

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 525 382**

51 Int. Cl.:

C12Q 1/68 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **29.03.2011 E 11710526 (2)**

97 Fecha y número de publicación de la concesión europea: **15.10.2014 EP 2553118**

54 Título: **Método para la predicción de recurrencia del cáncer de mama bajo tratamiento endocrino**

30 Prioridad:

31.03.2010 EP 10158561

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

23.12.2014

73 Titular/es:

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

72 Inventor/es:

**DARTMANN, MAREIKE;
FEDER, INKE SABINE;
GEHRMANN, MATHIAS;
HENNIG, GUIDO;
WEBER, KARSTEN;
VON TÖRNE, CHRISTIAN;
KRONENWETT, RALF y
PETRY, CHRISTOPH**

74 Agente/Representante:

DE ELZABURU MÁRQUEZ, Alberto

ES 2 525 382 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

Método para la predicción de recurrencia del cáncer de mama bajo tratamiento endocrino

Campo técnico

5 La presente invención se refiere a métodos, kits y sistemas para la prognosis del resultado de la enfermedad del cáncer de mama. Más específicamente, la presente invención se refiere a la prognosis del cáncer de mama basándose en mediciones de los niveles de expresión de genes marcadores en muestras tumorales de pacientes con cáncer de mama.

Antecedentes de la invención

10 El cáncer de mama es una de las causas principales de muerte por cáncer en mujeres en los países occidentales. Más específicamente, el cáncer de mama se cobra la vida de aproximadamente 40.000 mujeres y se diagnostica en aproximadamente 200.000 mujeres anualmente solo en los Estados Unidos. A lo largo de las últimas décadas, la terapia sistémica coadyuvante ha conducido a una supervivencia marcadamente mejorada en el cáncer de mama temprano. Esta experiencia clínica ha conducido a recomendaciones consenso que ofrecen una terapia sistémica coadyuvante para la inmensa mayoría de los pacientes con cáncer de mama (EBCAG). En el cáncer de mama se encuentra disponible una multitud de opciones de tratamiento que se pueden aplicar además de la eliminación quirúrgica del tumor realizada rutinariamente y la posterior radiación del lecho del tumor. Tres estrategias principales y conceptualmente diferentes son el tratamiento endocrino, la quimioterapia y el tratamiento con terapias dirigidas. El requisito previo para el tratamiento con agentes endocrinos es la expresión de receptores hormonales en el tejido tumoral, esto es, receptores de estrógeno, receptores de progesterona o ambos. Se encuentran disponibles varios agentes endocrinos con modos de acción diferentes y diferencias en el resultado de la enfermedad cuando se someten a ensayo en grandes cohortes de pacientes. El tamoxifeno ha sido el pilar del tratamiento endocrino durante las tres últimas décadas. Grandes pruebas clínicas demostraron que el tamoxifeno reduce significativamente el riesgo de recurrencia del tumor. Una opción de tratamiento adicional se basa en inhibidores de aromatasa que pertenecen a una nueva clase de fármacos endocrinos. En contraste con el tamoxifeno que es un inhibidor competitivo de la unión a estrógeno, los inhibidores de aromatasa bloquean la producción de estrógeno como tal, reduciendo de ese modo el estímulo de crecimiento de las células tumorales positivas para receptores de estrógeno. Aún así, algunos pacientes experimentan una recaída a pesar del tratamiento endocrino y en particular estos pacientes se podrían beneficiar de fármacos terapéuticos adicionales. Se ha demostrado que la quimioterapia con antraciclinas, taxanos y otros agentes resulta beneficiosa en la reducción de la recurrencia de la enfermedad en pacientes positivos para receptores de estrógeno así como negativos para receptores de estrógeno. El estudio de NSABP-20 comparó el tamoxifeno solo frente al tamoxifeno más quimioterapia en pacientes positivos para receptores de estrógeno con ganglios negativos y demostró que el tratamiento combinado era más eficaz que el tamoxifeno solo. Sin embargo, el estudio de IBCSG IX que comparaba el tamoxifeno solo con el tamoxifeno más quimioterapia no logró demostrar ningún beneficio significativo para la adición de agentes citotóxicos. Recientemente, se ha demostrado que un anticuerpo administrado sistémicamente dirigido contra el antígeno HER2/neu sobre la superficie de las células tumorales reduce el riesgo de recurrencia varias veces en pacientes con tumores que expresan en exceso Her2neu. Sin embargo, la mayoría, si no todos los diferentes tratamientos con fármacos tienen numerosos efectos adversos potenciales que pueden deteriorar gravemente la calidad de vida de los pacientes (Shapiro y Recht, 2001; Ganz et al., 2002). Esto obliga a seleccionar la estrategia de tratamiento basándose en una evaluación cuidadosa del riesgo para que el paciente individual evite sobre- así como infra-tratamiento. Puesto que el beneficio de la quimioterapia es relativamente grande en los tumores HER2/neu positivos y en los caracterizados por la ausencia de la expresión de HER2/neu y receptores de estrógeno (tipo basal), en comparación con los tumores negativos para HER2/neu y positivos para receptores de estrógeno (tipo luminal), la decisión del tratamiento más complicada está relacionada con los tumores luminales para los cuales los factores clínicos clásicos como la gradación, el tamaño del tumor o la implicación de los ganglios linfáticos no proporcionan una respuesta clara a la cuestión de si utilizar quimioterapia o no. Se han desarrollado herramientas moleculares más novedosas como un análisis del gen 21, un análisis del índice de grado genómico y otros para abordar esta necesidad médica.

50 Las pautas de tratamiento son desarrolladas normalmente por expertos reconocidos en la técnica. En Europa las pautas de St. Gallen desde el año 2009 recomiendan la quimioterapia para pacientes con cáncer de mama positivo para HER2 así como para pacientes con enfermedad negativa para HER2 y negativa para ER. La incertidumbre acerca de la utilidad de la quimioterapia se presenta en pacientes con enfermedad negativa para HER2 y positiva para ER. Con el fin de tomar una decisión de tratamiento equilibrada para el individuo, se utiliza la probabilidad de recurrencia del cáncer como el criterio más útil. Los criterios clínicos como el estado de los ganglios linfáticos, la gradación del tumor, el tamaño del tumor y otros son provechosos puesto que proporcionan información sobre el riesgo de recurrencia. Más recientemente, se ha demostrado que los análisis de múltiples genes proporcionan información superior o adicional a los factores de riesgo clínicos convencionales. Generalmente se reconoce, que los marcadores de proliferación parecen proporcionar la información pronóstica dominante. Los ejemplos prominentes de esos predictores son el ensayo Mammaprint de Agendia, el Relapse Score de Veridex y el Genomic Grade Index, desarrollado en el instituto Jules Bordet y con licencia para Ipsogen. Todos estos análisis se basan en la determinación de los niveles de expresión de al menos 70 genes y todos se han desarrollado para ARN no

severamente degradado por la fijación con formalina y la inclusión en parafina, pero aislado de tejido fresco (enviado en RNALater™). Otro análisis de múltiples genes prominente es el ensayo Recurrence Score de Genomic Health Inc. El ensayo determina el nivel de expresión de 16 genes relacionados con el cáncer y 5 genes de referencia tras la extracción de ARN de muestras de tejido fijadas con formalina e incluidas en parafina.

- 5 Sin embargo, las herramientas corrientes adolecen de una carencia de validez y utilidad clínicas en el grupo de riesgo clínico más importante, esto es, aquellos pacientes con cáncer de mama con un riesgo intermedio de recurrencia basado en parámetros clínicos convencionales. Por lo tanto, se necesitan mejores herramientas para optimizar las decisiones de tratamiento basándose en el pronóstico del paciente. Por la utilidad clínica de evitar la quimioterapia, se necesita un ensayo con una sensibilidad elevada y un valor predictivo negativo elevado, con el fin de no infratratarse a un paciente que eventualmente desarrolla una metástasis distante después de la cirugía.

Con respecto a la continua necesidad de materiales y métodos útiles en la toma de decisiones clínicas sobre una terapia coadyuvante, la presente invención satisface la necesidad de métodos avanzados para la prognosis del cáncer de mama basándose en datos clínicos y experimentales fácilmente accesibles.

- 15 El documento WO2006/119593 A describe un método para predecir el resultado del cáncer de mama luminal mediante la determinación de los niveles de expresión de numerosos genes. El cáncer de mama luminal es positivo para receptores de estrógenos y negativo para HER2. El conjunto de marcadores incluye p. ej. UBE2C y BIRC5. También se describe una firma de la expresión génica que se puede utilizar para predecir el resultado en el cáncer de mama luminal. Los marcadores incluyen UBE2C y BIRC5.

- 20 Sorlie et al. describen en PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES (PNAS), NATIONAL ACADEMY OF SCIENCE, US, vol. 98, núm. 19, 11 Septiembre 2001 (11-09-2001), páginas 10869-10874 un conjunto de 264 genes que se correlaciona con el resultado del cáncer de mama en el paciente en diversas categorías de muestras, una de las cuales es cáncer de mama luminal, esto es, cáncer de mama positivo para receptores de estrógeno y negativo para HER2. El conjunto de genes pronósticos, incluye, p. ej. proteína portadora de ubiquitina (UBE2C) y receptor de estrógeno 1 (ESR1).

- 25 Von Minckwitz et al. describen en CANCER RESEARCH, AMERICAN ASSOCIATION FOR CANCER RESEARCH, US, vol. 69, núm. 24, supl. 1 Diciembre 2009 (01-12-2009), página 635S, un método para predecir el resultado del cáncer de mama luminal utilizando marcadores de expresión génica.

- 30 Chanrion et al. describen en CLINICAL CANCER RESEARCH: AN OFFICIAL JOURNAL OF THE AMERICAN ASSOCIATION FOR CANCER RESEARCH 15 MAR 2008 LNKD-PUBMED:18347175, vol. 14, núm. 6, 15 Marzo 2008 (15-03-2008), páginas 1744-1752 un método para predecir el resultado de un tumor de mama positivo para receptores de estrógeno tratado con tamoxifeno utilizando como marcadores un conjunto de 36 genes. El conjunto de genes incluye, entre otros, los genes STC2 y UBE2C.

- 35 Paik et al. describen en NEW ENGLAND JOURNAL OF MEDICINE, MASSACHUSETTS MEDICAL SOCIETY, BOSTON, MA, US, vol. 351, núm. 27, 30 Diciembre 2004 (30-12-2004), páginas 2817-2826, la predicción del resultado clínico de tumores positivos para receptores de estrógeno y negativos para HER2 mediante la determinación de los niveles de expresión de genes. Se utiliza un conjunto de 21 genes que incluye, p. ej. los genes BIRC5 (survivalina), SCUBE2 y ESR1 (ER).

- 40 El documento WO2009/158143 A1 describe métodos para clasificar y para evaluar la prognosis de un sujeto que tiene cáncer de mama. Los métodos incluyen la predicción del subtipo de cáncer de mama utilizando un algoritmo supervisado entrenado para estratificar sujetos basándose en el subtipo intrínseco de cáncer de mama. El modelo de predicción se basa en el perfil de expresión génica de los genes intrínsecos descritos. Este modelo de predicción se puede utilizar para predecir exactamente el subtipo intrínseco de un sujeto diagnosticado de o que se sospecha que tiene cáncer de mama. Adicionalmente se describen composiciones y métodos para predecir el resultado o la respuesta a la terapia de un sujeto diagnosticado de o que se sospecha que tiene cáncer de mama. Estos métodos son útiles para guiar o determinar las opciones de tratamiento para un sujeto aquejado de cáncer de mama. Los métodos incluyen adicionalmente métodos para evaluar los perfiles de expresión génica, incluyendo micromatrices y análisis de reacción en cadena de la polimerasa cuantitativos, así como kits que comprenden reactivos para poner en práctica los métodos.

- 50 El documento WO 2010/003773 A describe métodos para predecir el resultado del cáncer en un paciente que padece cáncer, habiendo diagnosticado previamente que dicho paciente tiene ganglios positivos y habiéndolo tratado con quimioterapia citotóxica, comprendiendo dicho método determinar en una muestra biológica de dicho paciente un nivel de expresión de una pluralidad de genes seleccionados del grupo que consiste en ACTG1, CA12, CALM2, CCND1, CHPT1, CLEC2B, CTSB, CXCL13, DCN, DHRS2, EIF4B, ERBB2, ESR1, FBXO28, GABRP, GAPDH, H2AFZ, IGFBP3, IGHG1, IGKC, KCTD3, KIAA0101, KRT17, MLPH, MMP1, NAT1, NEK2, NR2F2, OAZ1, PCNA, PDLIM5, PGR, PPIA, PRC1, RACGAP1, RPL37A, SOX4, TOP2A, UBE2C y VEGF; ABCB1, ABCG2, ADAM15, AKR1C1, AKR1C3, AKT1, BANF1, BCL2, BIRC5, BRMS1, CASP10, CCNE2, CENPJ, CHPT1, EGFR, CTTN, ERBB3, ERBB4, FBLN1, FIP1L1, FLT1, FLT4, FNNTA, GATA3, GSTP1, Herstatina, IGF1R, IGHM, KDR, KIT, CKRT5, SLC39A6, MAPK3, MAPT, MKI67, MMP7, MTA1, FRAP1, MUC1, MYC, NCOA3, NFIB, OLFM1, TP53,

PCNA, PI3K, PPERLD1, RAB31, RAD54B, RAF1, SCUBE2, STAU, TINF2, TMSL8, VGLL1, TRA@, TUBA1, TUBB, TUBB2A.

5 El documento WO 2009/132928 A2 describe un método de identificación de un panel de genes informativo del resultado de la enfermedad que se pueden combinar en un algoritmo para un ensayo pronóstico o predictivo. El algoritmo hace uso de datos de expresión génica de muestras biológicas y clasifica los pacientes que tienen un riesgo elevado o un riesgo bajo, p. ej. en pacientes con cáncer en un grupo con un mal resultado por metástasis o uno con buen resultado. Los patrones de referencia de la expresión génica se obtienen para los grupos de alto riesgo y bajo riesgo, respectivamente. Una muestra de un paciente desconocido se analiza y se clasifica como perteneciente a un grupo de alto riesgo o bajo riesgo, respectivamente, dependiendo de la correlación con el patrón de referencia de alto riesgo o el patrón de referencia de bajo riesgo.

15 El documento WO 2008/006517 A2 describe métodos y kits para la predicción de un resultado probable de la quimioterapia en un paciente con cáncer. Más específicamente, la invención se refiere a la predicción de la respuesta del tumor a la quimioterapia basándose en 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 basada en taxano, tal como p. ej. quimioterapia basada en taxano-antraciclina-ciclofosfamida (p. ej. basada en Taxótero (docetaxel)-Adriamicina (doxorubicina)-Ciclofosfamida, esto es (TAC)).

20 El documento EP 2 163 649 A1 describe un clasificador molecular de 38 genes que consiste en los siguientes genes: BTG2, UBE2C, VEGFA, SLC25A5, FABPS, ENO1, PSMB7, TUBB6, TACC3, SURF4, RHOC, CHCHD2, SFRS5, TNFSF13, DDT, Clorf64, RAB10, XIST, TUBA4A, PPA1, C4B, TUBB2C, C3orf37, GATA3, PGK1, PPP1CA, BCL2A1, IL6ST, PSMB3, RPN1, ETVB, CYC1, POR, TUBB3, TMED9, TOE1, C17orf37, PSMA. El nivel de expresión de estos genes en el tejido tumoral de mama es un marcador útil para evaluar el riesgo de recaída metastásica en pacientes con cáncer de mama.

25 El documento WO 2008/079269 A2 describe grupos de genes que se pueden utilizar para diagnosticar y diferenciar entre tipos de enfermedades específicas tales como el cáncer de mama. Los grupos de genes se pueden utilizar adicionalmente para desarrollar kits de diagnóstico para las enfermedades específicas. Los kits de diagnóstico también pueden diferenciar entre subcategorías de una enfermedad para ayudar a identificar el régimen de tratamiento apropiado para un paciente.

30 D. Loussouarn et al. describen en el British Journal of Cancer (2009) 101, 166-173, la validación de la proteína UBE2C como marcador pronóstico en el cáncer de mama con ganglios positivos.

35 El documento WO 2010/076322 A1 describe un método de predicción de la respuesta a y/o el beneficio de la quimioterapia en un paciente que padece cáncer que comprende las etapas de (i) clasificar un tumor en al menos dos clases, (ii) determinar en una muestra tumoral la expresión de al menos un gen marcador indicativo de una respuesta a la quimioterapia para un tumor en cada respectiva clase, (iii) dependiendo de la expresión de dicho gen, predecir dicha respuestas y/o beneficio; en donde 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, 1GKC, 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 mismos.

Definiciones

A menos que se defina de otro modo, los términos técnicos y científicos utilizados en la presente memoria tienen los mismos significados comúnmente entendidos por los expertos en la técnica a la cual pertenece esta invención.

45 El término "cáncer" no se limita a cualquier fase, grado, característica histomorfológica, agresividad, o malignidad de un tejido afectado o una agregación celular.

Se pretende que el término "predecir un resultado" de una enfermedad, según se utiliza en la presente memoria, incluya tanto una predicción de un resultado de un paciente que está experimentando una terapia dada como una prognosis de un paciente que no está tratado. El término "predecir un resultado" puede, en particular, hacer referencia al riesgo de un paciente de desarrollar metástasis, recurrencia local o muerte.

50 El término "predicción", según se utiliza en la presente memoria, hace referencia a una evaluación individual de la malignidad de un tumor, o a la tasa de supervivencia esperada (OAS, supervivencia global o DFS, supervivencia libre de enfermedad) de un paciente, si el tumor es tratado con una terapia dada. En contraste con esto, el término "prognosis" se refiere a una evaluación individual de la malignidad de un tumor, o a la tasa de supervivencia esperada (OAS, supervivencia global o DFS, supervivencia libre de enfermedad) de un paciente, si el tumor sigue sin ser tratado.

