden aislar a partir de cualquier tipo de muestra de tumor, por ejemplo de muestras de biopsias, muestras de frotis, material tumoral extirpado, tejido de tumor reciente congelado o a partir de tejido de tumor incluido en parafina y fijado con formalina. En primer lugar, se determinan los niveles de ARN de genes que codifican combinaciones específicas de los genes UBE2C, BIRC5, DHCR7, RACGAP1, AURKA, PVALB, NMU, STC2, AZGP1, RBBP8, IL6ST, MGP, PTGER3, CXCL12, ABAT, CDH1 y PIP, o combinaciones específicas de los mismos, según se indica. A base de estos valores de expresión se calcula mediante una combinación matemática una puntuación predictiva, por ejemplo, de acuerdo con las fórmulas T5, T1, T4 o T5b (véase a continuación).

25 Un valor de la puntuación elevado indica una probabilidad aumentada de una respuesta patológica completa tras el tratamiento con la quimioterapia neoadyuvante, un valor de la puntuación bajo indica una probabilidad reducida de desarrollar una respuesta patológica completa tras el tratamiento neoadyuvante. En consecuencia, una puntuación elevada también indica que el paciente es un paciente de riesgo elevado que se beneficiará de una terapia más agresiva, por ejemplo de la quimioterapia citotóxica.

A continuación la Tabla 1 muestra las combinaciones de genes utilizados para cada algoritmo.

30 Tabla 1: Combinación de genes para los respectivos algoritmos:

Gen	Algo_T1	Algo_T4	Alg_T5	Algo_T5b
UBE2C			X	X
BIRC5	X	X	X	
DHCR7		X	X	X
RACGAP1		X		X
AURKA	X			
PVALB	X	X		
NMU	X			X
STC2	X	X	X	
AZGP1			X	X
RBBP8	X		X	X
IL6ST		X	X	X
MGP			X	X
PTGER3	X	X		
CXCL12	X	X		
ABAT		X		
CDH1	X			
PIP	X			

A continuación la Tabla 2 muestra el ID del conjunto de sondas Affy y el mapeo del ID de diseño de TaqMan de los genes marcadores de la presente invención.

Tabla 2: Símbolo del gen, ID del conjunto de sondas Affy y el mapeo del ID del diseño de TaqMan:

Gen	ID de diseño	ID del conjunto de sondas
UBE2C	R65	202954_at
BIRC5	SC089	202095_s_at
DHCR7	CAGMC3344	201791_s_at
RACGAP1	R125-2	222077_s_at
AURKA	CAGMC336	204092_s_at
PVALB	CAGMC339	205336_at
NMU	CAGMC331	206023_at
STC2	R52	203438_at
AZGP1	CAGMC372	209309_at
RBBP8	CAGMC347	203344_s_at
IL6ST	CAGMC312	212196_at
MGP	CAGMC383	202291_s_at
PTGER3	CAGMC315	213933_at
CXCL12	CAGMC342	209687_at
ABAT	CAGMC338	209460_at
CDH1	CAGMC335	201131_s_at

A continuación la Tabla 3 muestra los nombres completos, las ID de Entrez Gene, los números de registro del gene bank y el emplazamiento cromosómico de los genes marcadores de la presente invención

Símbolo oficial	Nombre completo oficial	ID de Entrez Gene	Número de referencia	Emplazamiento
UBE2C	enzima E2C de conjugación de ubiquitina	11065	U73379	20q13. 12
BIRC5	PIA baculovírica que contiene repeticiones 5	332	U75285	17q25
DHCR7	7-dihidrocolesterol reductasa	1717	AF034544	11q13. 4
STC2	estaniocalcina 2	8614	AB012664	5q35.2
RBBP8	proteína de unión de retinoblastoma 8	5932	AF043431	18q11. 2
IL6ST	transductor de señal de la interleucina 6	3572	M57230	5q11
MGP	proteína de la matriz Gla	4256	M58549	12p12. 3
AZGP1	alfa-2-glucoproteína 1, unión a cinc	563	BC005306	11q22. 1
RACGAP1	proteína de activación de la Rac GTPasa 1	29127	NM_013277	12q13
AURKA	aurora quinasa A	6790	BC001280	20q13
PVALB	parvalbúmina	5816	NM_002854	22q13. 1
NMU	neuromedina U	10874	X76029	4q12
PTGER3	receptor de la prostaglandina E 3 (subtipo EP3)	5733	X83863	1p31.2
CXCL12	ligando de la quimiocina (motivo C-X-C) 12 (factor derivado de células estromales 1)	6387	L36033	10q11.1
ABAT	4-aminobutirato aminotransferasa	18	L32961	16p13. 2
CDH1	cadherina 1, tipo 1, E-cadherina (epitelial)	999	L08599	16q22. 1
PIP	proteína inducida por prolactina	5304	NMM_002652	7q32-qter

5

Algoritmo de ejemplo T5:

El algoritmo T5 es una agrupación de cuatro miembros en donde cada miembro es una combinación lineal de dos genes. Las fórmulas matemáticas para T5 se muestran a continuación; la notación es la misma que para T1. T5 se puede calcular solo a partir de los datos de expresión de genes.

10 Miembro de riesgo 1 = $0,434039 [0,301..0,567] * (0,939 * BIRC5 -3,831)$
 $-0,491845 [-0,714..-0,270] * (0,707 * RBBP8 -0,934)$
Miembro de riesgo 2 = $0,488785 [0,302..0,675] * (0,794 * UBE2C -1,416)$
 $-0,374702 [-0,570..-0,179] * (0,814 * IL6ST -5,034)$

Miembro de riesgo 3 = $0,39169 [0,541..-0,242] * (0,674 * AZGP1 -0,777)$
 $+0,44229 [0,256..0,628] * (0,891 * DHCR7 -4,378)$

Miembro de riesgo 4 = $0,377752 [-0,543..-0,212] * (0,485 * MGP +4,330)$
 $-0,177669 [-0,267..-0,088] * (0,826 * STC2 -3,630)$

5 riesgo = Miembro de riesgo 1 + Miembro de riesgo 2 + Miembro de riesgo 3 + Miembro de riesgo 4

Los coeficientes a la izquierda de cada línea se calcularon como los coeficientes de regresión de los riesgos proporcionales de COX, los números entre corchetes indican los límites del intervalo de confianza del 95 % para estos coeficientes. En otras palabras, en lugar de multiplicar el término $(0,939 * BIRC5 -3,831)$ por 0,434039, se puede multiplicar por cualquier coeficiente entre 0,301 y 0,567, y todavía proporcionar un resultado predictivo dentro de los límites del intervalo de confianza del 95 %. Los términos entre paréntesis a la derecha de cada línea indican una transferencia de plataforma de la PCR a Affymetrix: Las variables PVALB, CDH1, ... indican las expresiones basadas en PCR normalizadas por los genes de referencia (valores delta Ct), el término entero dentro de los paréntesis corresponde al logaritmo (en base 2) de los valores de expresión de micromatriz de Affymetrix de los conjuntos de sonda correspondientes.

15 Algoritmo de ejemplo TSclin:

El algoritmo TSclin es una puntuación combinada que consiste en la puntuación T5 y los parámetros clínicos (estados de los ganglios y tamaño tumoral).

$$T5clin = 0,35 * t + 0,64 * n + 0,28 * s$$

20 en donde t codifica el tamaño tumoral (1: ≤ 1 cm, 2: >1 cm a ≤ 2 cm, 3: >2 cm a ≤ 5 cm, 4: > 5 cm), y n el estado de los ganglios (1: negativo, 2: 1 a 3 ganglios positivos, 3: 4 hasta 4 ganglios positivos, 4: >10 ganglios positivos).

