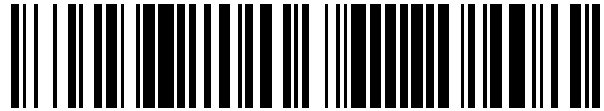


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 559 757**

51 Int. Cl.:

C12Q 1/68

(2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **01.09.2009 E 09810267 (6)**

97 Fecha y número de publicación de la concesión europea: **21.10.2015 EP 2331706**

54 Título: **Método para la determinación del porcentaje de células tumorales en una muestra**

30 Prioridad:

01.09.2008 EP 08163379

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

15.02.2016

73 Titular/es:

**AGENDIA N.V. (100.0%)
Science Park 406
1098 XH Amsterdam, NL**

72 Inventor/es:

**ROEPMAN, PAUL y
GLAS, ANNUSKA MARIA**

74 Agente/Representante:

LAZCANO GAINZA, Jesús

ES 2 559 757 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 determinación del porcentaje de células tumorales en una muestra

5 La invención se relaciona con el campo del diagnóstico de tumores, más específicamente a nuevos medios y métodos alternativos para determinar el porcentaje de células tumorales en una muestra. La invención permite la determinación del porcentaje de células tumorales basada en los niveles de expresión de un conjunto específico de genes. Los métodos y medios de la invención se puede usar adicionalmente a una revisión histopatológica de una muestra y son particularmente adecuados para determinar el porcentaje tumoral en las muestras que son demasiado pequeñas o no
10 adecuadas para la revisión histopatológica.

El diagnóstico de micromatrices de especímenes de tumor primario seccionados se basa en la medición de la expresión génica de un conjunto específico de genes predictivos o pronósticos. Comúnmente, las muestras tumorales globales se usan para el diagnóstico que consiste en células tumorales además de los tejidos que comprenden el estroma. Aunque el estroma tumoral probablemente juega un papel importante en el desarrollo de tumores y metástasis, los perfiles de expresión génica se generan típicamente usando tejidos tumorales que contienen suficientes células tumorales. Por ejemplo, se identificó el perfil de 70 genes pronósticos por Mammaprint usando originalmente muestras con al menos 50% de las células tumorales (van 't Veer y otros, 2002. Nature 415: 530-536). Si no se alcanza un porcentaje de células tumorales suficiente, no se tendrá en cuenta una muestra de tumor de este análisis.

20 Generalmente, la evaluación del porcentaje de células tumorales se realiza por un patólogo que usa una tinción tal como tinción de heamatoxilina-eosina (HE) para los análisis histopatológicos de una muestra. Sin embargo, a pesar de la grandes experiencias patológicas, la puntuación histopatológica del tumor continúa siendo una práctica subjetiva y que consume mucho tiempo y puede conducir a resultados no concluyentes (Furness y otros, 2003. Am J Surg Pathol 27:805-810; Ross y otros, 2007. Clin Cancer Res 13: 2831-2835] que incluyen la puntuación del porcentaje de células tumorales (Figura 1). Adicionalmente, la puntuación patológica de células tumorales es imposible para muestras que son muy pequeñas o inadecuadas para el análisis histopatológico tales como biopsias de núcleo y aspirados con aguja fina.

30 Golub y otros, 1999. Science 286, 531-537 discuten la posibilidad de usar firmas de genes para la predicción de clase en ciertos tipos de tumores. Este documento no describe que las firmas de genes se pueden usar para determinar un porcentaje de las células tumorales en una muestra. Similarmente, Bhattacharjee y otros, 2001. PNAS 98, 13790-13795 describe un método para la clasificación de los carcinomas de pulmón mediante perfiles de expresión. Además Bhattacharjee y otros, no describe que el perfil de expresión se puede usar para determinar el porcentaje de células tumorales. XP002513552 (Agilent Technologies 2004: "Agilent 012391 Whole Genome Oligo microarray" indica que la matriz del genoma humano completo 44K de Agilent comprende sondas para la identificación de 41 000 genes y transcritos únicos.