Un "resultado" dentro del significado de la presente invención es un estado definido alcanzado en el transcurso de la

enfermedad. Este resultado de la enfermedad puede ser, p. ej. un estado clínico tal como la "recurrencia de una enfermedad", el "desarrollo de metástasis", el "desarrollo de metástasis ganglionar", el "desarrollo de metástasis distante", la "supervivencia", la "muerte", la "tasa de emisión del tumor", una fase o un grado de la enfermedad o similar.

- 5 Se entiende que un "riesgo" es un número relacionado con la probabilidad de que un sujeto o paciente desarrolle o alcance un cierto resultado de la enfermedad. No se pretende que el término "riesgo" en el contexto de la presente invención signifique tener cualquier connotación positiva o negativa con respecto al bienestar de un paciente, sino simplemente se refiere a la probabilidad o verosimilitud de una incidencia o al desarrollo de una afección dada.

- 10 El término "datos clínicos" se refiere a la totalidad de los datos disponibles y la información relacionada con el estado de salud de un paciente incluyendo, pero no limitados a, la edad, sexo, peso, estado menopáusico/hormonal, datos etiopatológicos, datos de anamnesis, datos obtenidos mediante métodos diagnósticos in vitro tales como histopatología, análisis de sangre u orina, datos obtenidos mediante métodos de formación de imágenes, tales como rayos x, tomografía computarizada, MRI, PET, spect, ultrasonidos, datos electrofisiológicos, análisis genético, análisis de expresión génica, evaluación mediante biopsia, hallazgos intraoperatorios.

- 15 El término "de ganglios positivos", "diagnosticado como de ganglios positivos", "implicación de ganglios" o "implicación de ganglios linfáticos" representa un paciente que ha sido diagnosticado previamente de metástasis en los ganglios linfáticos. Esto incluirá tanto ganglios linfáticos de drenaje, ganglios linfáticos próximos, y metástasis de ganglios linfáticos distantes. Este diagnóstico previo por sí mismo no formará parte del método de la invención. En vez de eso es una condición previa para seleccionar a los pacientes cuyas muestras se pueden utilizar para una
 20 realización de la presente invención. Se puede alcanzar este diagnóstico previo mediante cualquier método adecuado conocido en la técnica, incluyendo, pero no limitado a eliminación de los ganglios linfáticos y análisis patológico, análisis por medio de biopsia, análisis in vitro de biomarcadores indicativos de metástasis, métodos de formación de imágenes (p. ej. tomografía computarizada, rayos X, formación de imágenes por resonancia magnética, ultrasonidos), y hallazgos intraoperatorios.

- 25 En el contexto de la presente invención una "muestra biológica" es una muestra que deriva de o ha estado en contacto con un organismo biológico. Los ejemplos de las muestras biológicas son: células, tejidos, fluidos corporales, fluidos de lavado, muestras de frotis, especímenes de biopsias, sangre, orina, saliva, esputo, plasma, suero, sobrenadante de cultivo celular, y otros.

- 30 Una "muestra de tumor" es una muestra biológica que contiene células tumorales, ya sea intactas o degradadas. La muestra puede ser de cualquier tejido o fluido biológicos. Tales muestras incluyen, pero no se limitan a, esputo, sangre, suero, plasma, células sanguíneas (p. ej., glóbulos blancos), tejido, muestras de biopsia por punción con aguja gruesa o fina, fluidos corporales que contienen células, orina, fluido peritoneal, y fluido pleural, líquido cefalorraquídeo, lágrimas, o células aisladas de los mismos. Esto también puede incluir secciones de tejidos tales como secciones congeladas o fijadas tomadas con fines histológicos o células microdisecionadas o porciones
 35 extracelulares de las mismas. La muestra de tumor que se va a analizar puede ser material tisular de una lesión neoplásica tomada por aspiración o punción, escisión o mediante cualquier otro método quirúrgico que conduzca a una biopsia o material celular resecaado. Esto comprende células tumorales o fragmentos de células tumorales obtenidos del paciente. Las células se pueden encontrar en un "frotis" celular recogido, por ejemplo, mediante aspiración del pezón, lavado ductal, biopsia con aguja fina o a partir de descarga del pezón provocada o
 40 espontánea. En otra realización, la muestra es un fluido corporal. Tales fluidos incluyen, por ejemplo, fluidos de la sangre, suero, plasma, linfa, fluidos ascíticos, fluidos ginecológicos, u orina, pero sin limitarse a estos fluidos.

Un "gen" es un conjunto de segmentos de ácido nucleico que contiene la información necesaria para producir un producto de ARN funcional. Un "producto génico" es una molécula biológica producida por medio de transcripción o expresión de un gen, p. ej. un ARNm, ADNc o la proteína traducida.

- 45 Un "ARNm" es el producto transcrito de un gen y deberá tener el significado normal entendido por un experto en la técnica. Una "molécula derivada de un ARNm" es una molécula que se obtiene químicamente o enzimáticamente a partir de un molde de ARNm, tal como ADNc.

- El término "nivel de expresión" hace referencia a un nivel determinado de expresión génica. Este puede ser un nivel determinado de expresión génica en forma de un valor absoluto o comparado con un gen de referencia (p. ej. un gen constitutivo), con la media de dos o más genes de referencia, o con un valor de expresión medio computado (p. ej. en un análisis con chip de ADN) o con otro gen informativo sin el uso de una muestra de referencia. El nivel de expresión de un gen se puede medir directamente, p. ej. obteniendo una señal en donde la intensidad de la señal se corresponde con la cantidad de transcritos de ARNm de ese gen o se puede obtener indirectamente a nivel de proteína, p. ej. mediante métodos de inmunohistoquímica, CISH, ELISA o RIA. El nivel de expresión también se
 50 puede obtener a título de una reacción competitiva para una muestra de referencia. Se puede asignar un valor de expresión que se determina midiendo algún parámetro físico en un análisis, p. ej. emisión de fluorescencia, a un valor numérico que se puede utilizar para el procesamiento adicional de la información.

Se entenderá que un "patrón de referencia de los niveles de expresión", dentro del significado de la invención es

cualquier patrón de niveles de expresión que se puede utilizar para la comparación con otro patrón de niveles de expresión. En una realización preferida de la invención, un patrón de referencia de niveles de expresión es, p. ej., un patrón medio de niveles de expresión observado en un grupo de individuos sanos, individuos enfermos, o individuos enfermos que han recibido un tipo concreto de terapia, que sirven como grupo de referencia, o individuos con un buen o mal resultado.

El término "combinar matemáticamente niveles de expresión", dentro del significado de la invención se entenderá como obtener un valor numérico de un nivel de expresión determinado de un gen y aplicar un algoritmo a uno o más de tales valores numéricos para obtener un valor numérico combinado o una puntuación combinada.

Un "algoritmo" es un proceso que realiza cierta secuencia de operaciones para producir información.

Una "puntuación" es un valor numérico que fue obtenido combinando matemáticamente niveles de expresión utilizando un algoritmo. También se puede obtener de niveles de expresión y otra información, p. ej. datos clínicos. Una puntuación puede estar relacionada con el resultado de la enfermedad de un paciente.

Una "función discriminante" es una función de un conjunto de variables utilizada para clasificar un objeto o evento. Una función discriminante permite de ese modo la clasificación de un paciente, una muestra o un evento en una categoría o una pluralidad de categorías de acuerdo con los datos o parámetros disponibles de dicho paciente, muestra o evento. Semejante clasificación es un instrumento convencional de análisis estadístico bien conocido por los expertos en la técnica. Por ejemplo, un paciente se puede clasificar como de "alto riesgo" o "bajo riesgo", "elevada probabilidad de metástasis" o "baja probabilidad de metástasis", "necesitado de tratamiento" o "no necesitado de tratamiento" de acuerdo con los datos obtenidos de dicho paciente, muestra o evento. La clasificación no está limitada a "elevada vs. baja", sino que se puede realizar en una pluralidad de categorías, gradación o similar. La clasificación también se deberá entender en un sentido más amplio como una puntuación discriminante, donde p. ej. una puntuación superior representa una mayor verosimilitud de metástasis distante, p. ej. el riesgo (global) de una metástasis distante. Los ejemplos de las funciones discriminantes que permiten una clasificación incluyen, pero no se limitan a, funciones definidas por máquinas de vectores de soporte (SVM), k vecinos más cercanos (kNN), modelos de Bayes (Naive), modelos de regresión lineal o funciones definidas a trozos tales como, por ejemplo, en el descubrimiento de un subgrupo, en árboles de decisión, en el análisis lógico de datos (LAD) y similares. En un sentido más amplio, los valores de puntuación continua de métodos matemáticos o algoritmos, tales como coeficientes de correlación, proyecciones, puntuaciones de máquinas de vectores de soporte, otros métodos basados en la similitud, combinaciones de estos y similares son ejemplos con fines ilustrativos.

Los términos "modalidad de terapia", "modo de terapia", "régimen" así como "régimen de terapia" se refieren a una administración secuencial o simultánea oportuna de agentes anti-tumorales, y/o anti-vasculares, y/o estimuladores inmunitarios, y/o de proliferación de células de la sangre, y/o terapia con radiación, y/o hipertermia, y/o hipotermia para la terapia del cáncer. La administración de estos se puede realizar en un modo con coadyuvante y/o neoadyuvante. La composición de semejante "protocolo" puede variar en la dosis del agente individual, el marco temporal de la aplicación y la frecuencia de administración dentro de una ventana de terapia definida. Actualmente están siendo investigadas diversas combinaciones de diversos fármacos y/o métodos físicos, y diversos programas.

El término "quimioterapia citotóxica" hace referencia a diversas modalidades de tratamiento que afectan a la proliferación y/o a la supervivencia celulares. El tratamiento puede incluir la administración de agentes alquilantes, antimetabolitos, antraciclinas, alcaloides de plantas, inhibidores de topoisomerasa, y otros agentes antitumorales, incluyendo anticuerpos monoclonales e inhibidores de quinasas. En particular, el tratamiento citotóxico puede relacionarse con un tratamiento con taxano. Los taxanos son alcaloides de plantas que bloquean la división celular evitando la función de los microtúbulos. El taxano prototipo es el producto natural paclitaxel, originalmente conocido como Taxol y derivado en primer lugar de la corteza del árbol Tejo del Pacífico. El docetaxel es un análogo semisintético del paclitaxel. Los taxanos potencian la estabilidad de los microtúbulos, evitando la separación de los cromosomas durante la anafase.

El término "tratamiento endocrino" o "tratamiento hormonal" (a veces también referido también como "tratamiento anti-hormonal") indica un tratamiento que dirige la señalización hormonal, p. ej. inhibición de hormonas, inhibición de receptores de hormonas, uso de agonistas o antagonistas de receptores de hormonas, uso de receptores captadores o huérfanos, uso de derivados de hormonas e interferencia en la producción de hormonas. Los ejemplos concretos son la terapia con tamoxifeno que modula la señalización del receptor de estrógeno, o el tratamiento con aromatasa que interfiere en la producción de hormonas esteroides.

El tamoxifeno es un modulador del receptor de estrógenos selectivo (SERM) activo oralmente que se utiliza en el tratamiento del cáncer de mama y actualmente es el fármaco que más se vende en el mundo para este fin. El tamoxifeno se comercializa con los nombres comerciales Nolvadex, Istubal, y Valodex. Sin embargo, el fármaco, incluso antes de la expiración de la patente, fue, y todavía es, ampliamente referido por su nombre genérico "tamoxifeno". El tamoxifeno y los derivados de tamoxifeno se unen competitivamente a receptores de estrógeno sobre los tumores y otras dianas tisulares, produciendo un complejo nuclear que disminuye la síntesis de ARN e inhibe los efectos de los estrógenos.

Los receptores de esteroides son receptores intracelulares (típicamente citoplásmicos) que realizan la transducción de señales para las hormonas esteroides. Los ejemplos incluyen Receptores de tipo I, en particular receptores de hormonas sexuales, p. ej. receptor de andrógeno, receptor de estrógeno, receptor de progesterona; receptor de glucocorticoide, receptor de mineralocorticoide; y Receptores de tipo II, p. ej. receptor de vitamina A, receptor de vitamina D, receptor de retinoide, receptor de hormona tiroidea.

El término "método basado en la hibridación", según se utiliza en la presente memoria, hace referencia a métodos que confieren un proceso de combinación de ácidos nucleicos de hebra sencilla, complementarios o análogos de nucleótidos en una única molécula de doble hebra. Los nucleótidos o análogos de nucleótidos se unirán a su complemento en condiciones normales, de manera que dos hebras perfectamente complementarias se unirán entre sí fácilmente. En bioanalítica, muy a menudo se utilizan sondas de hebra sencilla, marcadas con el fin de encontrar secuencias diana complementarias. Si tales secuencias existen en la muestra, las sondas hibridarán con dichas secuencias que a continuación pueden ser detectadas debido a la marca. Otros métodos basados en la hibridación comprenden métodos con micromatrices y/o biochips. Allí, las sondas son inmovilizadas sobre una fase sólida, que a continuación es expuesta a una muestra. Si existen ácidos nucleicos complementarios en la muestra, éstos hibridarán con las sondas y de ese modo se podrán detectar. Estos enfoques también son conocidos como "métodos basados en matrices". Otro método más basado en la hibridación es la PCR, que se describe más abajo. Cuando se trata de la determinación de los niveles de expresión, se pueden utilizar, por ejemplo, métodos basados en la hibridación para determinar la cantidad de ARNm para un gen dado.

Un oligonucleótido capaz de unirse específicamente a secuencias de un gen o fragmentos del mismo hace referencia a un oligonucleótido que hibrida específicamente con un gen o producto génico, tal como el ARNm o el ADNc del gen o con un fragmento del mismo. Para detectar específicamente el gen o producto génico, no es necesario detectar la secuencia génica completa. Un fragmento de aproximadamente 20-150 bases contendrá suficiente información específica de la secuencia como para permitir la hibridación específica.

El término "un método basado en la PCR" según se utiliza en la presente memoria hace referencia a métodos que comprenden una reacción en cadena de la polimerasa (PCR). Este es un método para amplificar exponencialmente ácidos nucleicos, p. ej. ADN por medio de replicación enzimática in vitro. Puesto que la PCR es una técnica in vitro, se puede llevar a cabo sin restricciones sobre la forma de ADN, y se puede modificar considerablemente para realizar una amplia gama de manipulaciones genéticas. Cuando se trata de la determinación de los niveles de expresión, se puede utilizar, por ejemplo, un método basado en la PCR para detectar la presencia de un ARNm dado por medio de (1) transcripción inversa de la reserva completa de ARNm (el dominado transcriptoma) a ADNc con la ayuda de una enzima transcriptasa inversa, y (2) detección de la presencia de un ADNc dado con la ayuda de los respectivos cebadores. Este enfoque es comúnmente conocido como PCR con transcriptasa inversa (rtPCR).

Por otra parte, los métodos basados en PCR comprenden p. ej. PCR en tiempo real, y, particularmente adecuada para el análisis de los niveles de expresión, PCR cinética o cuantitativa (qPCR).

El término "PCR Cuantitativa" (qPCR) hace referencia a cualquier tipo de método de PCR que permite la cuantificación del molde en una muestra. La PCR en tiempo real cuantitativa comprende diferentes técnicas de funcionamiento o detección del producto como por ejemplo la técnica TaqMan o la técnica LightCycler. La técnica TaqMan, por ejemplo, utiliza una sonda fluorogénica doblemente marcada. La PCR en tiempo real TaqMan mide la acumulación de un producto a través del fluoróforo durante las fases exponenciales de la PCR, en lugar de en el momento final como en la PCR convencional. El incremento exponencial del producto se utiliza para determinar el ciclo umbral, CT, esto es, el número de ciclos de PCR al cual se detecta un incremento exponencial significativo, y que está directamente correlacionado con el número de copias del molde de ADN presente en la reacción. El ajuste de la reacción es muy similar al de una PCR convencional, pero se lleva a cabo en un ciclador térmico a tiempo real que permite la medición de las moléculas fluorescentes en los tubos de PCR. A diferencia de la PCR ordinaria, en la PCR en tiempo real TaqMan se añade una sonda a la reacción, esto es, un oligonucleótido de hebra sencilla complementario a un segmento de 20-60 nucleótidos dentro del molde de ADN y localizado entre los dos cebadores. Un informador fluorescente o fluoróforo (p. ej., 6-carboxifluoresceína, acrónimo: FAM, o tetraclorofluoresceína, acrónimo: TET) y un extintor (p. ej., tetrametilrodamina, acrónimo: TAMRA, o tripéptido dihidrociclopirroloindol 'extintor de agujero negro', acrónimo: BHQ) están anclados covalentemente a los extremos 5' y 3' de la sonda, respectivamente [2]. La cercanía entre el fluoróforo y el extintor anclados a la sonda inhibe la fluorescencia del fluoróforo. Durante la PCR, a medida que comienza la síntesis de ADN, la actividad exonucleasa 5' a 3' de la polimerasa Taq degrada aquella proporción de la sonda que ha hibridado con el molde. La degradación de la sonda libera el fluoróforo de la misma y se rompe cerca del extintor, mitigando de ese modo el efecto extintor y permitiendo la fluorescencia del fluoróforo. Por consiguiente, la fluorescencia detectada en el ciclador térmico de la PCR en tiempo real es directamente proporcional al fluoróforo liberado y a la cantidad de molde de ADN presente en la PCR.