En una realización preferente, el umbral para la puntuación T5clin es 3,3.

Algoritmo de ejemplo T1:

El algoritmo T1 es una agrupación de tres miembros en donde cada miembro es una combinación lineal de hasta cuatro variables. En las variables generales pueden estar las expresiones de genes o las variables clínicas. En T1 la única variable que no es génica es el estado de los ganglios codificado como 0, si el paciente es negativo para ganglios linfáticos y como 1, si el paciente es positivo para ganglios linfáticos. Las fórmulas matemáticas para T1 se muestran a continuación.

Miembro de riesgo 1 = $+0,193935 [0,108..0,280] * (0,792 * PVALB -2,189)$
 $-0,240252 [-0,400..-0,080] * (0,859 * CDH1 -2,900)$
 $-0,270069 [-0,385..-0,155] * (0,821 * STC2 - 3,529)$
 $+1,2053 [0,534..1,877] * \text{estado de ganglios}$

Miembro de riesgo 2 = $-0,25051 [-0,437..-0,064] * (0,558 * CXCL12 +0,324)$
 $-0,421992 [-0,687..-0,157] * (0,715 * RBBP8 -1,063)$
 $+0,148497 [0,029..0,268] * (1,823 * NMU -12,563)$
 $+0,293563 [0,108..0,479] * (0,989 * BIRC5 -4,536)$

Miembro de riesgo 3 = $+0,308391 [0,074..0,543] * (0,812 * AURKA -2,656)$
 $-0,225358 [-0,395..-0,055] * (0,637 * PTGER3 +0,492)$
 $-0,116312 [-0,202..-0,031] * (0,724 * PIP +0,985)$

riesgo = + Miembro de riesgo 1 + Miembro de riesgo 2 + Miembro de riesgo 3

40 Los coeficientes a la izquierda de cada línea se calcularon como los coeficientes de regresión de los riesgos proporcionales de COX, los números entre corchetes indican los límites del intervalo de confianza del 95 % para estos coeficientes. Los términos entre paréntesis a la derecha de cada línea indican una transferencia de plataforma de la PCR a Affymetrix: Las variables PVALB, CDH1, ... indican las expresiones basadas en PCR normalizadas mediante los genes de referencia, el término entero dentro de los paréntesis corresponde al logaritmo (en base 2) de los valores de expresión de micromatriz de Affymetrix de los conjuntos de sonda correspondientes.

Algoritmo de ejemplo T4:

El algoritmo T4 es una combinación lineal de motivos. Los primeros 10 genes de varios análisis de los conjuntos de datos de Affymetrix y de los datos de PCR se agruparon en motivos. Los genes que no pertenecen a un grupo se utilizaron como motivos de genes individuales. En un análisis de multivarianza se encontraron los coeficientes de regresión de riesgos proporcionales de COX.

En los motivos generales pueden estar las expresiones de genes individuales o las expresiones de genes medias de genes correlacionados. Las fórmulas matemáticas para T4 se muestran a continuación.

$$\text{prolif} = ((0,84 [0,697..0,977] * \text{RACGAP1} -2,174) + (0,85 [0,713..0,988] * \text{DHCR7} -3,808) + (0,94 [0,786..1,089] * \text{BIRC5} -3,734)) / 3$$

$$\text{motiv2} = ((0,83 [0,693..0,96] * \text{IL6ST} -5,295) + (1,11 [0,930..1,288] * \text{ABAT} -7,019) + (0,84 [0,701..0,972] * \text{STC2} -3,857)) / 3$$

$$\text{ptger3} = (\text{PTGER3} * 0,57 [0,457..0,659] + 1,436)$$

$$\text{cxcl12} = (\text{CXCL12} * 0,53 [0,446..0,618] + 0,847)$$

$$\text{pvalb} = (\text{PVALB} * 0,67 [0,558..0,774] -0,466)$$

- 10 Los factores y desfases para cada gen indican una transferencia de plataforma de la PCR a Affymetrix: Las variables RACGAP1, DHCR7, ... indican expresiones basadas en PCR normalizadas por CALM2 y PPIA, el término entero dentro de los paréntesis corresponde al logaritmo (en base 2) de los valores de expresión de micromatriz de Affymetrix de los conjuntos de sondas correspondientes.

Los números entre corchetes indican los límites del intervalo de confianza del 95 % para estos factores.

- 15 Como el algoritmo se realiza incluso mejor en combinación con una variable clínica, se añadió el estado de ganglios. En T4 el estado de ganglios se codifica como 0, si el paciente es negativo para ganglios linfáticos y como 1, si el paciente es positivo para ganglios linfáticos. Con esto, el algoritmo T4 es:

$$\begin{aligned} \text{riesgo} = & -0,32 [-0,510..-0,137] * \text{motiv2} \\ & + 0,65 [0,411..0,886] * \text{prolif} \\ & - 0,24 [-0,398..-0,08] * \text{ptger3} \\ & - 0,05 [-0,225..0,131] * \text{cxcl12} \\ & + 0,09 [0,019..0,154] * \text{pvalb} \\ & + \text{estado de ganglios} \end{aligned}$$

- 25 Los coeficientes de riesgo se calcularon como los coeficientes de regresión de los riesgos proporcionales de COX, los números entre corchetes indican los límites del intervalo de confianza del 95 % para estos coeficientes.

El algoritmo T5b es una agrupación de dos miembros en donde cada miembro es una combinación lineal de cuatro genes. Las fórmulas matemáticas para T5b se muestran a continuación, la notación es la misma que para T1 y T5. En T5b una variable no génica es el estado de ganglios codificado como 0, si el paciente es negativo para ganglios linfáticos y como 1, si el paciente es positivo para ganglios linfáticos y como 0,5 si el estado de ganglios linfáticos es desconocido. T5b se define por:

$$\begin{aligned} \text{Miembro de riesgo 1} = & 0,359536 [0,153..0,566] * (0,891 * \text{DHCR7} -4,378) \\ & -0,288119 [-0,463..-0,113] * (0,485 * \text{MGP} + 4,330) \\ & +0,257341 [0,112..0,403] * (1,118 * \text{NMU} -5,128) \\ & -0,337663 [-0,499..-0,176] * (0,674 * \text{AZGP1} -0,777) \end{aligned}$$

$$\begin{aligned} \text{Miembro de riesgo 2} = & 0,374940 [-0,611..-0,139] * (0,707 * \text{RBBP8} -0,934) \\ & 0,387371 [-0,597..-0,178] * (0,814 * \text{IL6ST} -5,034) \\ & +0,800745 [0,551..1,051] * (0,860 * \text{RACGAP1} -2,518) \\ & +0,770650 [0,323..1,219] * \text{estado de ganglios} \end{aligned}$$

$$\text{riesgo} = \text{Miembro de riesgo 1} + \text{Miembro de riesgo 2}$$

El experto entenderá que estos logaritmos representan ejemplos particulares y que se basan en la información con respecto a la asociación de la expresión de genes con la predicción de la respuesta terapéutica.

Simplificación del algoritmo empleando subconjuntos de genes

“Algoritmo de ejemplo T5” es una agrupación predictiva que consiste en 4 miembros con 2 genes de interés cada uno. Cada miembro es un predictivo independiente y autocontenido de la recidiva a distancia y/o de la respuesta a la terapia, cada miembro adicional contribuye a la solidez y al poder predictivo del algoritmo. La ecuación a

continuación muestra el “Algoritmo de Ejemplo T5”; para facilitar la lectura, el número de dígitos tras la coma en los números decimales se ha limitado hasta 2; el intervalo en corchetes enumera el intervalo estimado de los coeficientes (media +/- 3 desviaciones típicas).