35 Por lo tanto, existe una necesidad de un método de puntuación del porcentaje del tumor que se puede usar adicionalmente, o en lugar de, los análisis histopatológicos.

40 La invención proporciona un método para determinar el porcentaje de células tumorales de mama en una muestra de células de un individuo, el método que comprende preparar el ARN a partir de una muestra de células de dicho individuo, dicha muestra de células que comprende las células tumorales de mama o sospechas de comprender células tumorales de mama; determinar los niveles de ARN para un conjunto de genes en dicha muestra de ARN; y determinar el porcentaje de células tumorales en dicha muestra de células sobre la base de los niveles de ARN determinados para el conjunto de genes, por lo cual el conjunto de genes comprende AK094860, identificado por la sec. con núm. de ident.: 3, y ENST00000300458, identificado por sec. con núm. de ident.: 14.

45 Dicho método permite determinar un porcentaje de las células tumorales (TCP) de una muestra de células, que es una medida de la cantidad de células tumorales en relación con la cantidad total de células en dicha muestra celular. Esta determinación complementa o reemplaza una determinación histopatológica. Sólo las muestras con un suficiente TCP se pueden usar para los análisis posteriores costosos y que consumen tiempo tales como los pronósticos basados en micromatrices, que ahorrarán tiempo y dinero.

50 Un método de la invención mejorará adicionalmente la eficiencia y productividad, ya que no requiere del análisis histopatológico para la puntuación inicial del porcentaje del tumor. Adicionalmente, un método de la invención permite el diagnóstico de la transcripción de pequeñas muestras clínicas, tales como aspirados con agujas finas.

55 Un método preferido de acuerdo con la descripción permite la discriminación entre una muestra que comprende un porcentaje mínimo específico de las células tumorales, y una muestra con puntuaciones más abajo del porcentaje mínimo, tales como la discriminación entre una muestra que comprende al menos el cincuenta por ciento de células tumorales (TCP ≥ 50), y una muestra que comprende menos del 50 por ciento de células tumorales (TCP < 50). Un método preferido de la invención permite la discriminación entre una muestra que comprende al menos treinta por ciento

de células tumorales (TCP ≥ 30), y una muestra que comprende menos del 30 por ciento de las células tumorales (TCP < 30).

5 Dicha muestra puede ser cualquier muestra que puede obtenerse a partir de un paciente que sufre a partir de, o se espera que sufra de, un tumor. Tumor, como se usa en esta solicitud se define como una neoplasia benigna o, preferido, maligna (cáncer). Un cáncer más preferido es un cáncer de mama tal como, pero sin limitarse a, carcinoma ductal y carcinoma lobular

10 Por lo tanto, una muestra preferida para una determinación del porcentaje de células tumorales de acuerdo con la invención es el tejido mamario, que comprende un tumor de mama o se sospecha que padece de este.

15 Una muestra preferida es una biopsia. Se prefiere que las biopsias tengan una profundidad de casi 10 milímetros, con mayor preferencia de casi 5 milímetros, y un diámetro de aproximadamente 2 milímetros, aproximadamente 3 milímetros, aproximadamente 4 milímetros, aproximadamente 5 milímetros, aproximadamente 6 milímetros, aproximadamente 7 milímetros, aproximadamente 8 milímetros, aproximadamente 9 milímetros, o aproximadamente 10 milímetros. Sin embargo, otras formas que son iguales en volumen también son posibles.

20 Una muestra preferida adicional se prepara a partir de una biopsia por aspiración con aguja, que es un procedimiento por el cual se inserta una aguja delgada en un tejido para extraer las células.