Por "matriz" o "array" (por su denominación en inglés) se quiere significar una disposición de localizaciones direccionables o "direcciones" sobre un dispositivo. Las localizaciones se pueden disponer en formaciones bidimensionales, formaciones tridimensionales, u otros formatos de matrices. El número de localizaciones puede oscilar de varios a al menos cientos de miles. Lo más importante, cada localización representa un sitio de reacción totalmente independiente. Las formaciones incluyen, pero no se limitan a, formaciones de ácidos nucleicos, formaciones de proteínas y formaciones de anticuerpos. Una "matriz de ácido nucleico" hace referencia a una matriz

que contiene sondas de ácido nucleico, tales como oligonucleótidos, análogos de nucleótidos, polinucleótidos, polímeros de análogos de nucleótidos, morfolinos o porciones más grandes de genes. El ácido nucleico y/o análogo de la matriz son preferiblemente de hebra sencilla. Las matrices en donde las sondas son oligonucleótidos son referidas como "matrices de oligonucleótidos" o "chips de oligonucleótidos". Una "micromatriz", en la presente memoria también hace referencia a un "biochip" o "chip biológico", una matriz de regiones que tienen una densidad de regiones discretas de al menos aproximadamente 100/cm², y preferiblemente al menos aproximadamente 1000/cm².

"Pares de cebadores" y "sondas", dentro del significado de la invención, tendrán el significado habitual de este término que es bien conocido por el experto en la técnica de la biología molecular. En una realización preferida de la invención se entenderá que "pares de cebadores" y "sondas" son moléculas de polinucleótidos que tienen una secuencia idéntica, complementaria, homóloga, u homóloga al complemento de las regiones de un polinucleótido diana que se va a detectar o cuantificar. En otra realización más, también está incluido el uso de los análogos de nucleótidos como cebadores y/o sondas. Las tecnologías de sondas utilizadas para aplicaciones de PCR cinética o en tiempo real podrían ser, p. ej., sistemas TaqMan® obtenibles de Applied Biosystems, sondas de extensión tales como Scorpion® Primers, Dual Hybridisation Probes, Amplifluor® obtenible de Chemicon International, Inc, o Minor Groove Binders.

Se deberá entender que "sondas marcadas individualmente", dentro del significado de la invención, son sondas moleculares que comprende un polinucleótido, oligonucleótido o análogo de nucleótido y una marca, útil en la detección o cuantificación de la sonda. Las marcas preferidas son moléculas fluorescentes, moléculas luminiscentes, moléculas radiactivas, moléculas enzimáticas y/o moléculas extintoras.

Se entenderá que "sondas en matriz ordenada", dentro del significado de la invención, son una colección de sondas inmovilizadas, preferiblemente dispuestas de forma ordenada. En una realización preferida de la invención, las "sondas en matriz ordenada" se pueden identificar por su respectiva posición sobre el soporte sólido, p. ej., sobre un "chip".

Cuando se utiliza en referencia a una secuencia de ácido nucleico de hebra sencilla, el término "sustancialmente homóloga" hace referencia a cualquier sonda que pueda hibridar (esto es, es el complemento de) la secuencia de ácido nucleico de hebra sencilla en condiciones de baja restricción como se ha descrito anteriormente.

Compendio de la invención

En términos generales, la presente invención proporciona un método para evaluar el riesgo de recurrencia de un paciente con cáncer de mama positivo para receptores de estrógeno y negativo para HER2/NEU, de ganglios negativos o positivos, en particular pacientes que reciben terapia endocrina, por ejemplo cuando se tratan con tamoxifeno. El estado de los receptores de estrógeno se determina generalmente utilizando inmunohistoquímica, el estado HER2/NEU (ERBB2) se determina generalmente utilizando inmunohistoquímica e hibridación in situ con fluorescencia. No obstante, el estado de los receptores de estrógeno y el estado de HER2/NEU (ERBB2) se pueden determinar, para los fines de la invención, mediante cualquier método adecuado, p. ej. inmunohistoquímica, hibridación in situ con fluorescencia (FISH), o análisis de la expresión del ARN.

La presente invención hace referencia a un método para predecir el resultado de un cáncer de mama en un paciente con cáncer de mama positivo para receptores de estrógeno y negativo para HER2, comprendiendo dicho método:

(a) determinar en una muestra tumoral de dicho paciente los niveles de expresión de ARN de los siguientes 8 genes: UBE2C, BIRC5, DHCR7, STC2, AZGP1, RBBP8, IL6ST, y MGP;

(b) combinar matemáticamente los valores de los niveles de expresión para los genes de dicho conjunto cuyos valores se determinaron en la muestra tumoral para proporcionar una puntuación combinada, en donde dicha puntuación combinada es indicativa de una prognosis de dicho paciente.

Se describe que

BIRC5 puede ser remplazado 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 con la condición de que después de un remplazo se seleccionan 8 genes diferentes; y

UBE2C puede ser remplazado 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 con la condición de que después de un remplazo se seleccionan 8 genes diferentes; y

DHCR7 puede ser remplazado por AURKA, BIRC5, UBE2C o por cualquier otro gen que pueda remplazar a BIRC5 o UBE2C con la condición de que después de un remplazo se seleccionan 8 genes diferentes; y

STC2 puede ser remplazado 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 con la condición de que después de un replazo se seleccionan 8 genes diferentes; y

AZGP1 puede ser remplazado por PIP o EPHX2 o PLAT o SEC14L2 o SCUBE2 o PGR con la condición de que después de un replazo se seleccionan 8 genes diferentes; y

- 5 **RBBP8** puede ser remplazado por CELSR2 o PGR o STC2 o ABAT o IL6ST con la condición de que después de un replazo se seleccionan 8 genes diferentes; y

IL6ST puede ser remplazado 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 con la condición de que después de un replazo se seleccionan 8 genes diferentes; y

- 10 **MGP** puede ser remplazado por APOD o IL6ST o EGFR con la condición de que después de un replazo se seleccionan 8 genes diferentes.

De acuerdo con un aspecto de la invención se proporciona un método como se ha descrito anteriormente, en donde dicha puntuación combinada es indicativa del beneficio de la quimioterapia citotóxica.

- 15 Utilizar un método de la invención antes de que un paciente reciba terapia endocrina permite una predicción de la eficacia de la terapia endocrina.

- 20 La siguiente Tabla 2 muestra si la expresión en exceso de cada uno de los genes marcadores anteriores es indicativa de un buen resultado o de un mal resultado en un paciente que recibe terapia endocrina. El experto puede construir de ese modo una combinación matemática, esto es, un algoritmo que tenga en consideración el efecto de un gen dado. Por ejemplo una suma o una suma ponderada de genes cuya expresión en exceso es indicativa de buenos resultados produce un algoritmo en donde una puntuación de riesgo elevado es indicativa de un buen resultado. La validez del algoritmo se puede examinar analizando muestras tumorales de pacientes con un registro clínico, en donde, p. ej., la puntuación para pacientes con un buen resultado y para pacientes con un mal resultado se puede determinar por separado y comparar. El experto en la técnica, un bioestadístico, sabrá aplicar métodos matemáticos adicionales, tales como funciones discriminantes para obtener algoritmos optimizados. Los algoritmos se pueden optimizar, p. ej. para la sensibilidad o la especificidad. Los algoritmos se pueden adaptar a la plataforma analítica concreta utilizada para medir la expresión génica de genes marcadores, tales como PCR cuantitativa.

- 25 De acuerdo con un aspecto de la invención, se proporciona un método como se ha descrito anteriormente, en donde dicha terapia endocrina comprende tamoxifeno o un inhibidor de aromatasa.

- 30 De acuerdo con un aspecto de la invención, se proporciona un método como se ha descrito anteriormente, en donde se predice el riesgo de desarrollar recurrencia.

De acuerdo con un aspecto de la invención, se proporciona un método como se ha descrito anteriormente, en donde dicho nivel de expresión se determina como un nivel de expresión sin proteína.

De acuerdo con un aspecto de la invención, se proporciona un método como se ha descrito anteriormente, en donde dicho nivel de expresión se determina como un nivel de expresión de ARN.

- 35 De acuerdo con un aspecto de la invención, se proporciona un método como se ha descrito anteriormente, en donde dicho nivel de expresión se determina por medio de al menos uno de

un método basado en PCR,

un método basado en microformaciones, y

un método basado en la hibridación.

- 40 De acuerdo con un aspecto de la invención, se proporciona un método como se ha descrito anteriormente, en donde dicha determinación de niveles de expresión es en una muestra tumoral fijada con formalina incluida en parafina o en una muestra tumoral recién congelada.

- 45 De acuerdo con un aspecto de la invención, se proporciona un método como se ha descrito anteriormente, en donde 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 medio computado.

De acuerdo con un aspecto de la invención, se proporciona un método como se ha descrito anteriormente, en donde dicha etapa de combinación matemática comprende una etapa de aplicación de un algoritmo a valores representativos de un nivel de expresión de un gen dado.

- 50 De acuerdo con un aspecto de la invención, se proporciona un método como se ha descrito anteriormente, en donde dicho algoritmo es una combinación lineal de dichos valores representativos de un nivel de expresión de un

gen dado.

De acuerdo con un aspecto de la invención, se proporciona un método como se ha descrito anteriormente, en donde un valor representativo de un nivel de expresión de un gen dado se multiplica por un coeficiente.

5 De acuerdo con un aspecto de la invención, se proporciona un método como se ha descrito anteriormente, en donde se determinan uno, dos o más umbrales para dicha puntuación combinada y se discriminan en grupos de riesgo elevado y bajo, riesgo elevado, intermedio y bajo, o más grupos de riesgo aplicando el umbral sobre la puntuación combinada.

10 De acuerdo con un aspecto de la invención, se proporciona un método como se ha descrito anteriormente, en donde una puntuación combinada elevada es indicativa del beneficio de una terapia más agresiva, p. ej. quimioterapia citotóxica. El experto en la técnica entiende que una "puntuación elevada" en este aspecto, se refiere a un valor de referencia o valor de corte. El experto entiende adicionalmente que dependiendo del algoritmo concreto utilizado para obtener la puntuación combinada, también una puntuación "baja" por debajo de un valor de corte o de referencia puede ser indicativa del beneficio de una terapia más agresiva, p. ej. quimioterapia citotóxica. Este es el caso cuando los genes tienen una correlación positiva con un factor de alto riesgo de metástasis en el algoritmo con un coeficiente positivo, de manera que una puntuación global alta indica una elevada expresión de genes que tienen una correlación positiva con un alto riesgo.

De acuerdo con un aspecto de la invención, se proporciona un método como se ha descrito anteriormente, en donde una información referente al estado ganglionar del paciente es procesada en la etapa de combinación matemática de los valores de los niveles de expresión para los genes para proporcionar una puntuación combinada.

20 De acuerdo con un aspecto de la invención, se proporciona un método como se ha descrito anteriormente, en donde dicha información referente al estado ganglionar es un valor numérico ≤ 0 si dicho estado ganglionar es negativo y dicha información es un valor numérico > 0 si dicho estado ganglionar es positivo o desconocido. En realizaciones ilustrativas de la invención, se asigna a un estado ganglionar negativo el valor de 0, a un estado ganglionar desconocido se le asigna el valor 0,5 y a un estado ganglionar positivo se le asigna el valor 1. Se pueden seleccionar otros valores para reflejar una ponderación diferente del estado ganglionar en un algoritmo.

La descripción se refiere adicionalmente a un kit para realizar un método como se ha descrito anteriormente, comprendiendo dicho kit un conjunto de oligonucleótidos susceptibles de unirse específicamente a secuencias o a secuencias de fragmentos de los genes de una combinación de genes, en donde

30 (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 10 genes BIRC5, AURKA, PVALB, NMU, STC2, RBBP8, PTGER3, CXCL12, CDH1, y PIP; o

(iii) dicha combinación comprende al menos los 9 genes BIRC5, DHCR7, RACGAP1, PVALB, STC2, IL6ST, PTGER3, CXCL12, y ABAT; o

35 (iv) dicha combinación comprende al menos los 9 genes DHCR7, RACGAP1, NMU, AZGP1, RBBP8, IL6ST, y MGP.

La invención se refiere adicionalmente al uso de un kit para realizar un método de cualquiera de las reivindicaciones 1 a 17, comprendiendo dicho kit un conjunto de oligonucleótidos capaz de unirse específicamente a secuencias o a secuencias de fragmentos de los genes en una combinación de genes, en donde

40 (i) dicha combinación comprende al menos los 8 genes UBE2C, BIRC5, DHCR7, STC2, AZGP1, RBBP8, IL6ST, y MGP. También se describen tales usos en donde

(ii) dicha combinación comprende al menos los 10 genes BIRC5, AURKA, PVALB, NMU, STC2, RBBP8, PTGER3, CXCL12, CDH1, y PIP; o

(iii) dicha combinación comprende al menos los 9 genes BIRC5, DHCR7, RACGAP1, PVALB, STC2, IL6ST, PTGER3, CXCL12, y ABAT; o

45 (iv) dicha combinación comprende al menos los 9 genes DHCR7, RACGAP1, NMU, AZGP1, RBBP8, IL6ST, y MGP; 19. Un producto informático capaz de procesar valores representativos de un nivel de expresión de los genes AKR1C3, MAP4 y SPP1 combinando matemáticamente dichos valores para proporcionar una puntuación combinada, en donde dicha puntuación combinada es indicativa del beneficio de la quimioterapia citotóxica de dicho paciente.

50 La invención describe adicionalmente un programa informático capaz de procesar valores representativos de un nivel de expresión de una combinación de genes combinando matemáticamente dichos valores para proporcionar una puntuación combinada, en donde dicha puntuación combinada es indicativa de la eficacia o el beneficio de la terapia endocrina de dicho paciente, de acuerdo con los métodos anteriores.

Dicho programa informático se puede almacenar en un soporte de datos o implementar en un sistema de diagnóstico capaz de dar salida a valores representativos de un nivel de expresión de un gen dado, tal como un sistema de PCR en tiempo real.

5 Si el programa informático se almacena en un soporte de datos o se ejecuta en un ordenador, el personal puede introducir los valores de expresión obtenidos para el nivel de expresión de los respectivos genes. El programa informático puede aplicar a continuación un algoritmo para producir una puntuación combinada indicativa del beneficio de la terapia citotóxica para un paciente dado.

10 Los métodos de la presente invención tienen la ventaja de proporcionar una predicción fiable del resultado de una enfermedad basándose solamente en el uso de un pequeño número de genes. Se ha descubierto que los métodos de la presente invención son especialmente adecuados para analizar la respuesta al tratamiento endocrino, p. ej. con tamoxifeno, de pacientes con tumores clasificados como ESR1 positivos y ERBB2 negativos.

Descripción detallada de la invención

La invención se explica junto con realizaciones ilustrativas y las figuras adjuntas:

15 La Figura 1 muestra un Gráfico de Forrest de la razón de riesgos instantáneos corregida con un intervalo de confianza de 95% de la puntuación T5 en la cohorte combinada, así como los grupos de tratamiento individuales de los estudios ABCSG06 y 08, utilizando la metástasis distante como criterio de valoración.

La Figura 2 muestra un Análisis de Kaplan Meier de pacientes ER+, HER-, N0-3 de las cohortes combinadas ABCSG06 y 08, estratificadas en alto o bajo riesgo de acuerdo con el valor de la Puntuación T5.

20 En la presente memoria se describen combinaciones únicas de genes marcadores que se pueden combinar en un algoritmo para el nuevo ensayo predictivo presentado aquí. Técnicamente, el método de la invención se puede poner en práctica utilizando dos tecnologías: 1.) Aislamiento del ARN total a partir de tejido tumoral fresco o fijado y 2.) RT-PCR Cinética de los ácidos nucleicos aislados. Alternativamente, se contempla la medición de niveles de expresión utilizando tecnologías alternativas, p. ej. por medio de una micromatriz o por medio de la medición a nivel de proteína.

25 Los métodos de la invención se basan en la determinación cuantitativa de especies de ARN aisladas del tumor con el fin de obtener valores de expresión y el subsiguiente análisis bioinformático de dichos valores de expresión determinados. Se podrían aislar especies de ARN de cualquier tipo de muestra tumoral, p. ej. muestras de biopsia, muestras de frotis, material tumoral resecado, tejido tumoral recién congelado o tejido tumoral fijado con formalina e incluido en parafina. En primer lugar, se determinan los niveles de ARN de los genes que codifican combinaciones específicas de los genes UBE2C, BIRC5, DHCR7, RACGAP1, AURKA, PVALB, NMU, STC2, AZGP1, RBBP8, 30 IL6ST, MGP, PTGER3, CXCL12, ABAT, CDH1, y PIP o combinaciones específicas de los mismos, según se ha indicado. Basándose en estos valores de expresión se calcula una puntuación pronóstica por medio de una combinación matemática, p. ej. de acuerdo con las fórmulas T5 T1, T4, o T5b (véase más abajo). Un valor de puntuación elevado indica un riesgo elevado de desarrollar metástasis distante, un valor de puntuación bajo indica un bajo riesgo de metástasis distante. Por consiguiente, una puntuación elevada también indica que el paciente es un paciente de alto riesgo que se beneficiará de una terapia más agresiva, p. ej. quimioterapia citotóxica.

40 Los presentes ejemplos se basan en la identificación de genes pronóstico utilizando tumores de pacientes tratados de manera homogénea en el entorno coadyuvante con tamoxifeno. Además, la identificación de genes relevantes se ha restringido a tumores clasificados como ESR1 positivos y ERBB2 negativos basándose en los niveles de expresión de ARN. Además, los genes que permiten la separación del riesgo intermedio, p. ej. tumores de grado 2 fueron considerados para el desarrollo del algoritmo. Finalmente, se realizó una transferencia de plataforma desde formaciones Affymetrix HG_U133a a PCR en tiempo real cuantitativa, así como una transferencia del tipo de muestra de tejido recién congelado a tejido FFPE para garantizar un funcionamiento robusto del algoritmo, independiente del tipo de plataforma y tejido. Como resultado, la determinación del nivel de expresión de las 45 especies de ARN del tumor primario y el posterior análisis complejo y multivariante descrito anteriormente proporcionan un método superior para la predicción de la verosimilitud de recurrencia de la enfermedad en pacientes diagnosticados de cáncer de mama temprano con ganglios negativos o positivos, cuando se tratan con tamoxifeno solamente en el entorno coadyuvante. De este modo, el ensayo se basa en menos genes que los de los competidores, pero proporciona información superior en cuanto a la alta sensibilidad y al valor predictivo negativo, en particular para tumores que se consideraba que mostraban un riesgo de recurrencia intermedio basándose en 50 factores clínicos convencionales.