Algoritmo T5:

$$\begin{aligned}
 &+0,41 [0,21..0,61] * BIRC5 -0,33 [-0,57..-0,09] * RBBP8 \\
 &+0,38 [0,15..0,61] * UBE2C -0,30 [-0,55..-0,06] * IL6ST \\
 &-0,28 [-0,43..-0,12] * AZGP1 +0,42 [0,16..0,68] * DHCR7 \\
 &-0,18 [-0,31..-0,06] * MGP -0,13 [-0,25..-0,02] * STC2
 \end{aligned}$$

índices c: conjunto de entrenamiento= 0,724,

10 Los nombres de los genes en el algoritmo indican la diferencia de la expresión de ARNm del gen en comparación con uno o más genes constitutivos, como se describe anteriormente.

Al analizar una cohorte distinta de la cohorte de búsqueda (234 muestras de tumor), fue sorprendente darse cuenta de que algunas simplificaciones del “algoritmo T5 original” todavía producían un desempeño diagnóstico no significativamente inferior al algoritmo TS original. La simplificación más directa fue reducir la agrupación predictiva a solo un miembro. Los ejemplos de desempeño de las “agrupaciones de un miembro” se muestran a continuación:

solo miembro 1:

$$+0,41 [0,21..0,61] * BIRC5 -0,33 [-0,57..-0,09] * RBBP8$$

índices c: conjunto de entrenamiento = 0,653, cohorte independiente = 0,681
solo miembro 2:

$$+0,38 [0,15..0,61] * UBE2C -0,30 [-0,55..-0,06] * IL6ST$$

índices c: conjunto de entrenamiento = 0,664, cohorte independiente = 0,696
solo miembro 3:

$$-0,28 [-0,43..-0,12] * AZGP1 +0,42 [0,16..0,68] * DHCR7$$

índices c: conjunto de entrenamiento = 0,666, cohorte independiente = 0,601
solo miembro 4:

$$-0,18 [-0,31..-0,06] * MGP -0,13 [-0,25..-0,02] * STC2$$

índices c: conjunto de entrenamiento = 0,668, cohorte independiente = 0,593

El desempeño de las agrupaciones de un miembro está notablemente reducido en comparación con el desempeño del algoritmo completo, como se muestra en una cohorte independiente de 234 muestras.

30 La combinación de forma gradual de más de uno pero menos de cuatro miembros para un nuevo algoritmo predictivo de agrupación pronóstico, con frecuencia conduce a un pequeño pero significativo aumento del desempeño diagnóstico en comparación con la agrupación de un miembro. Fue sorprendente darse cuenta de que hubo marcadas mejoras mediante algunas combinaciones de miembros de la agrupación, mientras que otras combinaciones tuvieron una producción cercana a la no mejora. De forma inicial, la hipótesis fue que una combinación de miembros que representa similares motivos biológicos, reflejado en que los genes empleados produjeron una mejora más pequeña que la combinación de miembros, lo que refleja claramente distintos motivos biológicos. No obstante, este no fue el caso. No se pudo identificar una regla para pronosticar una combinación de algunos genes para generar un algoritmo que presente más capacidad pronóstica que otra combinación de genes. Las combinaciones prometedoras solo pudieron seleccionarse a base de los datos experimentales. Las combinaciones identificadas de miembros de agrupación combinados para producir algoritmos simplificados pero altamente eficaces se muestran a continuación.

solo miembros 1 y 2:

$$\begin{aligned}
 &+0,41 [0,21..0,61] * BIRC5 -0,33 [-0,57..-0,09] * RBBP8 \\
 &+0,38 [0,15..0,61] * UBE2C -0,30 [-0,55..-0,06] * IL6ST
 \end{aligned}$$

45 índices c: conjunto de entrenamiento = 0,675, cohorte independiente = 0,712
solo miembros 1 y 3:

$$+0,41 [0,21..0,61] * BIRC5 -0,33 [-0,57..-0,09] * RBBP8$$

$$-0,28 [-0,43..-0,12] * AZGP1 +0,42 [0,16..0,68] * DHCR7$$

índices c: conjunto de entrenamiento n = 0,697, cohorte independiente = 0,688
solo miembros 1 y 4:

5

$$+0,41 [0,21..0,61] * BIRC5 -0,33 [-0,57..-0,09] * RBBP8$$

$$-0,18 [-0,31..-0,06] * MGP -0,13 [-0,25..-0,02] * STC2$$

índices c: conjunto de entrenamiento = 0,705, cohorte independiente = 0,697
solo miembros 2 y 3:

$$+0,38 [0,15..0,61] * UBE2C -0,30 [-0,55..-0,06] * IL6ST$$

10

$$-0,28 [-0,43..-0,12] * AZGP1 +0,42 [0,16..0,68] * DHCR7$$

índices c: conjunto de entrenamiento = 0,698, cohorte independiente = 0,670
solo miembros 1, 2 y 3:

$$+0,41 [0,21..0,61] * BIRC5 -0,33 [-0,57..-0,09] * RBBP8$$

15

$$+0,38 [0,15..0,61] * UBE2C -0,30 [-0,55..-0,06] * IL6ST$$

$$-0,28 [-0,43..-0,12] * AZGP1 +0,42 [0,16..0,68] * DHCR7$$

índices c: conjunto de entrenamiento = 0,701, cohorte independiente = 0,715

20 La no omisión de los miembros al completo de la agrupación, pero sí de un único gen o genes de distintos miembros de la misma, es posible pero requiere un reentrenamiento del algoritmo entero. Aun así, puede ser ventajoso realizarlo también. El desempeño de algoritmos simplificados generados omitiendo todos los miembros o genes individuales es en gran medida idéntico.

Variantes del algoritmo por reemplazo de genes

25 Los algoritmos descritos, tales como el “algoritmo de ejemplo T5” anterior, también puede modificarse reemplazando uno o más genes por uno o más de otros genes. El fin de tales modificaciones es reemplazar genes difíciles de medir en una plataforma específica por un gen que se ensaya de forma más directa en esta plataforma. Mientras que tal transferencia puede no necesariamente producir un desempeño mejorado, en comparación con un algoritmo inicial, puede generar el indicio para implantar el algoritmo pronóstico en una plataforma de diagnóstico particular. En general, el reemplazo de un gen por otro gen, aunque conserva la capacidad diagnóstica del algoritmo predictivo, puede realizarse mejor reemplazando un gen por un gen coexpresado con una elevada correlación (mostrado por ejemplo mediante el coeficiente de correlación de Pearson). Aun así, debe tenerse en cuenta que la expresión de

30 ARNm de dos genes elevadamente correlativa en una plataforma puede parecer bastante independiente una de la otra cuando se evalúa en otra plataforma. Por consiguiente, cuando se reduce a la práctica experimental, tal reemplazo aparentemente fácil puede producir resultados deficientes decepcionantes, así como sorprendentes buenos resultados, siempre dependiendo de la imponderabilidad de la plataforma empleada. Repitiendo este

35 procedimiento se pueden reemplazar varios genes.

La eficacia de tal estrategia se puede demostrar evaluando el desempeño predictivo de la puntuación del algoritmo T5 y de sus variantes en las cohortes de validación. La siguiente tabla muestra el índice c con respecto al criterio de valoración de recidiva a distancia en dos cohortes de validación.