El ARN total se extrajo con un método de aislamiento basado en cuentas de sílice y completamente automático, de Siemens, para el ARN de una sección de tejido FFPE completa de 10 μ m en un robot de manipulación de líquido Hamilton MICROLAB STARlet (17). El robot, los tampones y los agentes químicos eran parte de un Siemens VERSANT® kPCR Molecular System (Siemens Healthcare Diagnostics, Tarrytown, NY; no disponible en el mercado en los EEUU). En resumen, se añadieron 150 μ l de tampón FFPE (Tampón FFPE, reactivo de investigación, Siemens Healthcare Diagnostics) a cada sección y se incubaron 30 minutos a 80°C con sacudimiento para fundir la parafina. Después de enfriar, se añadió proteinasa K y se incubó durante 30 minutos a 65°C. Después de la lisis, se

retiraron los desechos de tejido residual del fluido de lisis por medio de una etapa de incubación de 15 minutos a 65°C con 40 µl de cuentas de óxido de hierro recubiertas con sílice. Las cuentas con los desechos tisulares unidos a la superficie se separaron con un imán y los productos lisados se transfirieron a una placa con pocillos (96 pocillos) profunda de 2 ml convencional. Allí, el ARN total y el ADN se unieron a 40 µl de cuentas sin utilizar y se incubaron a la temperatura ambiente. Se produjeron condiciones caotrópicas mediante la adición de 600 µl de tampón de lisis. A continuación, las cuentas se separaron magnéticamente y se descartaron los sobrenadantes. Después de eso, los ácidos nucleicos unidos a la superficie se lavaron tres veces seguido de magnetización, aspiración y eliminación de los sobrenadantes. Después de eso, los ácidos nucleicos se hicieron eluir mediante incubación de las cuentas con 100 µl de tampón de elución durante 10 minutos a 70°C con sacudimiento. Finalmente, las cuentas se separaron y el sobrenadante se incubó con 12 µl de DNase I Mix (2 µL de ADNasa I (libre de ARNasa); 10 µl de 10x tampón de ADNasa I; Ambion/Applied Biosystems, Darmstadt, Alemania) para retirar el ADN contaminante. Después de incubar durante 30 minutos a 37°C, se tomaron alícuotas de ARN total libre de ADN y se almacenaron a -80°C o se utilizaron directamente para el análisis de expresión del ARNm por medio de PCR cinética con transcripción inversa (RTkPCR). Todas las muestras se analizaron con una RT-kPCR de una etapa para determinar la expresión génica de hasta tres genes de referencia, (RPL37A, CALM2, OAZ1) y hasta 16 genes diana en un ABI PRISM® 7900HT (Applied Biosystems, Darmstadt, Alemania. Se utilizó SuperScript® III Platinum® One-Step Quantitative RT-PCR System con ROX (6-carboxi-X-rhodamina) (Invitrogen, Karlsruhe, Alemania) de acuerdo con las instrucciones del fabricante. Las sondas y los cebadores respectivos se muestran en la tabla 1. Las condiciones de la PCR fueron las siguientes: 30 minutos a 50°C, 2 minutos a 95°C seguido de 40 ciclos de 15 segundos a 95°C y 30 segundos a 60°C. Todos los análisis de PCR se realizaron por triplicado. Como marcador sustituto para el rendimiento de ARN, se utilizó el valor del ciclo umbral (Ct) del gen constitutivo, RPL37A como se describe en otro lugar (17). Los niveles de expresión génica relativos de los genes diana se calcularon mediante el método delta-Ct utilizando la fórmula:

$$20 - (Ct \text{ (diana)} - \text{media } (Ct \text{ (genes de referencia)}))$$

La transferencia de plataforma de formaciones Affymetrix HG_U133a (tejido recién congelado) a PCR en tiempo real cuantitativa (tejido FFPE) se calculó como sigue. Se midió material de 158 pacientes utilizando ambas plataformas para proporcionar muestras emparejadas. Se calcularon los valores Delta-Ct a partir de los datos de la PCR. Se calcularon los Log2 de las expresiones de los datos de Affymetrix aplicando un límite inferior (ajustando todos los valores por debajo del límite inferior al límite inferior) y a continuación calculando el logaritmo de base 2. La aplicación de un límite inferior reduce el efecto de aumento de ruido de la medición relativa para los genes poco expresados/muestras; se utilizó un límite inferior de 20, los límites inferiores entre 0,1 y 200 también funcionan bien. Se seleccionó un conjunto de sondas HG_U133a para cada gen medido mediante PCR maximizando el coeficiente de correlación de Pearson entre el valor delta-Ct (de la PCR) y el log2 de la expresión (de Affymetrix). Otras medidas de correlación también funcionarán bien, p. ej. el coeficiente de correlación de Spearman. En la mayor parte de los casos el conjunto de sondas mejor correlacionado pertenecía al gen pretendido, para el resto de los casos se eliminó el gen de la PCR para su procesamiento adicional. Aquellos genes que mostraban una mala correlación entre las plataformas también fueron eliminados, cuando se utilizó un umbral en el coeficiente de correlación de Pearson de 0,7 (valores entre 0,5 y 0,8) también funcionó bien. La transformación de plataforma finalizó calculando las transformadas Z sin supervisión para ambas plataformas y combinándolas; a continuación un único valor de PCR-delta-Ct se transforma a la escala de Affymetrix mediante las siguientes etapas: (i) aplicar la transformación lineal afín cuando los coeficientes se determinaron mediante la transformada z de los datos de la PCR, (ii) aplicar la transformación lineal afín inversa cuando los coeficientes se determinaron mediante la transformada z de los datos de Affymetrix, (iii) invertir el log2, esto es, calcular el exponencial con respecto a la base 2. Las alternativas a las transformadas z dobles son la regresión lineal o de orden superior, la regresión robusta o métodos basados en un componente principal, que también funcionarán bien.

Las secuencias de los cebadores y las sondas son las siguientes:

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

gen	sonda	Seq ID	cebador directo	Seq ID	cebador inverso	Seq ID
ABAT	TCGCCCTAAGAGGCTCTCCCTC	1	GGCAACTTGAGGCTGAGCTTTTG	2	GGCAGCTCACAAGTGGTGTGA	3
ADRA2A	TTGTCTTTCCCCCTCCGTGC	4	CCCCAAGAGCTGTTAGGATCAA	5	TCAATCACATGATCTCAACCAGAA	6
APOD	CATCAGCTCTCAACTCCTGGTTAAACA	7	ACTCACTAATGGAAAACGGAAAGATC	8	TCACCTTCGATTTGATTCACAGTT	9
ASPH	TGGGAGGAAGCAAGGTGCTCATC	10	TGTGCCAACGAGACCAAGAC	11	TCGTGCTCAAAGGAGTCATCA	12
AURKA	CCGTACGCTGTGCTAGGCAT	13	AATCTGGAGGCAAGGTTCTGA	14	TCTGGATTTGCCCTCCTCTGTGAA	15
BIRC5	AGCCAGATGACGACCCCATAGAGGAACA	16	CCCAGTGTCTTCTGCTTCAAG	17	CAACCGACGAATGCTTTTTT	18
CELSR2	ACTGACTTTCCTTCTGGAGCAGGTGGC	19	TCCAAGCATGATTCACAGACTTGT	20	TGCCACAGCCCTCTTTTTCT	21
CHPT1	CCACGGCCACCGAAGAGGCAC	22	CGCTCGTGTCTATCTCCTACT	23	CCCAGTGCACATAAAGGTATGTC	24
CXCL12	CCACAGCAGGTTTTCAGGTCC	25	GCCACATACCCCTCCTGAA	26	TCACCTTGCCAACAGTCTGAT	27
CYBRD1	AGGCATCGCCATCATCGTC	28	GTCACCGGCTTCGTCTCA	29	CAGGTCCACGGCAGTCTGT	30
DCN	TCTTTTCAGCAACCCGGTCCA	31	AAGGCTCTTATTCGGGTGTA	32	TGGATGGCTGTATCTCCCAGTA	33
DHCR7	TGAGCGCCACCCCTCTCGA	34	GGGCTCTGCTTCCCGATT	35	AGTCATAGGGCAAGCAGAAAATTC	36
E2F8	CAGGATACCTAATCCCTCTCACGCAG	37	AAATGCTCCCGCAACCTTGTTT	38	CTGCCCCAGGGATGAG	39
EPHX2	TGAAGCGGAGGACTTTTGTAAA	40	CGATGAGAGTGTTTTATCCATGCA	41	GCTGAGGCTGGCTCTCT	42
ESR1	ATGCCCTTTTGCCGATGCA	43	GCCAAATGTGTTTGATGGATTAA	44	GACAAAACCCGAGTCCACATCAGTAATAG	45
GJA1	TGCACAGCCTTTTGTATTTCCCCGAT	46	CGGGAAGCACCATCTCTAACTC	47	TTCATGTCCAGCAGCTAGTTTTT	48
HSPA2	CAAGTCAGCAAAACACGGCAAAA	49	CATGCACGAACTAATCAAAAATGC	50	ACATTAATCGAGGTTCTCTTTAATGC	51
IL6ST	CAAGTCCACCTTCCAAAGGACCT	52	CCCTGAATCCATAAAGGCATACC	53	CAGCTTCGTTTTTCCCTACTTTTT	54
INPP4B	TCCGAGCGCTGGATTGCATGAG	55	GCACAGTTACACAAGGACTTCTTT	56	TCCTATGCGGCACTCTCTCT	57
MAPT	AGACTATTTGCACACTGCCGCCT	58	GTGGCTCAAAGGATAATATCAAACAC	59	ACCTTGCTCAGGTCAACTGGTT	60
MGP	CCTTCAATATCCCTCAGCAGAGATGG	61	CCTTCAATTAACAGGAAATGCAA	62	ATTGAGCTCGTGGACAGGCTTA	63
NEK2	TCCTGAACAATGAATCGCATGTCCATCAA	64	ATTTGTTGGCACACCTTATACATGT	65	AAGCAGCCCAATGACCAGATA	66
PCNA	AAATACTAAAATGCCCGGCAATGA	67	GGCGTGAACCTCACCCAGTA	68	CTTCGGCCCTTAGTGAATGATAATC	69

PGR	TTGATAGAAACGCTGTGAGCTCGA	70	AGCTCATCAAGGCAATGGTTT	71	ACAAGATCATGCAAGTTATCAAGAAGTT	72
PIP	TGCATGGTGGTTAAAACCTTACCTCA	73	TGCTTGCAGTTCAAACAGAATTG	74	CACCTTGTAGAGGGANGCTGCTA	75
PLAT	CAGAAAGTGGCCATGCCACCCCTG	76	TGGAAAGACATGAATGCACACTA	77	GGAGGTTGGGCTTGTAGCTGAA	78
PRSS16	CACTGCCGGTCACCCACACCA	79	CTGAGGAGCACAGAACCTCAACT	80	CGAACTCGGTACATGTCTGATACAA	81
PTGER3	TCGGTCTGTGGTCTCCGCTCC	82	CTGANTGAAGATCATPTTCAACATCA	83	GACGGCCATTACAGCTTATGG	84
PTPRT	TTGGTCTTGACACCCCTCACA	85	GAGTTGTGGCCTCFACCAATTGC	86	GAGCGGGAACCTTGGGATAG	87
RACGAP1	ACTGAGAATCTCCACCCGGGCA	88	TCGCCAACTGGATAAATTGGA	89	GAATGTGCGGAAHCTGTTTGAG	90
RBP8	ACCGATTCGGTACATTCACCCAAC	91	AGAAATGGCTTCCCTGCTCAAG	92	AAAACCAACTTCCCAAAAATCTCT	93
SCUBE2	CTAGAGGGTTCAGAGTCCCATACGTGACATA	94	TGTGGATTCAGTTCAAAGTCCAATG	95	CCATCTCGAACTATGTCTTCAATGAGT	96
SEC14L2	TGGGAGGCATGCAACGCGTG	97	AGGTCTTACTAAGCAGTCCCATCTCT	98	CGACCGGCACCCTGAACTC	99
SOLE	TATGCGTCTCCAAAAGAAACACCTCG	100	GCAAAGCTTCCCTTCCCTTCA	101	CCTTTAGCAGTTTTTCTCCATAGTTTTATATC	102
TFAP2B	CAACACCACCTAACAGGCACACGTC	103	GGCATGGACAAGATGTTCTTGA	104	CCTCCTTGTCCGCCAGTTTTACT	105
TOP2A	CAGATCAGGACCAAGATGGTCCCCACAT	106	CATTAAGACGCTTCGTTATGG	107	CCAGTTGTGATGGATAAAATTAATCAG	108
TRIM29	TGCTGTCTCACTACCGGCCAATCTAG	109	TGGAANTCTGGCAAGCAGACT	110	CAATCCCGTTGCCCTTGTG	111
UBE2C	TGAACACACATGTGCCGAGCTCTG	112	CTTCTAGGAGAACCCCAACATTTGATAGT	113	GTTTTCTGCAGGTACTTCTTAAAAGCT	114
WNT5A	TATTCACATCCCTCAGTTGCAGTGAATTG	115	CTGTGGCTCTTAATTTATTTGCATAATG	116	TTAGTGTCTTTTGGCTTCAAGATCTT	117
STC2	TCTCACCTTGACCTCAGCCCAAG	118	ACATTTGACAAAATTTCCCTTAGGATT	119	CCAGGACGCAGCTTTACCAA	120
AZGP1	CACCAGCCACCCAGGCCCCAG	121	TCCTGGACCCGCAAGATC	122	TAGGCCAGGCACCTTCAGTCTC	123
CALM2	TCGCGTCTCGGAAACCGGTAGC	124	GAGCGAGCTGAGTGGTGTG	125	AGTCAGTTGGTCCAGCCATGCT	126
CDH1	CCTGCCAATCCCGATGAAATTTGGAAT	127	TGAGTGTCCCCCGGTATCTTC	128	TCAGCCGCTTTCAGATTTTCA	129
NMU	ACCTGTGACCTTCTTCCATTCGGT	130	AGAAATTTGGCTTCCCTGCTCAAG	131	AAAACCAACTTCCCAAAAATCTCT	132
OAZ1	TGCTTCCACAAGAACCCGCGAGGA	133	CGAGCCGACCATGTCTTTCAT	134	AAGCCCCAAAAGCTGAAAGTT	135
PVALB	AAGTTCTTCCAAATGGTCCGGCC	136	CCGACTCCTTCGACCCAAA	137	CATCATCCGCACCTCTTTTTCTTC	138
RPL37A	TGGCTGGCGGTGCCCTGGGA	139	TGTGGTCTCTGCAATGAAGACA	140	GTGACAGCGGAAAGTGGTATTTGTAC	141

La Tabla 2, de más abajo, enumera los genes utilizados en los métodos de la invención y en las realizaciones concretas T5, T1, T4, y T5b. La Tabla 2 también muestra si la expresión en exceso de un gen dado es indicativa de un buen o mal resultado bajo terapia con Tamoxifeno. La Tabla 2 enumera la función del gen, la localización del compartimento dentro de la célula y los procesos celulares en los que está implicado.

5 Tabla 2 Lista de genes de los algoritmos T5, T1, T4, y T5b:

Gen	Nombre	Expresión Elevada	Función	Componente	Proceso
UBE2C	Enzima E2C conjugada con ubiquitina	Mal resultado	Unión a ATP	citósol	división celular
BIRC5	IAP baculoviral que contiene 5 repeticiones	Mal resultado	Unión a GTPasa Ran	citósol	ciclo celular
DHCR7	7-Deshidrocolesterol reductasa	Mal resultado	Actividad 7-deshidrocolesterol reductasa	membrana retículo endoplasmático	regulación de la proliferación celular
RACGAP1	Proteína 1 activadora de GTPasa Rac	Mal resultado	Actividad activadora de GTPasa	citoplasma	ciclo celular
AURKA	Auroraquinasa A	Mal resultado	Unión a ATP	centrosoma	ciclo celular mitótico
PVALB	Parvalbúmina	Mal resultado	Unión ión calcio		
NMU	Neuromedina U	Mal resultado	Unión a receptor	región extracelular	transducción de señal
STC2	Estanniocalcina 2	Buen resultado	Actividad hormonal	región extracelular	transducción de señal ligada a receptor superficie celular
AZGP1	Alfa-2-glicoproteína 1	Buen resultado	Actividad transportadora proteína transmembrana	región extracelular	regulación negativa de la proliferación celular
RBBP8	Proteína 8 de unión a retinoblastoma	Buen resultado	Unión a proteínas	núcleo	punto de control del ciclo celular
IL6ST	Transductor de señal de interleuquina 6	Buen resultado	Actividad receptora	región extracelular	transducción de señal
MGP	Proteína Gla de la matriz	Buen resultado	Constituyente extracelular de la matriz	región extracelular	diferenciación celular
PTGER3	Receptor 3 de prostaglandina E	Buen resultado	Actividad receptora dependiente del ligando	membrana plasmática	transducción de señal

ES 2 525 382 T3

Gen	Nombre	Expresión Elevada	Función	Componente	Proceso
CXCL12	Ligando 12 de quimioquina (motivo C-XC)	Buen resultado	Actividad quimioquina	Región extracelular	transducción de señal
ABAT	4-Aminobutirato aminotransferasa	Buen resultado	Actividad transferasa	mitocondria	Proceso catabólico del ácido gamma-aminobutírico
CDH1	Cadherina 1	Buen resultado	Unión a molécula de adherencia celular	Membrana plasmática	Adherencia a células homofílicas
PIP	Proteína inducida por prolactina	Buen resultado	Unión a actina	Región extracelular	
CALM2					Gen de Referencia
OAZ1					Gen de Referencia
RPL37A					Gen de Referencia

La Tabla 3, de más abajo, muestra las combinaciones de genes utilizadas para cada algoritmo.

Tabla 3: Combinación de genes para los respectivos algoritmos:

Gen	Ref. Algo T1	Ref. Algo T4	Inv. Algo T5	Ref. Algo T5b
UBE2C			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		

ES 2 525 382 T3

Gen	Ref. Algo T1	Ref. Algo T4	Inv. Algo T5	Ref. Algo T5b
CXCL12	X	X		
ABAT		X		
CDH1	X			
PIP	X			
Ref. = Ejemplo de referencia; Inv. = de acuerdo con la invención				

La Tabla 4, de más abajo, muestra el ID del conjunto de sondas "probeset" de Affy y el mapeo con ID de diseño Taqman de los genes marcadores de la presente invención.

Tabla 4: Símbolo del Gen, ID grupo de sondas Affy y mapeo con ID diseño TaqMan:

Gen	ID Diseño	ID Grupo de sondas
UBE2C	R65	202954_at
BIRC5	SC089	202095_s_at
DHCR7	CAGMC334	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

La Tabla 5, de más abajo, muestra los nombres completos, Entrez GeneID, número de acceso Gene bank y localización cromosómica de los genes marcadores de la presente invención

Símbolo oficial	Nombre completo oficial	Entrez GeneID	Núm. de Acceso	Localización
UBE2C	Enzima conjugada con ubiquitina E2C	11065	U73379	20q13. 12
BIRC5	IAP baculoviral que contiene 5 repeticiones	332	U75285	17q25
DHCR7	7-Deshidrocolesterol reductasa	1717	AF034544	11q13. 4
STC2	Estaniocalcina 2	8614	AB012664	5q35.2
RBBP8	Proteína de unión a retinoblastoma 8	5932	AF043431	18q11. 2
IL6ST	Transductor de señal de interlequina 6	3572	M57230	5q11
MGP	Proteína Gla de la matriz	4256	M58549	12p12. 3
AZGP1	Alfa-2-glicoproteína 1, unión a cinc	563	BC005306	11q22. 1
RACGAP1	Proteína de activación de GTPasa Rac 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 prostaglandina E 3 (subtipo EP3)	5733	X83863	1p31.2
CXCL12	Quimioquina (motivo C-X-C) ligando 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

Ejemplo algoritmo T5:

- 5 El Algoritmo T5 es un comité de cuatro miembros donde cada miembro es una combinación lineal de dos genes. Las fórmulas matemáticas para T5 se muestran más abajo; la notación es la misma que para T1. T5 se puede calcular a partir de los datos de expresión génica solamente.

$$\text{riesgoMiembro1} = 0,434039 [0,301..0,567] * (0,939 * \text{BIRC5} - 3,831) \\ - 0,491845 [-0,714..-0,270] * (0,707 * \text{RBBP8} - 0,934)$$

10

$$\text{riesgoMiembro2} = 0,488785 [0,302..0,675] * (0,794 * \text{UBE2C} - 1,416) \\ - 0,374702 [-0,570..-0,179] * (0,814 * \text{IL6ST} - 5,034)$$

riesgoMiembro3 = -0,39169 [-0,541..-0,242] * (0,674 * AZGP1 -0,777)
 +0,44229 [0,256..0,628] * (0,891 * DHCR7 -4,378)

5 riesgoMiembro4 = -0,377752 [-0,543..-0,212] * (0,485 * MGP +4,330)
 -0,177669 [-0,267..-0,088] * (0,826 * STC2 -3,630)

riesgo = riesgoMiembro1 + riesgoMiembro2 + riesgoMiembro3 + riesgoMiembro4

10 Los coeficientes a la izquierda de cada línea se calcularon en forma de coeficientes de regresión para los riesgos
 proporcionales de COX, los números entre corchetes indican límites de confianza de 95% para estos coeficientes.
 En otras palabras, en lugar de multiplicar el término (0,939 * BIRC5 -3,831) por 0,434039, éste se puede multiplicar
 por cualquier coeficiente entre 0,301 y 0,567 y todavía tener un resultado predictivo en los límites de confianza de
 15 95%. Los términos entre paréntesis a la derecha de cada línea indican una transferencia de plataforma de PCR a
 Affymetrix: Las variables PVALB, CDH1, ... indican expresiones basadas en la PCR normalizadas por los genes de
 referencia (valores delta Ct), el término completo dentro del paréntesis corresponde al logaritmo (base 2) de los
 valores de expresión de la micromatriz de Affymetrix de los correspondientes grupos de sondas.

20 El funcionamiento del algoritmo T5 se sometió a ensayo en pacientes tratados con Tamoxifeno o Anastrozol con no
 más de 3 ganglios linfáticos positivos y tumores ER+, HER2-, que participaron en las pruebas clínicas aleatorizadas
 ABCSG06 (n = 332) o ABCSG08(n = 1244). Como se muestra en la figura 1, el análisis de regresión de Cox revela,
 que la puntuación de T5 tiene una asociación significativa con el desarrollo de metástasis distante en todas las
 cohortes sometidas a ensayo.

25 Se realizó el análisis de Kaplan Meier, después de clasificar a los pacientes de las cohortes ABCSG combinadas
 utilizando un corte predefinido para la puntuación de T5. Los pacientes con un bajo riesgo de desarrollar una
 metástasis distante tuvieron una puntuación de T5 ≤ -9,3, mientras los pacientes con un riesgo elevado de
 desarrollar metástasis distante tuvieron una puntuación de T5 por encima de -9.3. Como se muestra en la figura 2,
 se observa una separación muy significativa de ambos grupos de riesgo.

30 De manera importante, la puntuación de T5 se evaluó y se comparó frente a "Adjuvant! Online", una herramienta en
 Internet para ayudar a la selección de terapia basada en la entrada de parámetros clínicos tales como el tamaño del
 tumor, el grado del tumor y el estadio ganglionar. Cuando la puntuación de T5 se sometió a ensayo mediante la
 regresión de Cox bivariante frente a la puntuación del Adjuvant!Online Relapse Risk, ambas puntuaciones
 continuaron en asociación significativa con el desarrollo de metástasis distante. La regresión de Cox bivariante que
 utiliza datos dicotomizados, que se estratificaron de acuerdo con T5 (corte -9.3) con respecto a Adjuvant!Online
 (corte 8), de nuevo proporcionó asociaciones altamente significativas e independientes con el tiempo de metástasis
 como criterio de valoración clínico.

35 Tabla 6: Regresión de Cox bivariante von T5 und Adjuvant!Online

Variable	Razón de Riesgo	CI 95%*	P
Adjuvant!Online	2,36	1,58-3,54	<0,0001
Firma de expresión génica (grupo de riesgo)	2,62	1,71-4,01	<0,0001
Adjuvant!Online (puntuación)	1,04	1,02-1,06	<0,0001
Firma de expresión génica (grupo de riesgo)	1,35	1,21-1,49	<0,0001

Con HR = Razón de Riesgo, CI 95% = Intervalo de Confianza de 95%, p = valor de P.

Las curvas de Kaplan Meyer ilustrativas se muestran en las Figs. 1 en donde Elevado = Grupo de Riesgo Elevado,
 Bajo = Grupo de Riesgo Bajo de acuerdo con un corte previamente definido

Un valor elevado de la puntuación de T5 indica un aumento del riesgo de aparición de metástasis distante en un

período de tiempo dado.

Se ha demostrado que este es el caso para pacientes que han sido tratados con tamoxifeno y también para pacientes que han sido tratados con inhibidores de aromatasa.

Ejemplo de algoritmo T1:

5 El algoritmo T1 es un comité de tres miembros donde cada miembro es una combinación lineal de hasta cuatro variables. En general, las variables pueden ser expresiones de genes o variables clínicas. En T1 la única variable no génica es el estado ganglionar codificado como 0, si el paciente tiene ganglios linfáticos negativos y 1, si el paciente tiene ganglios linfáticos positivos. Las fórmulas matemáticas para T1 se muestran más abajo.

$$\text{riesgoMiembro1} = +0,193935 [0,108..0,280] * (0,792 * \text{PVALB} -2,189)$$

$$10 \quad -0,240252 [-0,400..-0,080] * (0,859 * \text{CDH1} -2,900)$$

$$-0,270069 [-0,385..-0,155] * (0,821 * \text{STC2} -3,529)$$

$$+1,2053 [0,534..1,877] * \text{estadoGanglionar}$$

$$\text{riesgoMiembro2} = -0,25051 [-0,437..-0,064] * (0,558 * \text{CXCL12} +0,324)$$

$$15 \quad -0,421992 [-0,687..-0,157] * (0,715 * \text{RBBP8} -1,063)$$

$$+0,148497 [0,029..0,268] * (1,823 * \text{NMU} -12,563)$$

$$+0,293563 [0,108..0,479] * (0,989 * \text{BIRC5} -4,536)$$

$$\text{riesgoMiembro3} = +0,308391 [0,074..0,543] * (0,812 * \text{AURKA} -2,656)$$

$$20 \quad -0,225358 [-0,395..-0,055] * (0,637 * \text{PTGER3} + 0,492)$$

$$-0,116312 [-0,202..-0,031] * (0,724 * \text{PIP} +0,985)$$

riesgo = +riesgoMiembro1 + riesgoMiembro2 + riesgoMiembro3

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

30 **Ejemplo algoritmo T4:**

El Algoritmo T4 es una combinación lineal de motivos. Los 10 genes principales de varios análisis de grupos de datos de Affymetrix y datos de PCR se agruparon en motivos. Los genes no pertenecientes a una agrupación se utilizaron como motivos de un solo gen. Los coeficientes de regresión de riesgos proporcionales de COX se encontraron en un análisis de múltiples variantes.

35 En general, los motivos pueden ser expresiones de un solo gen o expresiones génicas medias de genes correlacionados. Las fórmulas matemáticas para T4 se muestran a continuación.

$$\text{prolif} = ((0,84 [0,697..0,977] * \text{RACGAP1} -2,174) + (0,85 [0,713..0,988]$$

$$* \text{DHCR7} -3,808) + (0,94 [0,786..1,089] * \text{BIRC5} -3,734)) / 3$$

$$\text{motiv2} = ((0,83 [0,693..0,96] * \text{IL6ST} -5,295) + (1,11 [0,930..1,288] *$$

$$40 \quad \text{ABAT} -7,019) + (0,84 [0,701..0,972] * \text{STC2} -3,857)) / 3$$

$$\text{ptger3} = (\text{PTGER3} * 0,57 [0,475..0,659] + 1,436)$$

$$\text{cxc112} = (\text{CXCL12} * 0,53 [0,446..0,618] + 0,847)$$

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

5 Los factores y las compensaciones para cada gen indican una transferencia de plataforma de PCR a Affymetrix: Las variables RACGAP1, DHCR7, ... indican expresiones basadas en PCR normalizadas por CALM2 y PPIA, el término completo entre paréntesis corresponde al logaritmo (base 2) de los valores de expresión de microformaciones de Affymetrix de los conjuntos de sondas correspondientes.

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

Como el algoritmo funcionó incluso mejor combinado con una variable clínica, se añadió el estado ganglionar. En T4 el estado ganglionar se codifica como 0, si el paciente tiene ganglios linfáticos negativos y como 1, si el paciente tiene ganglios linfáticos positivos. Con esto, el algoritmo T4 es:

$$10 \text{ riesgo} = -0,32 [-0,510..-0,137] * \text{motiv2} \\ + 0,65 [0,411..0,886] * \text{prolif} \\ - 0,24 [-0,398..-0,08] * \text{ptger3} \\ - 0,05 [-0,225..0,131] * \text{cxcl12} \\ + 0,09 [0,019..0,154] * \text{pvalb} \\ 15 + \text{estadoGanglionar}$$

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

20 El Algoritmo T5b es un comité de dos miembros donde cada miembro es una combinación lineal de cuatro genes. Las fórmulas matemáticas para T5b se muestran más abajo, la notación es la misma que para T1 y T5. En T5b una variable no génica es el estado ganglionar codificado como 0, si el paciente tiene ganglios linfáticos negativos y como 1, si el paciente tiene ganglios linfáticos positivos y 0,5 si el estado de los ganglios linfáticos es desconocido. T5b se define mediante:

$$25 \text{ riesgoMiembro1} = 0,359536 [0,153..0,566] * (0,891 * \text{DHCR7} -4,378) \\ -0,288119 [-0,463..-0,113] * (0,485 * \text{MGP} +4,330) \\ +0,257341 [0,112..0,403] * (1,118 * \text{NMU} -5,128) \\ -0,337663 [-0,499..-0,176] * (0,674 * \text{AZGP1} -0,777)$$

$$30 \text{ riesgoMiembro2} = -0,374940 [-0,611..-0,139] * (0,707 * \text{RBBP8} -0,934) \\ -0,387371 [-0,597..-0,178] * (0,814 * \text{IL6ST} -5,034) \\ +0,800745 [0,551..1,051] * (0,860 * \text{RACGAP1} -2,518) \\ +0,770650 [0,323..1,219] * \text{estadoGanglionar}$$

$$\text{riesgo} = \text{riesgoMiembro1} + \text{riesgoMiembro2}$$

35 El experto en la técnica entiende que estos algoritmos representan ejemplos concretos y que basándose en la información referente a la asociación de la expresión génica con el resultado proporcionado en la tabla 2, se pueden establecer algoritmos alternativos utilizando técnicas rutinarias.

Simplificación del algoritmo empleando subconjuntos de genes

40 El "Ejemplo de algoritmo T5" es un comité predictor que consiste en 4 miembros con 2 genes de interés cada uno. Cada miembro es un predictor independiente y auto-contenido de recurrencia distante, cada miembro adicional contribuye a la robustez y poder predictivo del algoritmo para predecir el tiempo hasta la metástasis, el tiempo hasta la muerte o la verosimilitud de supervivencia para un paciente con cáncer de mama. La ecuación de más abajo muestra el "Ejemplo de Algoritmo T5"; para facilitar la lectura, el número de dígitos después de la coma decimal se ha acordado a 2; el intervalo entre corchetes enumera el intervalo estimado de los coeficientes (media +/- 3 desviaciones típicas).

45

Algoritmo T5:

+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

5 -0,18 [-0,31..-0,06] * MGP -0,13 [-0,25..-0,02] * STC2

índices c: grupo de Entrenamiento = 0,724,

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

10 Al analizar una cohorte diferente de la cohorte de resultados (234 muestras de tumor) fue sorprendente descubrir que algunas simplificaciones del "algoritmo T5 original" todavía proporcionaban un rendimiento diagnóstico no significativamente inferior al algoritmo T5 original. La simplificación más sencilla consistió en reducir el comité predictor a solamente un miembro. Los ejemplos del rendimiento de los "comités de un miembro" se muestran más abajo:

solamente miembro 1:

15 +0,41 [0,21..0,61] * BIRC5 -0,33 [-0,57..-0,09] * RBBP8

índices c: grupo de Entrenamiento = 0,653, Cohorte independiente = 0,681

solamente miembro 2:

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

20 índices c: grupo de Entrenamiento = 0,664, Cohorte independiente = 0,696

solamente miembro 3:

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

índices c: grupo de Entrenamiento = 0,666, Cohorte independiente = 0,601

solamente miembro 4:

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

índices c: grupo de Entrenamiento = 0,668, Cohorte independiente = 0,593

30 El rendimiento de los comités de un miembro mostrado en una cohorte independiente de 234 muestras se reduce notablemente en comparación con el rendimiento del algoritmo completo. Aún así, la utilización de un comité que consiste en menos miembros permite una estimación menos costosa, más simple del riesgo de recurrencia del cáncer de mama o muerte por cáncer de mama que podría ser aceptable para ciertos fines diagnósticos.

35 La combinación gradual de más de uno pero menos de cuatro miembros con un nuevo algoritmo predictor del comité pronóstico, frecuentemente conduce a un incremento pequeño pero significativo en el rendimiento del diagnóstico en comparación con un comité de un miembro. Resultó sorprendente averiguar que había mejoras marcadas por la combinación de miembros del comité mientras otras combinaciones casi no proporcionaban mejora. Inicialmente, la hipótesis fue que una combinación de miembros que representaban motivos biológicos similares a los reflejados por los genes empleados proporcionó una mejora más pequeña que la combinación de miembros que reflejaban motivos biológicos claramente diferentes. Sin embargo, este no era el caso. No se pudo identificar ninguna norma para predecir la combinación de algunos genes para generar un algoritmo que muestre más poder pronóstico que otras combinaciones de genes. Las combinaciones prometedoras solamente pudieron ser seleccionadas basándose en datos experimentales. Las combinaciones identificadas de los miembros del comité combinados para proporcionar algoritmos todavía potentes, simplificados se muestran a continuación.

miembros 1 y 2 solamente:

+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

45 índices c: grupo de Entrenamiento = 0,675, Cohorte independiente = 0,712

miembros 1 y 3 solamente:

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

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

índices c: grupo de Entrenamiento = 0,697, Cohorte independiente = 0,688

5 miembros 1 y 4 solamente:

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

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

índices c: grupo de Entrenamiento = 0,705, Cohorte independiente = 0,679

miembros 2 y 3 solamente:

10 $+0,38 [0,15..0,61] * UBE2C -0,30 [-0,55..-0,06] * IL6ST$

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

índices c: grupo de Entrenamiento = 0,698, Cohorte independiente = 0,670

miembros 1, 2 y 3 solamente:

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

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: grupo de Entrenamiento = 0,701, Cohorte independiente = 0,715

20 También es posible no omitir miembros de comités completos sino un único gen o genes de diferentes miembros del comité, pero se requiere una readaptación del algoritmo completo. Con todo, también puede resultar ventajoso llevarlo a cabo. El rendimiento de los algoritmos simplificados generados omitiendo miembros completos o genes individuales es en gran medida idéntico.