Variante	Estudio de validación A	Estudio validación B
algoritmo T5 original	índice c = 0,718	índice c = 0,686
omisión de BIRC5 (ajustando la expresión a alguna constante)	índice c = 0,672	índice c = 0,643
reemplazando BIRC5 por UBE2C (sin ajuste del coeficiente)	índice c = 0,707	índice c = 0,678

40 Se puede observar que la omisión de uno de los genes de T5, mostrado aquí por ejemplo para BIRC5, reduce de forma notable el desempeño predictivo. El reemplazo por otro gen produce aproximadamente el mismo desempeño.

Un procedimiento mejor de reemplazo de un gen es reentrenar el algoritmo. Dado que T5 consiste en cuatro miembros de la agrupación independientes, se debe reentrenar solo el miembro que contiene el gen reemplazo. Las

siguientes ecuaciones demuestran los reemplazos de genes del algoritmo T5 mostrado anteriormente, entrenados en una cohorte de 234 pacientes de cáncer de mama. A continuación se muestra solo un miembro, para el cálculo del índice c se utilizaron los miembros restantes sin cambios, a partir del algoritmo T5 original. El intervalo entre corchetes enumera el intervalo estimado de coeficientes: media +/- 3 desviaciones típicas.

5 Miembro 1 de T5:

Miembro 1 original:

$$+0,41 [0,21..0,61] * BIRC5 -0,33 [-0,57..-0,09] * RBBP8$$

índices c = conjunto de entrenamiento == 0,724, cohorte independiente = 0,705
reemplazo de BIRC5 por TOP2A en el miembro 1:

10 $+0,47 [0,24..0,69] * TOP2A -0,34 [-0,58..-0,10] * RBBP8$

índices c = conjunto de entrenamiento = 0,734, cohorte independiente = 0,694
reemplazo de BIRC5 por RACGPA1 en el miembro 1:

$$+0,69 [0,37..1,00] * RACGAP1 -0,33 [-0,57..-0,09] * RBBP8$$

15 índices c = conjunto de entrenamiento = 0,736, cohorte independiente = 0,743
reemplazo de RBBP8 por CELSR2 en el miembro 1:

$$+0,38 [0,19..0,57] * BIRC5 -0,18 [-0,41..0,05] * CELSR2$$

índices c = conjunto de entrenamiento = 0,726, cohorte independiente = 0,680
reemplazo de RBBP8 por PGR en el miembro 1:

$$+0,35 [0,15..0,54] * BIRC5 -0,09 [-0,23..0,05] * PGR$$

20 índices c = conjunto de entrenamiento = 0,727, cohorte independiente = 0,731

Miembro 2 de T5:

Miembro 2 original:

$$+0,38 [0,15..0,61] * UBE2C -0,30 [-0,55..-0,06] * IL6ST$$

25 índices c = conjunto de entrenamiento = 0,724, cohorte independiente = 0,725
reemplazo de UBE2C por RACGAP1 en el miembro 2:

$$+0,65 [0,33..0,96] * RACGAP1 -0,38 [-0,62..-0,13] * IL6ST$$

índices c = conjunto de entrenamiento = 0,735, cohorte independiente = 0,718
reemplazo de UBE2C por TOP2A en el miembro 2:

$$+0,42 [0,20..0,65] * TOP2A -0,38 [-0,62..-0,13] * IL6ST$$

30 índices c = conjunto de entrenamiento = 0,734, cohorte independiente = 0,700
reemplazo de IL6ST por INPP4B en el miembro 2:

$$+0,40 [0,17..0,62] * UBE2C -0,25 [-0,55..0,05] * INPP4B$$

índices c = conjunto de entrenamiento = 0,725, cohorte independiente = 0,686
reemplazo de IL6ST por MAPT en el miembro 2:

35 $+0,45 [0,22..0,69] * UBE2C -0,14 [-0,28..0,01] * MAPT$

índices c = conjunto de entrenamiento = 0,727, cohorte independiente = 0,711

Miembro 3 de T5:

Miembro 3 original:

$$-0,28 [-0,43..-0,12] * AZGP1 +0,42 [0,16..0,68] * DHCR7$$

40 índices c = conjunto de entrenamiento = 0,724, cohorte independiente = 0,705
reemplazo de AZGP1 por PIP en el miembro 3:

$$-0,10 [-0,18..-0,02] * PIP +0,43 [0,16..0,70] * DHCR7$$

índices c = conjunto de entrenamiento = 0,725, cohorte independiente = 0,692
reemplazo de AZGP1 por EPHX2 en el miembro 3:

$$-0,23 [-0,43..-0,02] * EPHX2 +0,37 [0,10..0,64] * DHCR7$$

5 índices c = conjunto de entrenamiento = 0,719, cohorte independiente = 0,698
reemplazo de AZGP1 por PLAT en el miembro 3:

$$-0,23 [-0,40..-0,06] * PLAT +0,43 [0,18..0,68] * DHCR7$$

índices c = conjunto de entrenamiento = 0,712, cohorte independiente = 0,715
reemplazo de DHCR7 por AURKA en el miembro 3:

$$-0,23 [-0,39..-0,06] * AZGP1 +0,34 [0,10..0,58] * AURKA$$

10 índices c = conjunto de entrenamiento = 0,716, cohorte independiente = 0,733

Miembro 4 de T5:

Miembro 4 original:

$$-0,18 [-0,31..-0,06] * MGP -0,13 [-0,25..-0,02] * STC2$$

15 índices c = conjunto de entrenamiento = 0,724, cohorte independiente = 0,705
reemplazo de MGP por APOD en el miembro 4:

$$-0,16 [-0,30..-0,03] * APOD -0,14 [-0,26..-0,03] * STC2$$

índices c = conjunto de entrenamiento = 0,717, cohorte independiente = 0,679
reemplazo de MGP por EGFR en el miembro 4:

$$-0,21 [-0,37..-0,05] * EGFR -0,14 [-0,26..-0,03] * STC2$$

20 índices c = conjunto de entrenamiento = 0,715, cohorte independiente = 0,708
reemplazo de STC2 por INPP4B en el miembro 4:

$$-0,18 [-0,30..-0,05] * MGP -0,22 [-0,53..0,08] * INPP4B$$

índices c = conjunto de entrenamiento = 0,719, cohorte independiente = 0,693
reemplazo de STC2 por SEC14L2 en el miembro 4:

25 $-0,18 [-0,31..-0,06] * MGP -0,27 [-0,49..-0,06] * SEC14L2$

índices c = conjunto de entrenamiento = 0,718, cohorte independiente = 0,681

Se puede observar que los reemplazos de genes individuales identificados de forma experimental por una cuantificación con PCR cuantitativa normalmente afecta el desempeño predictivo del algoritmo T5, evaluado mediante el índice c solo de forma no significativa.

30 La siguiente tabla muestra los candidatos potenciales de genes de reemplazo para los genes del algoritmo T5. Cada gen candidato se muestra en una celda de la tabla: el nombre del gen está seguido entre paréntesis del coeficiente de correlación de Pearson absoluto de la expresión del gen original en el algoritmo TS y el candidato de reemplazo, y la ID del conjunto de sondas HG-U133A.