Variantes de algoritmos mediante sustitución de genes

25 Los algoritmos descritos anteriormente, tales como el "Ejemplo de algoritmo T5", también pueden ser modificados reemplazando uno o más genes por uno o más genes diferentes. El propósito de semejantes modificaciones es reemplazar genes difíciles de medir en una plataforma específica por un gen más sencillo de analizar en esta plataforma. Si bien semejante transferencia puede no proporcionar necesariamente un mejor rendimiento en comparación con el algoritmo de partida, puede proporcionar la pista para implantar el algoritmo pronóstico en una plataforma de diagnóstico concreta. En general, el remplazo de un gen por otro gen a la vez que se conserva el poder diagnóstico del algoritmo predictivo se puede lograr mejor reemplazando un gen por un gen expresado

30 simultáneamente con una elevada correlación (mostrada p. ej. por el coeficiente de correlación de Pearson). Sin embargo, se debe tener en cuenta que las expresiones del ARNm de dos genes muy correlativos en una plataforma pueden mostrarse bastante independientes entre sí cuando se evalúan en una otra plataforma. Por consiguiente, semejante remplazo aparentemente fácil cuando se reduce a la práctica experimental, puede proporcionar resultados decepcionantemente pobres así como resultados sorprendentemente sólidos, siempre dependiendo de

35 los imponderables de la plataforma empleada. Repitiendo este procedimiento se pueden reemplazar diversos genes.

La eficacia de semejante enfoque se puede demostrar evaluando el rendimiento predictivo de la puntuación del algoritmo T5 y sus variantes en las cohortes de validación. La siguiente tabla muestra el índice c con respecto al criterio de valoración de recurrencia distante en dos cohortes de validación.

Tabla 7

Variante	Estudio de Validación A	Estudio de Validación B
Algoritmo original T5	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

Variante	Estudio de Validación A	Estudio de Validación B
Sustituyendo BIRC5 por UBE2C (sin ajuste del coeficiente)	índice c = 0,707	índice c = 0,678

Se puede observar que la omisión de uno de los genes T5, aquí mostrada para BIRC5 por ejemplo, reduce notablemente el rendimiento predictivo. Reemplazándolo por otro gen proporciona aproximadamente el mismo rendimiento.

- 5 Un método mejor para reemplazar un gen consiste en la readaptación del algoritmo. Puesto que T5 consiste en cuatro miembros del comité independientes, solamente hay que readaptar el miembro que contiene el gen reemplazado. Las siguientes ecuaciones demuestran reemplazos de genes del algoritmo T5 mostrado más arriba entrenado en una cohorte de 234 pacientes con cáncer de mama. Solamente un miembro se muestra más abajo, para el cálculo del índice c se utilizaron los restantes miembros inalterados con respecto al Algoritmo T5 original. El intervalo entre corchetes enumera el intervalo estimado de los coeficientes: media +/- 3 desviaciones típicas.

Miembro 1 de T5:

Miembro original 1:

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

índices c: grupo de Entrenamiento = 0,724, Cohorte independiente = 0,705

- 15 reemplazar BIRC5 por TOP2A en el miembro 1:

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

índices c: grupo de Entrenamiento = 0,734, Cohorte independiente = 0,694

reemplazar BIRC5 por RACGAP1 en el miembro 1:

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

- 20 índices c: grupo de Entrenamiento = 0,736, Cohorte independiente = 0,743

reemplazar RBBP8 por CELSR2 en el miembro 1:

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

índices c: grupo de Entrenamiento = 0,726, Cohorte independiente = 0,680

reemplazar RBBP8 por PGR en el miembro 1:

- 25 $+0,35 [0,15..0,54] * BIRC5 -0,09 [-0,23..0,05] * PGR$

índices c: grupo de Entrenamiento = 0,727, Cohorte independiente = 0,731

Miembro 2 de T5:

Miembro original 2:

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

- 30 índices c: grupo de Entrenamiento = 0,724, Cohorte independiente = 0,725

reemplazar UBE2C por RACGAP1 en el miembro 2:

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

índices c: grupo de Entrenamiento = 0,735, Cohorte independiente = 0,718

reemplazar UBE2C por TOP2A en el miembro 2:

- 35 $+0,042 [0,20..0,65] * TOP2A -0,38 [-0,62..-0,13] * IL6ST$

índices c: grupo de Entrenamiento = 0,734, Cohorte independiente = 0,700

ES 2 525 382 T3

reemplazar IL6ST por INPP4B en el miembro 2:

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

índices c: grupo de Entrenamiento = 0,725, Cohorte independiente = 0,686

reemplazar IL6ST por MAPT en el miembro 2:

5 +0,45 [0,22..0,69] * UBE2C -0,14 [-0,28..0,01] * MAPT

índices c: grupo de Entrenamiento = 0,727, Cohorte independiente = 0,711

Miembro 3 de T5:

Miembro original 3:

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

10 índices c: grupo de Entrenamiento = 0,724, Cohorte independiente = 0,705

reemplazar AZGP1 por PIP en el miembro 3:

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

índices c: grupo de Entrenamiento = 0,725, Cohorte independiente = 0,692

reemplazar AZGP1 por EPHX2 en el miembro 3:

15 -0,23 [-0,43..-0,02] * EPHX2 +0,37 [0,10..0,64] * DHCR7

índices c: grupo de Entrenamiento = 0,719, Cohorte independiente = 0,698

reemplazar AZGP1 por PLAT en el miembro 3:

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

índices c: grupo de Entrenamiento = 0,712, Cohorte independiente = 0,715

20 reemplazar DHCR7 por AURKA en el miembro 3:

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

índices c: grupo de Entrenamiento = 0,716, Cohorte independiente = 0,733

Miembro 4 de T5:

Miembro original 4:

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

índices c: grupo de Entrenamiento = 0,724, Cohorte independiente = 0,705

reemplazar MGP por APOD en el miembro 4:

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

índices c: grupo de Entrenamiento = 0,717, Cohorte independiente = 0,679

30 reemplazar MGP por EGFR en el miembro 4:

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

índices c: grupo de Entrenamiento = 0,715, Cohorte independiente = 0,708

reemplazar STC2 por INPP4B en el miembro 4:

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

35 índices c: grupo de Entrenamiento = 0,719, Cohorte independiente = 0,693

reemplazar STC2 por SEC14L2 en el miembro 4:

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

ES 2 525 382 T3

índices c: grupo de Entrenamiento = 0,718, Cohorte independiente = 0,681

Se puede observar que los remplazos de genes individuales identificados experimentalmente para una cuantificación con PCR cinética normalmente afectan al rendimiento predictivo del algoritmo T5, evaluado mediante el índice c solo de manera insignificante.

- 5 La siguiente tabla (Tab. 8) muestra el remplazo potencial de genes candidato para los genes del algoritmo T5. Cada gen candidato se muestra en una celda de la tabla: El nombre del gen está seguido del coeficiente de correlación de Pearson absoluto entre corchetes de la expresión del gen original en el Algoritmo T5 y el candidato del remplazo, y el ID del conjunto de sondas HG-U133A.

Tabla 8

BIRC5	RBBP8	UBE2C	IL6ST	AZGP1	DHCR7	MGP	STC2
UBE2C (0,775), 202954_at	CELSR2 (0,548), 204029_at	BIRC5 (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), 222077_s_at	STC2 (0,450), 203438_at	EPHX2 (0,369), 209368_at	BIRC5 (0,323), 202095_s_at	IL6ST (0,327), 212196_at	IL65T (0,450), 212196_at
RACGAP1 (0,704), 222077_s_at	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 (0,662), 219990_at	PGR (0,377), 208305_at	PGR (0,302), 208305_at			ABAT (0,409), 209459_s_at
PCNA (0,544), 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		ADRA2A (0,391), 209569_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

ES 2 525 382 T3

BIRC5	RBBP8	UBE2C	IL6ST	AZGP1	DHCR7	MGP	STC2
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	RBBP8 (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					
EGFR (0,346), 201983_s_at		PRSS16 (0,326), 208165_s_at					
CCND1 (0,331), 208712_at		EGFR (0,320), 201983_s_at					
TRIM29 (0,325), 202504_at		DHCR7 (0,315), 201791_s_at					
DHCR7 (0,323), 201791_s_at		EPHX2 (0,315), 209368_at					
PIP (0,308), 206509_at		TRIM29 (0,311), 202504_at					
TFAP2B (0,306), 214451_at							
WNT5A (0,303), 205990_s_at							
APOD (0,301), 201525_at							
PTPRT (0,301), 205948_at							

La siguiente tabla (Tab. 9) enumera las secuencias de cebadores y sondas de qRT-PCR utilizadas para la tabla anterior.

Tabla 9

gen	sonda	cebador directo	cebador inverso
ABAT	TCGCCCTAAGAGGCTCTTCCTC	GGCAACTTGAGGTCTGACTTTTG	GGTCAGCTACAAGTGGTGTA
ADRA2A	TTGTCCTTTCCCCCTCCGTGC	CCCCAAGAGCTGTAGGTATCAA	TCAATGACATGATCTCAACCAGAA
APOD	CATCAGCTCTCAACTCCTGGTTAACA	ACTCATAATGAAAAACGGAAAGATC	TCACCTTCGATTTGATTCACAGTT
ASPH	TGGGAGGAAGGCAAGGTGCTCATC	TGTGCCAACGAGACCAAGAC	TCGTGCTCAAAGGAGTCTATCA
AURKA	CCGTCAGCCTGTGCTAGGCAT	AATCTGGAGGCAAGGTTCGA	TCTGGATTTGCCTCTGTGAA
BIRC5	AGCCAGATGACGACCCCATAGAGGAACA	CCCAAGTGTTCCTCTGCTTCAAG	CAACCGGACGAATGCTTTTT
CCND1			
CELSR2	ACTGACTTTCCTTCTGGAGCAGGTGGC	TCCAAGCATGTATTCAGACTTGT	TGCCCCACGCCTCTTTTTCT
CHPT1	CCACGGCCACCGAAGAGGCAC	CGCTCGTCTCATCTCCTACT	CCCAGTGACATAAAAAGGTATGC
CXCL12	CCACAGCAGGGTTTCAGGTCC	GCCACTACCCCTCCTGAA	TCACCTTGCCAACAGTTCTGAT
CYBRD1	AGGGCATCGCCATCATCGTC	GTCACCGGCTTCGTCTCA	CAGGTCCACGGCAGTCTGT
DCN	TCTTTTCAGCAACCCGGTCCA	AAGGCTTCTTATTCGGGTGTA	TGGATGGCTGTATCTCCAGTA
DHCR7	TGAGCGCCACCTCTCGA	GGGCTCTGCTCCCGATT	AGTCATAGGGCAAGCAGAAAAATTC
E2F8	CAGGATACCTAATCCCTCTCACGCAG	AAATGTCTCCGCAACCTTGTTT	CTGCCCCAGGGATGAG
EGFR			
EPHX2	TGAAGCGGGAGGACTTTTTGTA	CGATGAGAGTGTTCATCCATGCA	GCTGAGGCTGGGCTCTTCT
ESR1	ATGCCCTTTTGCCGATGCA	GCCAAATTTGTGTTGATGGATTAA	GACAAAACCGAGTCACATCAGTAATAG
GJA1	TGCACAGCCTTTTGATTTCCCGAT	CGGGAAGCACCATCTCTAACTC	TTTCATGTCCAGCAGCTAGTTTTT
HSPA2	CAAGTCAGCAAACAGCAAAA	CATGCACGAATAATCAAAAATGC	ACATTATTCGAGGTTTCTCTTAATGC
IL6ST	CAAGCTCCACCTTCCAAGGACCT	CCCTGAATCCATAAAGGCATACC	CAGCTTCGTTTTCCCTACTTTTT
INPP4B	TCCGAGCGCTGGATTGCATGAG	GCACACAGTTACACAAGGACTTCTTT	TCTCATGCGGCATCCTTCTC
MAPT	AGACTATTGACACTGCCGCCT	GTGGCTCAAAGGATAATATCAAACAC	ACCTTGCTCAGGTCAACTGGTT
MGP	CCTTCATATCCCTCAGCAGAGATGG	CCTTCATTAACAGGAGAAATGCAA	ATTGAGCTCGTGACAGGCTTA
NEK2	TCTGAACAAATGAATCGCATGCTCAACA	ATTTGTTGGCACACCTTATTCATGT	AAGCAGCCCAATGACCAGATa
PCNA	AAATACTAAAAATGCGCCGGCAATGA	GGGCGTGAACCTCACCAGTA	CTTCGGCCCTTAGTGTAATGATATC
PGR	TTGATAGAAACGCTGTGAGCTCGA	AGCTCATCAAGGCAATTTGTTT	ACAAGATCATGCAAGTTATCAAGAAGTT
PIP	TGCATGGTGGTTAAAACCTTACCTCA	TGCTTGAGTTCAAACAGAATTG	CACCTTGATAGGGGATGCTGCTA
PLAT	CAGAAAGTGGCCATGCCACCCTG	TGGGAAGACATGAATGCACACTA	GGAGGTTGGGCTTTAGCTGAA
PRSS16	CACTGCCGGTCAACCAACCA	CTGAGGAGCACAGAACCTCAACT	CGAACTCGGTACATGCTGATACAA
PTGER3	TCGGTCTGCTGGTCTCCGCTCC	CTGATTGAAGATCATTTCAACATCA	GACGGCCATTAGCTTATGG
PTPR	TTGGCTCTGGACACCTCACA	GAGTGTGGCCTTACCATTGC	GAGCGGGAACCTTGGGATAG
RACGAP1	ACTGAGAATCTCCACCCGGCGCA	TCGCCAACTGGATAAATTGGA	GAATGTGCGGAATCTGTTTGAG
RBBP8	ACCGATTCCGCTACATTCACCCAAC	AGAAATTTGGCTTCTGCTCAAG	AAAACCAACTTCCAAAATTTCTCT
SCUBE2	CTAGAGGGTCCAGGTCCCATACGTGACATA	TGTGGATTGAGTTCAAGTCCAATG	CCATCTCGAACTATGCTTCAATGAGT
SEC14L2	TGGGAGGCATGCAACGCGTG	AGGTCTTACTAAGCAGTCCCATCTCT	CGACCGCACCTGAACCTC
SQLE	TATGCGTCTCCAAAAGAAGAACACCTCG	GCAAGCTTCTTCTCTCTTCA	CCTTTAGCAGTTTTCTCCATAGTTTTATATC
STC2	TCTCACCTTGACCCTCAGCCAAG	ACATTTGACAAATTTCCCTTAGGATT	CCAGGACGCAGCTTACCAA
TFAP2B	CAACACCACCACTAACAGGCACACGTC	GGCATGGACAAGATGTTCTTGA	CCTCTTGTGCGCAGTTTACT
TOP2A	CAGATCAGGACCAAGATGGTTCCACAT	CATTGAAGACGCTTCGTTATGG	CCAGTTGTGATGGATAAAAATTAATCAG
TRIM29	TGCTGTCTCACTACCGCCATTCTACG	TGGAAATCTGGCAAGCAGACT	CAATCCCGTTGCCTTTGTTG
UBE2C	TGAACACACATGCTGCCGAGCTCTG	CTTCTAGGAGAACCACATTGATAGT	GTTTCTGCAGGTACTTCTAAAAGCT
WNT5A	TATTCACATCCCCTCAGTTGCAAGTGAATTG	CTGTGGCTCTAATTTATGCATAATG	TTAGTGCTTTTTGCTTCAAGATCTT

5 Una segunda alternativa para la selección sin supervisión de posibles candidatos para el remplazo de genes se basa en los datos de Affymetrix solamente. Esto tiene la ventaja de que se puede realizar basándose solo en datos ya publicados (p. ej. de www.ncbi.nlm.nih.gov/geo/). La siguiente tabla (Tab. 10) enumera los candidatos de remplazo del grupo de sondas HG-U133a para los grupos de sondas utilizados en los algoritmos T1 - T5. Esto se basa en datos de entrenamiento de estos algoritmos. El encabezamiento de la columna contiene el nombre del gen y el ID del conjunto de sondas en negrita. A continuación, se enumeran los 10 conjuntos de sondas mejor correlacionados, donde cada celda de la tabla contiene el ID del conjunto de sondas, el coeficiente de correlación entre corchetes y el nombre del gen.

10

ES 2 525 382 T3

Tabla 10

IUBE2C	BIRC5	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	208683_at (-0,33) CAPN2	205347_s_at (0,45) TMSL8	203439_s_at (0,88) STC2
202095_s_at (0,82) BIRC5	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) PHC1	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) JMJD2B
204092_s_at (0,80) STK6	202580_x_at (0,77) FOXM1	218009_s_at (0,45) PRC1	203362_5_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
AZGP1	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) ARHGDI1B	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) FLJ11259	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	203685_at (0,38)	200648_s_at	200748_s_at	203439_s_at	203131_at (0,78)	212195_at (0,45)	210715_s_at

ES 2 525 382 T3

IUBE2C	BIRC5	DHCR7	RACGAP1	AURKA	PVALB	NMU	STC2
(0,42) ---	BCL2	(0,52) GLUL	(0,37) FTH1	(0,46) STC2	PDGFRA	IL6ST	(0,33) SPINT2
204288_s_at (0,41) SORBS2	203304_at (- 0,38) 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) TGFBR2	215867_x_at (0,45) CA12	218074_at (0,33) FAM96B

Después de la selección de un gen o un conjunto de sondas hay que definir un mapeo matemático entre los valores del gen que se va a reemplazar y los del nuevo gen. Existen diversas alternativas que se comentan aquí basadas en el ejemplo "reemplazar los valores de delta-Ct de BIRC5 por RACGAP1". En los datos de entrenamiento la distribución conjunta de expresiones tiene el aspecto de la Figura 3.