BIRCS	RBBP8	UBE2C	IL6ST	AZGP1	DHCR7	MGP	STC2
UBE2C (0,775), 202954_at	CELSR2 (0,548), 204029_at	BIRCS (0,775), 202095_s_at	INPP4B (0,477), 205376_at	PIP (0,530), 206509_at	AURKA (0,345), 204092_s_at	APOD (0,368), 201525_at	INPP4B (0,500), 205376_at
TOP2A (0,757), 201292_at	PGR (0,392), 208305_at	RACGAP1 (0,756), 204029_at	STC2 (0,450), 203438_at	EPHX2 (0,369), 209368_at	BIRCS (0,323), 202095_s_at	IL6ST (0,327), 212196_at	IL6ST (0,450), 212196_at
RACGAP1 (0,704),	STC2 (0,361), 203438_at	TOP2A (0,753), 201292_at	MAPT (0,440), 206401_s_at	PLAT (0,366), 201860_s_at	UBE2C(0,315), 202954_at	EGFR (0,308), 201983_s_at	SEC14L2 (0,417), 204541_at
AURKA (0,681), 204092_s_at	ABAT (0,317), 209459_s_at	AURKA (0,694), 204092_s_at	SCUBE2 (0,418), 219197_s_at	SEC14L2 (0,351), 204541_at			MAPT (0,414), 206401_s_at
NEK2 (0,680), 204026_s_at	IL6ST (0,311), 212196_at	NEK2 (0,684), 204026_s_at	ABAT (0,389), 209459_s_at	SCUBE2 (0,331), 219197_s_at			CHPT1 (0,410), 221675_s_at
E2F8 (0,640), 219990_at		E2F8 (0,652), 219990_at	PGR (0,377), 208305_at	PGR (0,302), 208305_at			ABAT (0,409), 209459_s_at
201202_at		PCNA (0,589), 201202_at	SEC14L2 (0,356), 204541_at				SCUBE2 (0,406), 219197_s_at
CYBRD1 (0,462), 217889_s_at		CYBRD1 (0,486), 217889_s_at	ESR1 (0,353), 205225_at				ESR1 (0,394), 205225_at
DCN (0,439), 209335_at		AORA2A (0,391), 209869_at	GJA1 (0,335), 201667_at				RBBP8 (0,361), 203344_s_at
ADRA2A (0,416), 209869_at		DCN (0,384), 209335_at	MGP (0,327), 202291_s_at				PGR (0,347), 208305_at
SQLE (0,415), 209218_at		SQLE (0,369), 209218_at	EPHX2 (0,313), 209368_at				PTPRT (0,343), 205948_at
CXCL12 (0,388), 209687_at		CCND1 (0,347), 208712_at	R88P8 (0,311), 203344_s_at				HSPA2 (0,317), 211538_s_at
EPHX2 (0,362), 209368_at		ASPH (0,344), 210896_s_at	PTPRT (0,303), 205948_at				PTGER3 (0,314), 210832_x_at
ASPH (0,352), 210896_s_at		CXCL12 (0,342), 209687_at	PLAT (0,301), 201860_s_at				
PRSS16 (0,352), 208165_s_at		PIP (0,328), 206509_at					

BIRCS	RBBP8	UBE2C	IL6ST	AZGP1	DHCR7	MGP	STC2
EGFR (0,346), 201983_s_at CCND1 (0,331), 208712_at TRIM29 (0,325), 202504_at DHCR7 (0,323), 201791_s_at PIP (0,308), 206509_at TFAP28 (0,306), 214451_at WNTSA (0,303), 205990_s_at APOD (0,301), 201525_at PTPRT (0,301), 205948_at		PRSS16 (0,326), 208165_s_at EGFR (0,320), 201983_s_at DHCR7 (0,315), 201791_s_at EPHX2 (0,315), 209368_at TRIM29 (0,311), 202504_at					

Las secuencias de los cebadores y sondas son como sigue:

Tabla 1 Secuencias de cebadores y sonda para los genes respectivos:

gen	sonda	Seq ID	Seq ID	Seq ID	Seq ID	Seq ID	Seq ID
ABAT	TCGCCCTAAGAGGCTCT- TCCTC	1	2	3	4	5	6
ADRA2A	TTGTCTTTTCCCCCTC- CGTGC	4	5	6	7	8	9
APOD	CATCAGCTCTCAACTCCT- GGTTAACA	7	8	9	10	11	12
ASPH	TGGGAGGAAGGCAAGGT- GCTCATC	10	11	12	13	14	15
AURKA	CCGTCAGCCTGTGCTAG- GCAT	13	14	15	16	17	18
BIRC5	AGCCAGATGACGAC- CCCATAGAGGAACA	16	17	18	19	20	21
CELSR2	ACTGACTTTCCTTCT- GGAGCAGGTGGC	19	20	21	22	23	24
CHPT1	CCACGGCCACCAGAG- GCAC	22	23	24	25	26	27
CXCL12	CCACAGCAGGGTTTCAG- GTTCC	25	26	27	28	29	30
CYBRD1	AGGGCATGCCAT- CATCGTC	28	29	30	31	32	33
DCN	TCTTTTCAGCAACCCG- GTCCA	31	32	33	34	35	36
DHCR7	TGAGCGCCACCCCTCTC- GA	34	35	36	37	38	39
E2F8	CAGGATACCTAATC- CCTCTCACGCAG	37	38	39	40	41	42
EPHX2	TGAAGCGGAG- GACTTTTGTAAA	40	41	42	43	44	45
ESR1	ATGCCCTTTTGCCGATGCA	43	44	45	46	47	48
GJA1	TGCACAGCCTTTTGATTTTC- CCCGAT	46	47	48	49	50	51
HSPA2	CAAGTCAGCAAAACACG- CAAAA	49	50	51	52	53	54

(continuación)						
gen	sonda	Seq ID	cebador directo	Seq ID	cebador inverso	Seq ID
IL6ST	CAAGCTCCACCTTC-	52	CCCTGAATCCATAAAG-	53	CAGCTTCGTTTTTC-	54
	CAAAGGACCT		GCATACC		CCTACTTT-TT	
INPP4B	TCCGAGCGCTGGATT-	55	GCACCAAGTTACACAAG-	56	TCTCTATGCGGCATCCT-	57
	GCATGAG		GACTTCTTT.		TCTC	
MAPT	AGACTATTTGCACACT-	58	GTGGCTCAAAGGA-	59	ACCTTGCTCAGGTCAACT-	60
	GCCGCCCT		TAATATCAAACAC		GGTT	
MGP	CCTTCATATCCCCTCAG-	61	CCTTCATTAAACAGGA-	62	ATTGAGCTCGTGGACAG-	63
	CAGAGATGG		GAAATGCAA		GCTTA	
NEK2	TCCTGAACAAATGAATCG-	64	ATTGTTGGCACACCTTAT-	65	AAGCAGCCCAATGACCA-	66
	CATGTCCTACAA		TACATGT		GATa	
PCNA	AAATACTAAAATGCGCCG-	67	GGCGTGAACCTCAC-	68	CTTCGGCCCTTAGTG-	69
	GCAATGA		CAGTA		TAATGATATC	
PGR	TTGATAGAAACGCTGT-	70	AGCTCATCAAGGCAATT-	71	ACAAGATCATGCAAGT-	72
	GAGCTCGA		GGTTT		TATCAAGAAGTT	
PIP	TGCATGGTGGTTAAAACT-	73	TGCTTGCAGTTCAAACA-	74	CACCTTGTAGAGGGAT-	75
	TACCTCA		GAATTG		GCTGCTA	
PLAT	CAGAAAGTGGCCAT-	76	TGGGAAGACATGAAT-	77	GGAGGTTGGGCTTTAGdT-	78
	GCCACCCCTG		GCACACTA		GAA	
PRSS16	CACTGCCGGTCACT-	79	CTGAGGAGCACAGAACCCT-	80	CGAACTCGGTACATGTCT-	81
	CCACACCA		CAACT		GATACAA	
PTGER3	TCGGTCTGCTGGTCTC-	82	CTGATTGAAGATCATTT-	83	GACGGCCATTGAGCTTAT-	84
	CGCTCC		TCAACATCA		GG	
PTPRT	TTGGCTTCTGGACACCCT-	85	GAGTTGTGGCCTCTAC-	86	GAGCGGGAACCTTGGA-	87
	CACA		CATTGC		TAG	
RACGAP1	ACTGAGAATCTCCACCCG-	88	TCGCCAACTGGATAAAATT-	89	GAATGTGCGGAATCT-	90
	GCGCA		GGA		GTTTGAG	
RBBP8	ACCGATTCCGCTACATT-	91	AGAAATTGGCTTCCTGCT-	92	AAACCAACTTC-	93
	CACCCCAAC		CAAG		CCAAAAATTCTCT	
SCUBE2	CTAGAGGGTTCAGGTC-	94	TGTGGATTTCAGT-	95	CCATCTCGAACTATGTCT-	96
	CCATACGTGACATA		TCAAGTCCAATG		TCAATGAGT	
SEC14L2	TGGGAGGCAT-	97	AGGTCTTACTAAGCAGTC-	98	CGACCGGCACCTGAACTC.	99
	GCAACGCGTG		CCATCTCT			
SQLE	TATGCGTCTCCCAAAA-	100	GCAAGCTTCTTCTCCTCT-	101	CCTTTAGCAGTTTTTCTC-	102
	GAAGAACACCTCG		TCA		CATAGTTTTATATC	

(continuación)

gen	sonda	Seq ID	cebador directo	Seq ID	cebador inverso	Seq ID
TFAP2B	CAACACCAACCACTAACAG- GCACACGTC	103	GGCATGGACAAGATGT- TCTTGA	104	CCTCCTTGTGCGCCAGTTT- TACT	105
TOP2A	CAGATCAGGACCAAGAT- GGTTCCACAT	106	CATTGAAGACGCTTCGT- TATGG	107	CCAGTTGTGATGGA- TAAATTAATCAG	108
TRIM29	TGCTGTCTCACTACCG- GCCATTCTACG	109	TGGAAA-TCTGGCAAGCA- GACT	110	CAATCCCGTTGCCTTT- GTTG	111
UBE2C	TGAACACACATGCT- GCCGAGCTCTG	112	CTTCTAGGAGAACCCTAA- CATTGATAGT	113	GTTTCTTGCAGGTACTTCT- TAAAAGCT	114
WNT5A	TATTCACATCCCCTCAGTT- GCAGTGAATTG	115	CTGTGGCTCTTAATTTATT- GCATAATG	116	TTAGTGCTTTTGTGCTT- TCAAGATCTT	117
STC2	TCTCACCTTGACCCT- CAGCCAAG	118	ACATTTGACAAAATTTCCCT- TAGGATT	119	CCAGGACGCAGCTTTTAC- CAA	120

5 Una segunda alternativa para selección no supervisada de posibles candidatos de reemplazo de genes es a base solo de los datos de Affymetrix. Esto tiene la ventaja de que puede realizarse a base solo de datos ya publicados (por ejemplo procedentes de www.ncbi.nlm.nih.gov/geo/). Las siguientes tablas enumeran los candidatos de reemplazo del conjunto de sondas HG-U133a para los conjuntos de sondas utilizados en los algoritmos T1-T5. Esto es a base de los datos de entrenamiento de estos algoritmos. El encabezado de la columna contiene el nombre del gen y la ID del conjunto de sondas en negrita. Después, se enumeran los 10 conjuntos de sonda mejor correlacionados, en donde cada celda de la tabla contiene la ID del conjunto de sondas, el coeficiente de correlación entre paréntesis y el nombre del gen.

UBE2C	BIRCS	DHCR7	RACGAP1	AURKA	PVALB	NMU	STC2
202954_at	202095_s_at	201791_s_at	222077_s_at	204092_s_at	205336_at	206023_at	203438_at
210052_s_at (0,82) TPX2	202954_at (0,82) UBE2C	201790_s_at (0,66) DHCR7	218039_at (0,79) NUSAP1	208079_s_at (0,89) STK6	208663_at (-0,33) CAPN2	205347_s_at (0,45) TMSL8	203439_s_at (0,88) STC2
202095_s_at (0,82) BIRCS	218039_at (0,81) NUSAP1	202218_s_at (0,48) FADS2	214710_s_at (0,78) CCNB1	202954_at (0,80) UBE2C	219682_s_at (0,30) TBX3	203764_at (0,45) DLG7	212496_s_at (0,52) JMJD2B
218009_s_at (0,82) PRC1	218009_s_at (0,79) PRC1	202580_x_at (0,47) FOXM1	203764_at (0,77) DLG7	210052_s_at (0,77) TPX2	218704_at (0,30) FU20315	203554_x_at (0,44) PTTG1	219440_at (0,52) RAI2
203554_x_at (0,82) PTTG1	202705_at (0,78) CCNB2	208944_at (-0,46) TGFB2	204026_s_at (0,77) ZWINT	202095_s_at (0,77) BIRC5		204962_s_at (0,44) CENPA	215867_x_at (0,51) CA12
208079_s_at (0,81) STK6	204962_s_at (0,78) CENPA	202954_at (0,46) UBE2C	218009_s_at (0,76) PRC1	203554_x_at (0,76) PTTG1		204825_at (0,43) MELK	214164_x_at (0,50) CA12
202705_at (0,81) CCNB2	203554_x_at (0,78) PTTG1	209541_at (-0,45) IGF1	204641_at (0,76) NEK2	218009_s_at (0,75) PRC1		209714_s_at (0,41) CDKN3	204541_at (0,50) SEC14L2
218039_at (0,81) NUSAP1	208079_s_at (0,78) STK6	201059_at (0,45) CTTN	204444_at (0,75) KIF11	201292_at (0,73) TOP2A		219918_s_at (0,41) ASPM	203963_at (0,50) CA12
202870_s_at (0,80) CDC20	210052_s_at (0,77) TPX2	200795_at (-0,45) SPARCL1	202705_at (0,75) CCNB2	214710_s_at (0,73) CCNB1		207828_s_at (0,41) CENPF	212495_at (0,50) JMJD28
204092_s_at (0,80) STK6	202580_x_at (0,77) FOXM1	218009_s_at (0,45) PRC1	203362_s_at (0,75) MAD2L1	204962_s_at (0,73) CENPA		202705_at (0,41) CCNB2	208614_s_at (0,49) FLNB
209408_at (0,80) KIF2C	204092_s_at (0,77) STK6	218542_at (0,45) C10orf3	202954_at (0,75) UBE2C	218039_at (0,73) NUSAP1		219787_s_at (0,40) ECT2	213933_at (0,49) PTGER3