El coeficiente de correlación de Pearson es 0,73.

Un enfoque consiste en crear una función de mapeo a partir de RACGAP1 a BIRC5 por regresión. La regresión lineal es la primera elección y proporciona en este ejemplo

$$BIRC5 = 1,22 * RACGAP1 - 2,85.$$

10 Utilizando esta ecuación se puede reemplazar fácilmente la variable BIRC5 p. ej. en el algoritmo T5 por el lado derecho. En otros ejemplos se pueden adecuar la regresión robusta, la regresión polinomial o pre-transformaciones no lineales univariantes.

15 El método de regresión da por sentado ruido en la medición sobre BIRC5, pero no ruido sobre RACGAP1. Por lo tanto el mapeo no es simétrico con respecto a la intercambiabilidad de las dos variables. Un enfoque de mapeo asimétrico se basaría en dos transformadas z univariantes

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

y

$$z = (RACGAP1 - (RAGCAP1) \text{ media}) / (RACGAP1) \text{ std}$$

$$z = (BRIC5 - 8,09) / 1,29 = (RACGAP1 - 8,95) / 0,77$$

20 $BRC5 = 1,67 * RACGAP1 + -6,89$

De nuevo, en otros ejemplos, pueden ser adecuadas otras transformaciones: normalización por medio de mediana y/o mad (desviación absoluta con respecto a la mediana), mapeos no lineales, u otros.

LISTA DE SECUENCIAS

	<110> Sividon Diagnostics GmbH	
5	<120> Método para la predicción de recurrencia del cáncer de mama bajo tratamiento endocrino	
	<130> Sividon	
	<160> 141	
10	<170> PatentIn versión 3.3	
	<210> 1	
	<211> 22	
15	<212> DNA	
	<213> Homo sapiens	
	<400> 1	
20	tcgccctaag aggctctcc tc	22
	<210> 2	
	<211> 23	
	<212> DNA	
25	<213> Homo sapiens	
	<400> 2	
	ggcaactga ggtctgact ttg	23
	<210> 3	
30	<211> 22	
	<212> DNA	
	<213> Homo sapiens	
	<400> 3	
35	ggtcagctca caagtgggtg ga	22
	<210> 4	
	<211> 22	
	<212> DNA	
40	<213> Homo sapiens	
	<400> 4	
	ttgccttc cccctccgt gc	22
45	<210> 5	
	<211> 23	
	<212> DNA	
	<213> Homo sapiens	
50	<400> 5	
	ccccaagagc tgtaggtat caa	23
	<210> 6	
	<211> 24	
55	<212> DNA	
	<213> Homo sapiens	
	<400> 6	
60	tcaatgacat gatctcaacc agaa	24
	<210> 7	
	<211> 27	
	<212> DNA	
	<213> Homo sapiens	

ES 2 525 382 T3

	<400> 7 catcagctct caactcctgg ttaaca	27
5	<210> 8 <211> 26 <212> DNA <213> Homo sapiens	
10	<400> 8 actcactaat ggaaaacgga aagatc	26
15	<210> 9 <211> 24 <212> DNA <213> Homo sapiens	
20	<400> 9 tcaccttcca ttgattcac agtt	24
25	<210> 10 <211> 24 <212> DNA <213> Homo sapiens	
30	<400> 10 tgggaggaag gcaaggtgct catc	24
35	<210> 11 <211> 20 <212> DNA <213> Homo sapiens	
40	<400> 11 tgtgccaacg agaccaagac	20
45	<210> 12 <211> 21 <212> DNA <213> Homo sapiens	
50	<400> 12 tcgtgctcaa aggagtcac a	21
55	<210> 13 <211> 21 <212> DNA <213> Homo sapiens	
60	<400> 13 ccgtcagcct gtgctaggca t	21
65	<210> 14 <211> 20 <212> DNA <213> Homo sapiens	
	<400> 14 aatctggagg caaggttcga	20
	<210> 15 <211> 21 <212> DNA <213> Homo sapiens	
	<400> 15	

ES 2 525 382 T3

	tctggatttg cctcctgtga a	21
5	<210> 16 <211> 28 <212> DNA <213> Homo sapiens	
10	<400> 16 agccagatga cgaccccata gaggaaca	28
15	<210> 17 <211> 23 <212> DNA <213> Homo sapiens	
	<400> 17 cccagtgttt ctctgcttc aag	23
20	<210> 18 <211> 20 <212> DNA <213> Homo sapiens	
25	<400> 18 caaccggacg aatgctttt	20
30	<210> 19 <211> 27 <212> DNA <213> Homo sapiens	
35	<400> 19 actgactttc ctctggagc aggtggc	27
40	<210> 20 <211> 24 <212> DNA <213> Homo sapiens	
	<400> 20 tccaagcatg tattccagac ttgt	24
45	<210> 21 <211> 20 <212> DNA <213> Homo sapiens	
50	<400> 21 tgcccacagc ctcttttct	20
55	<210> 22 <211> 21 <212> DNA <213> Homo sapiens	
	<400> 22 ccacggccac cgaagaggca c	21
60	<210> 23 <211> 21 <212> DNA <213> Homo sapiens	
65	<400> 23 cgctcgtgct catctctac t	21

ES 2 525 382 T3

<210> 24
 <211> 24
 <212> DNA
 5 <213> Homo sapiens

 <400> 24
 cccagtgac ataaaaggta tgtc 24

 10 <210> 25
 <211> 22
 <212> DNA
 <213> Homo sapiens

 15 <400> 25
 ccacagcagg gtttcaggtt cc 22

 <210> 26
 <211> 19
 20 <212> DNA
 <213> Homo sapiens

 <400> 26
 gccactacc cctcctgaa 19

 25 <210> 27
 <211> 22
 <212> DNA
 <213> Homo sapiens

 30 <400> 27
 tcacctgcc aacagttctg at 22

 <210> 28
 <211> 20
 35 <212> DNA
 <213> Homo sapiens

 <400> 28
 40 agggcatcgc catcatcgtc 20

 <210> 29
 <211> 19
 <212> DNA
 45 <213> Homo sapiens

 <400> 29
 gtcaccggct tcgtttca 19

 50 <210> 30
 <211> 19
 <212> DNA
 <213> Homo sapiens

 55 <400> 30
 caggtccacg gcagtctgt 19

 <210> 31
 <211> 21
 60 <212> DNA
 <213> Homo sapiens

 <400> 31
 65 tctttcagc aaccgggtcc a 21

 <210> 32

ES 2 525 382 T3

<211> 22
 <212> DNA
 <213> Homo sapiens

5 <400> 32
 aaggcttctt attcgggtgt ga 22

<210> 33
 <211> 22
 10 <212> DNA
 <213> Homo sapiens

<400> 33
 15 tggatggctg tatctcccag ta 22

<210> 34
 <211> 19
 <212> DNA
 <213> Homo sapiens

20 <400> 34
 tgagcgcca cctctcga 19

<210> 35
 <211> 18
 <212> DNA
 <213> Homo sapiens

25 <400> 35
 30 gggctctgct tcccgatt 18

<210> 36
 <211> 24
 <212> DNA
 <213> Homo sapiens

35 <400> 36
 agtcataggg caagcagaaa attc 24

<210> 37
 <211> 26
 <212> DNA
 <213> Homo sapiens

40 <400> 37
 45 caggatacct aatccctctc acgcag 26

<210> 38
 <211> 22
 <212> DNA
 <213> Homo sapiens

50 <400> 38
 55 aaatgtctcc gcaacctgt tc 22

<210> 39
 <211> 17
 <212> DNA
 <213> Homo sapiens

60 <400> 39
 ctgccccag ggatgag 17

<210> 40
 <211> 24
 <212> DNA

65

ES 2 525 382 T3

	<213> Homo sapiens	
	<400> 40	
5	tgaagcggga ggacttttg taaa	24
	<210> 41	
	<211> 24	
	<212> DNA	
10	<213> Homo sapiens	
	<400> 41	
	cgatgagagt gtttatcca tgca	24
	<210> 42	
15	<211> 19	
	<212> DNA	
	<213> Homo sapiens	
	<400> 42	
20	gctgaggctg ggctcttct	19
	<210> 43	
	<211> 19	
	<212> DNA	
25	<213> Homo sapiens	
	<400> 43	
	atgcccttt gccgatgca	19
30	<210> 44	
	<211> 24	
	<212> DNA	
	<213> Homo sapiens	
35	<400> 44	
	gccaaattgt gttgatgga ttaa	24
	<210> 45	
	<211> 27	
40	<212> DNA	
	<213> Homo sapiens	
	<400> 45	
45	gacaaaaccg agtcacatca gtaatag	27
	<210> 46	
	<211> 25	
	<212> DNA	
50	<213> Homo sapiens	
	<400> 46	
	tgcacagcct ttgatttcc ccgat	25
	<210> 47	
55	<211> 22	
	<212> DNA	
	<213> Homo sapiens	
	<400> 47	
60	cggaagcac catctctaac tc	22
	<210> 48	
	<211> 24	
	<212> DNA	
65	<213> Homo sapiens	

ES 2 525 382 T3

	<400> 48 ttcatgtcca gcagctagtt tttt	24
5	<210> 49 <211> 21 <212> DNA <213> Homo sapiens	
10	<400> 49 caagtcagca aacacgcaaa a	21
15	<210> 50 <211> 24 <212> DNA <213> Homo sapiens	
20	<400> 50 catgcacgaa ctaatcaaaa atgc	24
25	<210> 51 <211> 27 <212> DNA <213> Homo sapiens	
30	<400> 51 acattattcg aggtttctct ttaatgc	27
35	<210> 52 <211> 24 <212> DNA <213> Homo sapiens	
40	<400> 52 caagctccac ctccaaagg acct	24
45	<210> 53 <211> 23 <212> DNA <213> Homo sapiens	
50	<400> 53 ccctgaatcc ataaaggcat acc	23
55	<210> 54 <211> 24 <212> DNA <213> Homo sapiens	
60	<400> 54 cagcttcggtt tttccctact tttt	24
65	<210> 55 <211> 22 <212> DNA <213> Homo sapiens	
70	<400> 55 tccgagcgct ggattgcatg ag	22
75	<210> 56 <211> 25 <212> DNA <213> Homo sapiens	
80	<400> 56 gcaccagtta cacaaggact tcttt	25

<210> 57
 <211> 21
 <212> DNA
 5 <213> Homo sapiens

 <400> 57
 tctctatgcg gcatccttct c 21

 10 <210> 58
 <211> 23
 <212> DNA
 <213> Homo sapiens

 15 <400> 58
 agactatttg cacactgccg cct 23

 <210> 59
 <211> 26
 20 <212> DNA
 <213> Homo sapiens

 <400> 59
 gtggctcaaa ggataatc aaacac 26
 25
 <210> 60
 <211> 22
 <212> DNA
 <213> Homo sapiens
 30
 <400> 60
 accttgctca ggtcaactgg tt 22

 <210> 61
 <211> 26
 <212> DNA
 <213> Homo sapiens
 35
 <400> 61
 ccttcatac ccctcagcag agatgg 26
 40
 <210> 62
 <211> 24
 <212> DNA
 <213> Homo sapiens
 45
 <400> 62
 ccttcattaa caggagaaat gcaa 24
 50
 <210> 63
 <211> 22
 <212> DNA
 <213> Homo sapiens
 55
 <400> 63
 attgagctcg tggacaggct ta 22

 <210> 64
 <211> 30
 60 <212> DNA
 <213> Homo sapiens

 <400> 64
 tcctgaacaa atgaatcgca tgcctacaa 30
 65
 <210> 65

<211> 26
 <212> DNA
 <213> Homo sapiens

5 <400> 65
 atttggtggc acaccttatt acatgt 26

<210> 66
 <211> 21
 10 <212> DNA
 <213> Homo sapiens

<400> 66
 15 aagcagccca atgaccagat a 21

<210> 67
 <211> 25
 <212> DNA
 <213> Homo sapiens

20 <400> 67
 aaatactaaa atgcgccggc aatga 25

<210> 68
 <211> 20
 <212> DNA
 <213> Homo sapiens

25 <400> 68
 30 gggcgtgaac ctcaccagta 20

<210> 69
 <211> 25
 <212> DNA
 <213> Homo sapiens

35 <400> 69
 cttcggccct tagtgtaatg atatc 25

40 <210> 70
 <211> 24
 <212> DNA
 <213> Homo sapiens

45 <400> 70
 ttgatagaaa cgctgtgagc tcga 24

<210> 71
 <211> 22
 <212> DNA
 <213> Homo sapiens

50 <400> 71
 55 agctcatcaa ggcaattggt tt 22

<210> 72
 <211> 28
 <212> DNA
 <213> Homo sapiens

60 <400> 72
 acaagatcat gcaagttatc aagaagtt 28

<210> 73
 <211> 25
 <212> DNA

65

<213> Homo sapiens
 <400> 73
 5 tgcatggtgg ttaaaactta cctca 25
 <210> 74
 <211> 23
 <212> DNA
 <213> Homo sapiens
 10
 <400> 74
 tgcttgcaagt tcaaacagaa ttg 23
 <210> 75
 <211> 23
 <212> DNA
 <213> Homo sapiens
 15
 <400> 75
 20 cacctttagt agggatgctg cta 23
 <210> 76
 <211> 23
 <212> DNA
 <213> Homo sapiens
 25
 <400> 76
 cagaaagtgg ccatgccacc ctg 23
 <210> 77
 <211> 23
 <212> DNA
 <213> Homo sapiens
 30
 <400> 77
 35 tgggaagaca tgaatgcaca cta 23
 <210> 78
 <211> 21
 <212> DNA
 <213> Homo sapiens
 40
 <400> 78
 45 ggaggtggg cttagctga a 21
 <210> 79
 <211> 21
 <212> DNA
 <213> Homo sapiens
 50
 <400> 79
 cactgccggt cacccacacc a 21
 <210> 80
 <211> 23
 <212> DNA
 <213> Homo sapiens
 55
 <400> 80
 60 ctgaggagca cagaacctca act 23
 <210> 81
 <211> 25
 <212> DNA
 <213> Homo sapiens
 65

	<400> 81 cgaactcggg acatgtctga taaa	25
5	<210> 82 <211> 22 <212> DNA <213> Homo sapiens	
10	<400> 82 tcggctgct ggtctccgct cc	22
15	<210> 83 <211> 26 <212> DNA <213> Homo sapiens	
20	<400> 83 ctgattgaag atcatttca acatca	26
25	<210> 84 <211> 20 <212> DNA <213> Homo sapiens	
30	<400> 84 gacggccatt cagcttatgg	20
35	<210> 85 <211> 22 <212> DNA <213> Homo sapiens	
40	<400> 85 ttggcttctg gacaccctca ca	22
45	<210> 86 <211> 22 <212> DNA <213> Homo sapiens	
50	<400> 86 gagttgtggc ctctaccatt gc	22
55	<210> 87 <211> 20 <212> DNA <213> Homo sapiens	
60	<400> 87 gagcgggaac cttgggatag	20
65	<210> 88 <211> 23 <212> DNA <213> Homo sapiens	
70	<400> 88 actgagaatc tccaccggc gca	23
75	<210> 89 <211> 21 <212> DNA <213> Homo sapiens	
80	<400> 89 tcgccaactg gataaattgg a	21

<210> 90
 <211> 22
 <212> DNA
 5 <213> Homo sapiens

 <400> 90
 gaatgtgCGg aatctgtttg ag 22

 10 <210> 91
 <211> 26
 <212> DNA
 <213> Homo sapiens

 15 <400> 91
 accgattccg ctacattcca cccaac 26

 <210> 92
 <211> 22
 20 <212> DNA
 <213> Homo sapiens

 <400> 92
 25 agaaattggc ttctgtca ag 22

 <210> 93
 <211> 25
 <212> DNA
 <213> Homo sapiens
 30
 <400> 93
 aaaaccaact tcccaaaaat tctct 25

 <210> 94
 <211> 31
 <212> DNA
 <213> Homo sapiens
 35
 <400> 94
 40 ctagagggtt ccaggGCCa tacgtgacat a 31

 <210> 95
 <211> 24
 <212> DNA
 45 <213> Homo sapiens

 <400> 95
 tgtggattca gttcaagtcc aatg 24
 50
 <210> 96
 <211> 27
 <212> DNA
 <213> Homo sapiens
 55
 <400> 96
 ccatctcGaa ctatgtcttc aatgagt 27

 <210> 97
 <211> 20
 60 <212> DNA
 <213> Homo sapiens

 <400> 97
 65 tgggaggcat gcaacgcgtg 20

 <210> 98

<211> 26
 <212> DNA
 <213> Homo sapiens
 5 <400> 98
 aggtcttact aagcagtgccc atctct 26
 <210> 99
 <211> 18
 10 <212> DNA
 <213> Homo sapiens
 <400> 99
 cgaccggcac ctgaactc 18
 15 <210> 100
 <211> 29
 <212> DNA
 <213> Homo sapiens
 20 <400> 100
 tatgctctc ccaaagaag aacacctcg 29
 <210> 101
 <211> 21
 <212> DNA
 <213> Homo sapiens
 25 <400> 101
 gcaagcttc ttctccttc a 21
 <210> 102
 <211> 31
 <212> DNA
 <213> Homo sapiens
 35 <400> 102
 ccttagcag tttctccat agtttatat c 31
 40 <210> 103
 <211> 27
 <212> DNA
 <213> Homo sapiens
 45 <400> 103
 caacaccacc actaacaggc acacgtc 27
 <210> 104
 <211> 22
 <212> DNA
 <213> Homo sapiens
 50 <400> 104
 ggcatggaca agatgttct ga 22
 55 <210> 105
 <211> 22
 <212> DNA
 <213> Homo sapiens
 60 <400> 105
 cctcctgtc gccagttta ct 22
 <210> 106
 <211> 28
 <212> DNA

<213> Homo sapiens
 <400> 106
 5 cagatcagga ccaagatggt tccacat 28
 <210> 107
 <211> 22
 <212> DNA
 <213> Homo sapiens
 10 <400> 107
 cattgaagac gcttcgttat gg 22
 <210> 108
 <211> 27
 <212> DNA
 <213> Homo sapiens
 15 <400> 108
 20 ccagttgtga tggataaaat taatcag 27
 <210> 109
 <211> 27
 <212> DNA
 25 <213> Homo sapiens
 <400> 109
 tgctgtctca ctaccggcca ttctacg 27
 30 <210> 110
 <211> 21
 <212> DNA
 <213> Homo sapiens
 35 <400> 110
 tggaaatctg gcaagcagac t 21
 <210> 111
 <211> 20
 <212> DNA
 40 <213> Homo sapiens
 <400> 111
 caatcccgtt gcctttgtg 20
 45 <210> 112
 <211> 25
 <212> DNA
 <213> Homo sapiens
 50 <400> 112
 tgaacacaca tgctgccgag ctctg 25
 <210> 113
 <211> 27
 <212> DNA
 <213> Homo sapiens
 55 <400> 113
 60 ctctaggag aaccaacat tgatagt 27
 <210> 114
 <211> 27
 <212> DNA
 65 <213> Homo sapiens

	<400> 114 gtttctgca ggtactctt aaaagct	27
5	<210> 115 <211> 30 <212> DNA <213> Homo sapiens	
10	<400> 115 tattcacatc ccctcagtg cagtgaattg	30
15	<210> 116 <211> 27 <212> DNA <213> Homo sapiens	
20	<400> 116 ctgtggctct taatttattg cataatg	27
25	<210> 117 <211> 26 <212> DNA <213> Homo sapiens	
25	<400> 117 ttagtgcttt tgcttcaa gatctt	26
30	<210> 118 <211> 23 <212> DNA <213> Homo sapiens	
35	<400> 118 tctcaccttg accctcagcc aag	23
40	<210> 119 <211> 26 <212> DNA <213> Homo sapiens	
40	<400> 119 acattgaca aatttcctt aggatt	26
45	<210> 120 <211> 20 <212> DNA <213> Homo sapiens	
50	<400> 120 ccaggacgca gctttaccaa	20
55	<210> 121 <211> 20 <212> DNA <213> Homo sapiens	
55	<400> 121 caccagccac caggccccag	20
60	<210> 122 <211> 18 <212> DNA <213> Homo sapiens	
65	<400> 122 tcctggaccg gcaagatc	18

<210> 123
 <211> 21
 <212> DNA
 5 <213> Homo sapiens

 <400> 123
 taggccaggc acttcagttt c 21

 10 <210> 124
 <211> 22
 <212> DNA
 <213> Homo sapiens

 15 <400> 124
 tcgctctcg gaaaccgga gc 22

 <210> 125
 <211> 20
 20 <212> DNA
 <213> Homo sapiens

 <400> 125
 gagcgagctg agtggttg 20
 25
 <210> 126
 <211> 21
 <212> DNA
 <213> Homo sapiens
 30
 <400> 126
 agtcagttgg tcagccatgc t 21

 <210> 127
 <211> 27
 <212> DNA
 <213> Homo sapiens
 35
 <400> 127
 cctgccaatc ccgatgaaat tggaaat 27

 <210> 128
 <211> 21
 <212> DNA
 45 <213> Homo sapiens

 <400> 128
 tgagtgtccc ccggtatctt c 21

 50 <210> 129
 <211> 21
 <212> DNA
 <213> Homo sapiens

 55 <400> 129
 tcagccgctt tcagatttc a 21

 <210> 130
 <211> 26
 60 <212> DNA
 <213> Homo sapiens

 <400> 130
 accctgctga ccttctcca ttcgt 26
 65
 <210> 131

ES 2 525 382 T3

<211> 22
 <212> DNA
 <213> Homo sapiens

5 <400> 131
 agaaattggc ttctgctca ag 22

<210> 132
 <211> 25
 10 <212> DNA
 <213> Homo sapiens

<400> 132
 15 aaaaccaact tcccaaaaat tctct 25

<210> 133
 <211> 23
 <212> DNA
 <213> Homo sapiens

20 <400> 133
 tgcttcaca agaaccgca gga 23

<210> 134
 <211> 20
 <212> DNA
 <213> Homo sapiens

25 <400> 134
 cgagccgacc atgttcat 20

<210> 135
 <211> 21
 <212> DNA
 <213> Homo sapiens

30 <400> 135
 aagcccaaaa agctgaaggt t 21

35 <210> 136
 <211> 22
 <212> DNA
 <213> Homo sapiens

40 <400> 136
 aagttctcc aaatggtcgg cc 22

<210> 137
 <211> 19
 <212> DNA
 <213> Homo sapiens

45 <400> 137
 ccgactcctt cgaccacaa 19

50 <210> 138
 <211> 23
 <212> DNA
 <213> Homo sapiens

55 <400> 138
 catcatccgc actcttttc ttc 23

60 <210> 139
 <211> 18
 <212> DNA

65

ES 2 525 382 T3

	<213> Homo sapiens	
	<400> 139	
5	tggtggcgg tgcctgga	18
	<210> 140	
	<211> 21	
	<212> DNA	
10	<213> Homo sapiens	
	<400> 140	
	tgtggctcct gcatgaagac a	21
	<210> 141	
15	<211> 24	
	<212> DNA	
	<213> Homo sapiens	
	<400> 141	
20	gtgacagcgg aagtgtatt gtac	24

REIVINDICACIONES

1. Un método para predecir el resultado del cáncer de mama en un tumor positivo para receptores de estrógeno y negativo para HER2 de un paciente con cáncer de mama, comprendiendo dicho método:
 - 5 (a) determinar en una muestra tumoral de dicho paciente los niveles de expresión de ARN de los siguientes 8 genes: UBE2C, BIRC5, DHCR7, STC2, AZGP1, RBBP8, IL6ST, y MGP
 - (b) combinar matemáticamente los valores de los niveles de expresión para los genes de dicho conjunto cuyos valores se determinaron en la muestra tumoral para proporcionar una puntuación combinada, en donde dicha puntuación combinada es indicativa de una prognosis de dicho paciente.
- 10 2. El método de acuerdo con la reivindicación 1, en donde dicho paciente ha recibido terapia endocrina o se contempla que vaya a recibir tratamiento endocrino.
3. El método de la reivindicación 2, en donde dicha terapia endocrina comprende tamoxifeno o inhibidor de aromatasas.
4. El método de acuerdo con las reivindicaciones 1 a 3 en donde se predice el riesgo de desarrollar recurrencia del cáncer de mama o muerte relacionada con el cáncer de mama.
- 15 5. El método de acuerdo con cualquiera de las reivindicaciones 1 a 4, en donde dicho nivel de expresión se determina en forma de nivel de expresión de ARN Mensajero.
6. El método de acuerdo con la reivindicación 5, en donde dicho nivel de expresión se determina mediante al menos uno de
 - un método basado en PCR,
 - 20 un método basado en micromatriz, y
 - un método basado en hibridación.
7. El método de una cualquiera de las reivindicaciones anteriores, en donde dicha determinación de los niveles de expresión es en una muestra tumoral fijada con formalina incluida en parafina o en una muestra tumoral recién congelada.
- 25 8. El método de cualquiera de las reivindicaciones anteriores, en donde se determina el nivel de expresión de al menos un gen marcador en forma de un patrón de expresión relativo a al menos un gen de referencia o a un valor de expresión medio computado.
9. El método de una cualquiera de las reivindicaciones anteriores, en donde dicha etapa de combinación matemática comprende una etapa de aplicación de un algoritmo a valores representativos de los niveles de
 - 30 expresión de genes dados, en particular en donde dicho algoritmo es una combinación lineal de dichos valores representativos de los niveles de expresión de genes dados.
10. El método de la reivindicación 9, en donde un valor para un valor representativo de un nivel de expresión de un gen dado se multiplica por un coeficiente.
11. El método de una cualquiera de las reivindicaciones anteriores, en donde uno, dos o más umbrales se
 - 35 determinan para dicha puntuación combinada, que discriminan en riesgo elevado y bajo, riesgo elevado, intermedio y bajo, o más grupos de riesgo aplicando el umbral sobre la puntuación combinada en particular en donde una puntuación combinada elevada es indicativa de los beneficios de la quimioterapia citotóxica.
12. El método de una cualquiera de las reivindicaciones anteriores, en donde la información referente al estado ganglionar del paciente se procesa en la etapa de combinación matemática de los valores de los niveles de
 - 40 expresión para los genes para proporcionar una puntuación combinada.
13. El método de la reivindicación 11 o 12, en donde dicha información referente al estado ganglionar es un valor numérico si dicho estado ganglionar es negativo y dicha información es un valor numérico diferente si dicho estado ganglionar es positivo y un número diferente o idéntico si dicho estado ganglionar es desconocido.
14. El uso de un kit para llevar a cabo un método de cualquiera de las reivindicaciones 1 a 13, comprendiendo dicho
 - 45 kit un conjunto de oligonucleótidos capaz de unirse específicamente a secuencias o a secuencias de fragmentos de los genes en una combinación de genes, en donde
 dicha combinación comprende los 8 genes: UBE2C, BIRC5, DHCR7, STC2, AZGP1, RBBP8, IL6ST, y MGP.

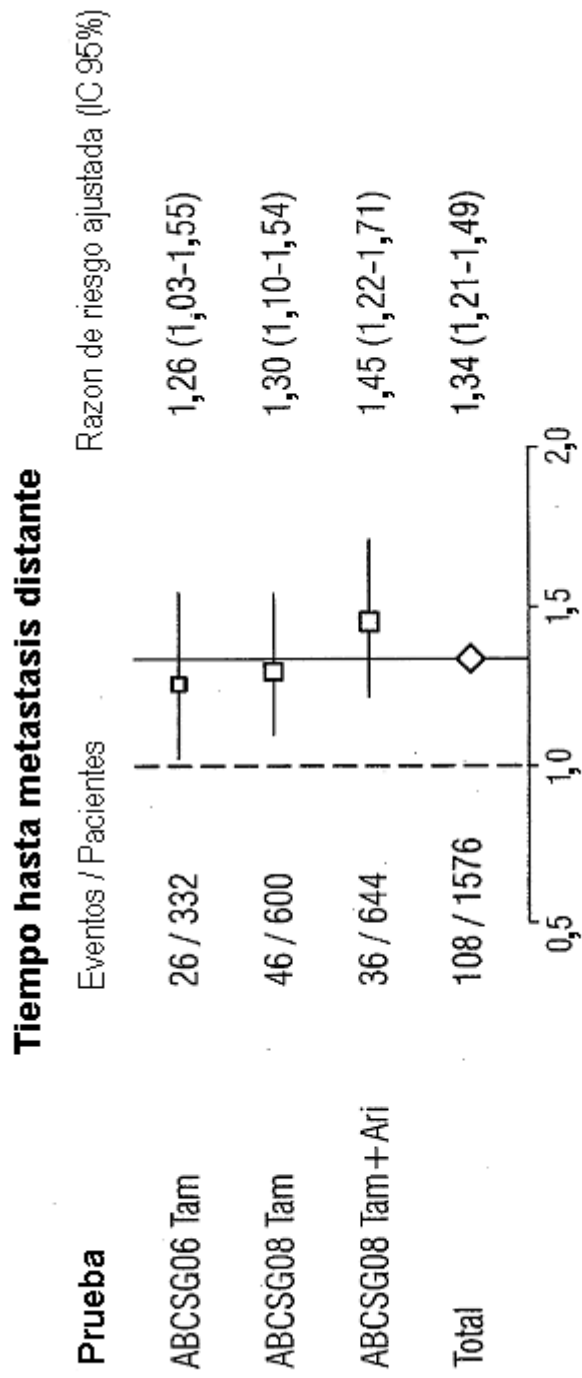


Fig.1

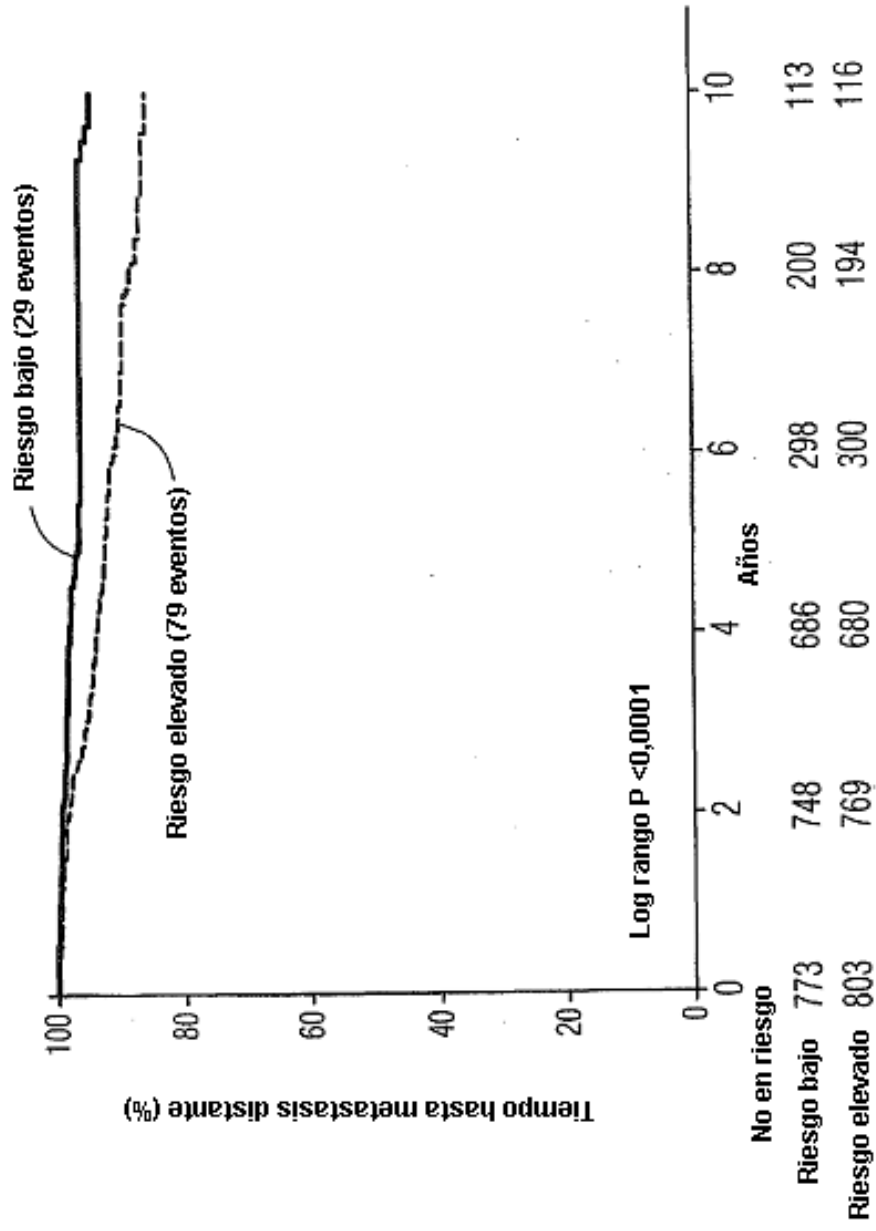


Fig.2

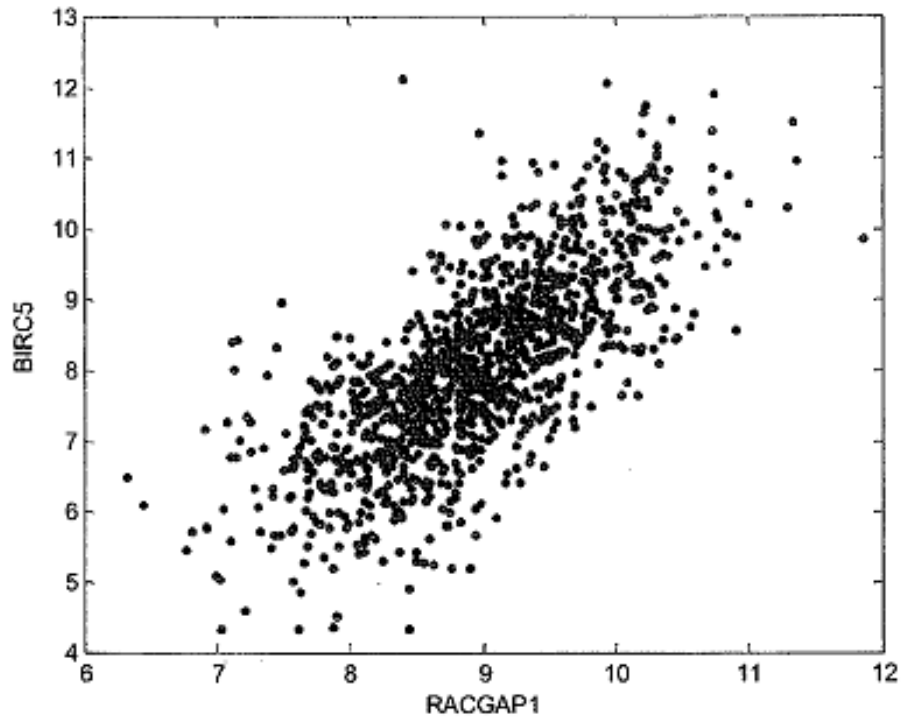


Fig.3