4ZGP1	RBBP8	IL6ST	MGP	PTGER3 CXCL12 ABAT			CDH1
209309_at	203344_s_at	212196_at	202291_s_at	213933_at	209687_at	209460_at	201131_s_at
217014_s_at (0,92) AZGP1	36499_at. (0,49) CELSR2	212195_at (0,85) IL6ST	201288_at (0,46) ARHGDIB	210375_at (0,74) PTGER3	204955_at (0,81) SRPX	209459_s_at (0,92) ABAT	201130_s_at (0,57) CDH1
206509_at (0,52) PIP	204029_at (0,45) CELSR2	204864_s_at (0,75) IL6ST	219768_at (0,42) VTCN1	210831_s_at (0,74) PTGER3	209335_at (0,81) DCN	206527_at (0,63) ABAT	221597_s_at (0,40) HSPC171
204541_at (0,46) SEC14L2	208305_at (0,45) PGR	211000_s_at (0,68) IL6ST	202849_x_at (-0,41) GRK6	210374_x_at (0,73) PTGER3	211896_s_at (0,81) DCN	213392_at (0,54) MGC35048	203350_at (0,38) AP1G1
200670_at (0,45) XBP1	205380_at (0,43) PDZK1	214077_x_at (0,61) MEIS4	205382_s_at (0,40) DF	210832_x_at (0,73) PTGER3	201893_x_at (0,81) DCN	221666_s_at (0,49) PYCARD	209163_at (0,36) CYB561
209368_at (0,45) EPHX2	203303_at (0,41) TCTE1L	204863_s_at (0,58) IL6ST	200099_s_at (0,39) RPS3A	210834_s_at (0,55) PTGER3	203666_at (0,80) CXCL12	218016_s_at (0,48) POLR3E	210239_at (0,35) IRX5
218627_at (-0,43) FU11259	205280_at (0,38) GLRB	202089_s_at (0,57) SLC39A6	221591_s_at (-0,37) FAM64A	210833_at (0,55) PTGER3	211813_x_at (0,80) DCN	214440_at (0,46) NAT1	200942_s_at (0,34) HSBP1
202286_s_at (0,43) TACSTD2	205279_s_at (0,38) GLRB	210735_s_at (0,56) CA12	214629_x_at (0,37) RTN4	203438_at (0,49) STC2	208747_s_at (0,79) C1S	204981_at (0,45) SLC22A18	209157_at (0,34) DNAJA2
213832_at (0,42)	203685_at (0,38) BCL2	200648_s_at (0,52) GLUL	200748_s_at (0,37) FTH1	203439_s_at (0,46) STC2	203131_at (0,78) PDGFRA	212195_at (0,45) IL6ST	210715_s_at (0,33) SPINT2
204288_s_at (0,41) SORBS2	203304_at (-0,39) BAMBI	214552_s_at (0,52) RABEP1	209408_at (-0,37) KIF2C	212195_at (0,41) IL6ST	202994_s_at (0,78) FBLN1	204497_at (0,45) ADCY9	203219_s_at (0,33) APRT
202376_at (0,41) SERPINA3	205862_at (0,36) GREB1	219197_s_at (0,51) SCUBE2	218726_at (-0,36) DKFZp762E1312	217764_s_at (0,40) RAB31	208944_at (0,78) TGFB2	215867_x_at (0,45) CA12	218074_at (0,33) FAM96B

Tras la selección de un gen o de un conjunto de sondas se tiene que definir un mapeo matemático entre los valores de expresión del gen a reemplazar y los del gen nuevo. Existen varias alternativas que se discuten en el presente documento a base del ejemplo “valores delta Ct del reemplazo de BIRC5 por RACGAP1”. En los datos de entrenamiento la distribución conjunta de las expresiones se asemeja a la Figura 3.

- 5 El coeficiente de correlación de Pearson es 0,73.

Una estrategia es crear una función de mapeo de RACGAP1 a BIRC5 mediante regresión. La regresión lineal es la primera opción y se produce en este ejemplo

$$\text{BIRC5} = 1,22 * \text{RACGAP1} - 2,85$$

- 10 Usando la ecuación se puede reemplazar fácilmente la variable BIRC5 en, por ejemplo, el algoritmo T5 por el lado derecho. En otros ejemplos puede ser adecuada la regresión robusta, la regresión polinomial o las pre-transformaciones no lineales de una variable.

El procedimiento de regresión asume el ruido de medición en BIRC5, pero no hay ruido en RACGAP1. Por lo tanto, el mapeo no es simétrico con respecto a la intercambiabilidad de las dos variables. Una estrategia de mapeo simétrico sería a base de dos transformaciones de z de una variable.

- 15
- $$z = (\text{BIRC5} - (\text{BIRC5}) \text{ media}) / (\text{BIRC5}) \text{ típico}$$
- $$z = (\text{RACGAP1} - (\text{RACGAP1}) \text{ media}) / (\text{RACGAP1}) \text{ típico}$$
- $$z = (\text{BIRC5} - 8,09) / 1,29 = (\text{RACGAP1} - 8,95) / 0,77$$
- $$\text{BIRC5} = 1,67 * \text{RACGAP1} + -6,89$$

- 20 Otra vez, en otros ejemplos pueden ser adecuadas otras transformaciones: la normalización mediante la mediana y/o mad, mapeos no lineales, u otros.

$$z = (\text{BIRC5} - (\text{BIRC5}) \text{ media}) / (\text{BIRC5}) \text{ típico}$$

$$z = (\text{RACGAP1} - (\text{RACGAP1}) \text{ media}) / (\text{RACGAP1}) \text{ típico}$$

$$z = (\text{BIRC5} - 8,09) / 1,29 = (\text{RACGAP1} - 8,95) / 0,77$$

$$\text{BIRC5} = 1,67 * \text{RACGAP1} + -6,89$$

- 25 Otra vez, en otros ejemplos pueden ser adecuadas otras transformaciones: la normalización mediante la mediana y/o mad, el mapeo no lineal, u otros.

REIVINDICACIONES

1. Un procedimiento para predecir una respuesta a y/o el beneficio de la quimioterapia con taxano o antraciclina, incluyendo quimioterapia no adyuvante, en un paciente que padece, o está en riesgo de desarrollar, cáncer de mama recurrente, comprendiendo dicho procedimiento las etapas de:
 - 5 (a) determinar en una muestra de tumor de mama procedente de dicho paciente los niveles de expresión de ARN de los siguientes 8 genes: UBE2C, RACGAP1, DHCR7, STC2, AZGP1, RBBP8, IL6ST y MGP, indicativos de una respuesta a la quimioterapia para un tumor, o
 - (b) determinar en una muestra de tumor de mama procedente de dicho paciente los niveles de expresión de ARN de los siguientes 8 genes: UBE2C, BIRC5, DHCR7, STC2, AZGP1, RBBP8, IL6ST y MGP; indicativos de una
 - 10 (c) combinar de forma matemática los valores del nivel de expresión para los genes de dicho conjunto, cuyos valores se determinaron en la muestra de tumor para producir una puntuación combinada, en la que dicha puntuación combinada está prediciendo dicha respuesta y/o el beneficio de la quimioterapia.
2. El procedimiento de una cualquiera de las reivindicaciones anteriores, en el que dicho nivel de expresión se
- 15 determina como no-proteína, tal como un nivel de expresión de gen.
3. El procedimiento de una cualquiera de las reivindicaciones anteriores, en el que dicho nivel de expresión se determina mediante al menos uno de
 - un procedimiento basado en PCR,
 - un procedimiento basado en micromatriz, o
 - 20 un procedimiento basado en hibridación, un enfoque de secuenciación y/o de secuenciación de última generación.
4. El procedimiento de una cualquiera de las reivindicaciones anteriores, en el que dicha determinación de los niveles de expresión es en una muestra de tumor fijada con formalina incluida en parafina o en una muestra de tumor reciente congelada.
- 25 5. El procedimiento de una cualquiera de las reivindicaciones anteriores, en el que el nivel de expresión de dicho al menos un gen marcador se determina como un patrón de expresión con respecto a al menos un gen de referencia o con respecto a un valor de expresión promedio calculado.
6. El procedimiento de una cualquiera de las reivindicaciones anteriores, en el que dicha etapa de combinar de forma matemática comprende una etapa de aplicación de un algoritmo a valores representativos de un nivel de
- 30 expresión de un gen dado, en particular en el que dicho algoritmo es una combinación lineal de dichos valores representativos de un nivel de expresión de un gen dado, o en el que un valor para un nivel de expresión representativo de un gen dado se multiplica por un coeficiente.
7. El procedimiento de una cualquiera de las reivindicaciones anteriores, en el que se determinan uno, dos o más umbrales para dicha puntuación combinada y se discrimina en riesgo elevado y bajo, riesgo elevado, intermedio y
- 35 bajo, o más grupos de riesgo mediante la aplicación del umbral en la puntuación combinada.
8. El procedimiento de una cualquiera de las reivindicaciones anteriores, en el que una puntuación combinada elevada es indicativa del beneficio de una terapia más agresiva, por ejemplo la quimioterapia citotóxica.
9. El procedimiento de una cualquiera de las reivindicaciones anteriores, en el que la información con respecto al estado de los ganglios del paciente se procesa en la etapa de la combinación de forma matemática de los valores
- 40 del nivel de expresión para los genes, para producir una puntuación combinada.
10. El procedimiento de una cualquiera de las reivindicaciones anteriores, en el que dicha información con respecto al tamaño del tumor del paciente se procesa en la etapa de combinar de forma matemática los valores del nivel de expresión para los genes, para producir una puntuación combinada.
11. El procedimiento de una cualquiera de las reivindicaciones anteriores, en el que dicha información con respecto al estado de los ganglios y el tamaño del tumor del paciente se procesa en la etapa de combinar de forma
- 45 matemática los valores del nivel de expresión para los genes, para producir una puntuación combinada.
12. Uso de un kit para realizar el procedimiento de al menos una de las reivindicaciones 1 a 11, comprendiendo dicho kit un conjunto de oligonucleótidos que tiene la capacidad de unirse de forma específica a secuencias o a secuencias de fragmentos de los genes en una combinación de genes, en el que
 - 50 (i) dicha combinación comprende al menos los 8 genes UBE2C, BIRC5, DHCR7, STC2, AZGP1, RBBP8, IL6ST y MGP; o
 - (ii) dicha combinación comprende al menos los 8 genes UBE2C, RACGAP1, DHCR7, STC2, AZGP1, RBBP8, IL6ST y MGP.

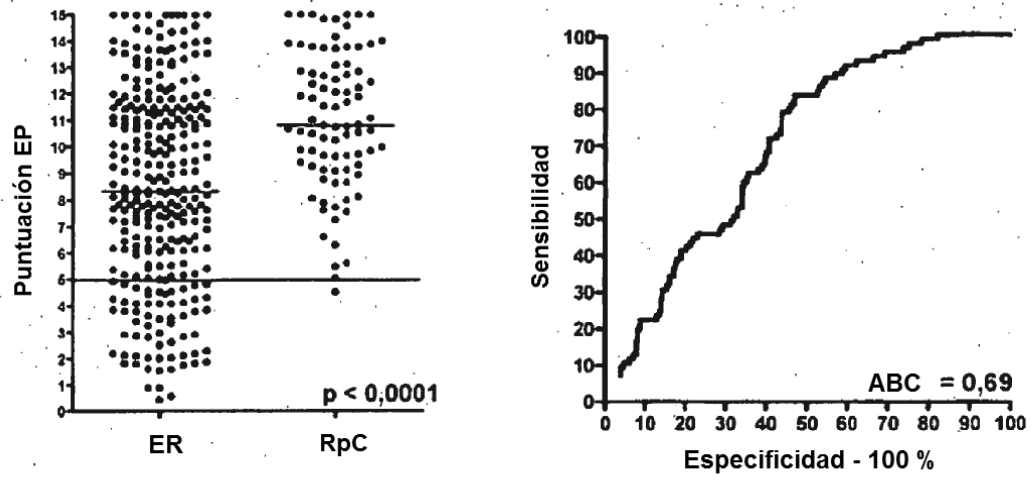


Figura 1

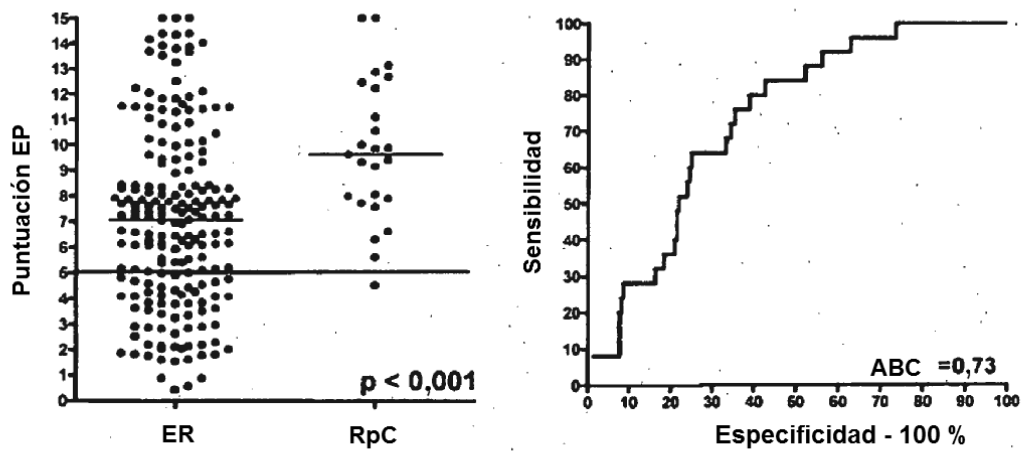


Figura 2

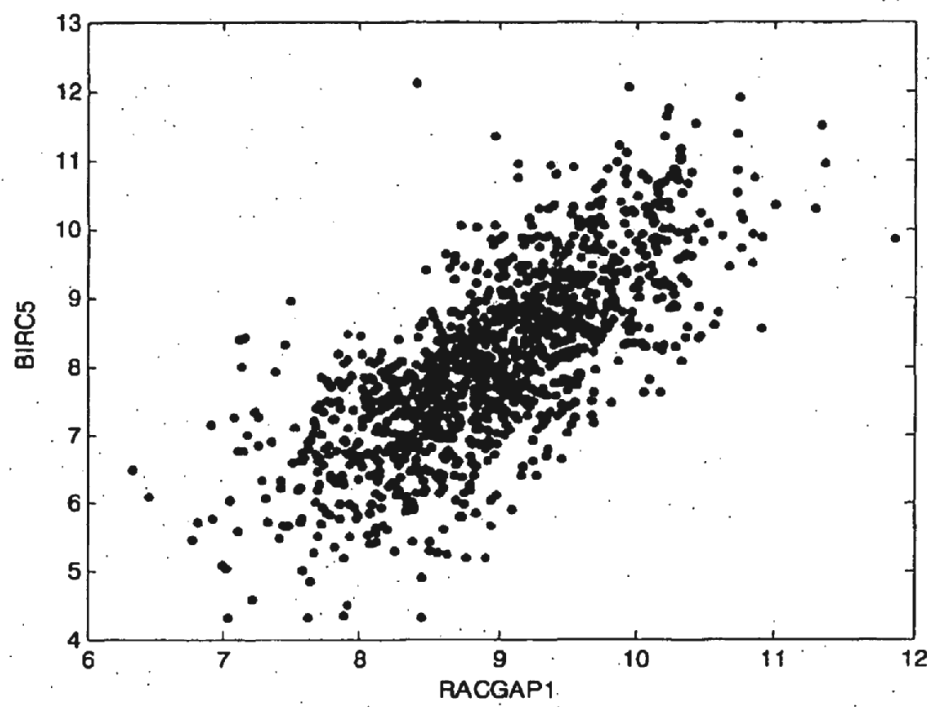


Figura 3