25 Una muestra puede procesarse en numerosas maneras, como es conocido para una persona experta. Por ejemplo, ellas se pueden preparar frescas a partir de células o tejidos en el momento de la cosecha, o se pueden preparar a partir de biopsias quirúrgicas que se almacenan a -70°C hasta su procesamiento para la preparación de la muestra. Alternativamente, las biopsias de tejidos o quirúrgicas se pueden almacenar bajo condiciones protectoras que conservan la calidad del ARN. Ejemplos de estas condiciones conservantes son la fijación usando, por ejemplo, formalina, inhibidores de ARN tales como ARNsin (Pharmingen) ARNlater (Ambion) o ARNsecure (Ambion). Alternativamente, las sales específicas, tales como una sal que comprende un sulfato amónico o sulfato de cesio se pueden añadir a la muestra como un agente conservante de ARN.

30 El ARN puede prepararse a partir de la muestra fresca, congelada o conservada usando los métodos y medios conocidos. La ruptura celular tiene lugar, preferentemente, en una solución desnaturalizante fuerte que comprende isotiocianato de guanidinio, cloruro de litio, dodecilsulfato de sodio, y/o fenol. El ARN total posteriormente se aísla por uno de varios métodos que incluyen la precipitación con alcohol o LiCl, ultracentrifugación, por ejemplo, a través de la extracción con cloruro de cesio, fenol-cloroformo o alternativas, conocidas por lo tanto, en la materia, o por absorción a una matriz de afinidad, tales como por ejemplo matrices o filtros basados en sílice o vidrio, o cualquier combinación de éstos.

35 El ARN mensajero (ARNm) se aísla usando cromatografía de afinidad para inmovilizar los oligo(dT) u oligo (dU), o por la eliminación específica del ARN ribosomal. Si es necesario, el ADNC puede generarse a partir del ARN total o ARNm usando uno o más iniciadores específicos tales como un iniciador basado en oligo(dT), y/o iniciadores aleatorios tales como iniciadores hexámicos e iniciadores nonámicos.

40 Como se describe, el porcentaje de células tumorales se determina sobre la base de los niveles de ARN determinados para un conjunto de genes que comprende al menos dos de los genes enumerados en la Tabla 1.

45 Los niveles de ARN se pueden determinar por cualquier método conocido para una persona experta, que incluyen, pero sin limitarse a la transferencia (de tipo) Northern, análisis de membrana de transferencia puntual, PCR cuantitativa, análisis de protección de ARN, tecnología de ADN ramificado (Breen et al. (1997) Cell Immunol 178:91-98), y análisis de micromatrices.

50 El análisis de micromatrices proporciona un método preferido, que implica el uso de biomoléculas seleccionadas que se inmovilizan sobre una superficie. Una micromatriz comprende generalmente, moléculas de ácido nucleico denominadas sondas, que son capaces de hibridar a los productos de expresión de ácidos nucleicos. Las sondas se exponen a la muestra de ácido nucleico marcado, se hibridan, y se determina la abundancia de productos de expresión de ácidos nucleicos en la muestra que son complementarios a una sonda. Las sondas en una micromatriz pueden comprender secuencias de ADN, secuencias de ARN, o secuencias de copolímero de ADN y ARN. Las sondas pueden comprender también, análogos de ADN y/o ARN tales como, por ejemplo, análogos de nucleótidos o moléculas de ácido nucleico péptido (PNA), o combinaciones de éstos. Las secuencias de las sondas pueden ser fragmentos completos o parciales de ADN genómico. Las secuencias de la sonda son preferentemente secuencias de nucleótidos sintéticos, tales como secuencias de oligonucleótidos sintéticos.

60 Se proporciona, un método preferido adicional mediante la reacción en cadena de la polimerasa cuantitativa (qPCR). La qPCR se realiza mediante mediciones de punto final, en la cual se determina la cantidad de un producto final, o

preferentemente mediante PCR en tiempo real (rtPCR), en el cual se controla la cantidad del producto durante la reacción.

5 Como conoce una persona experta, la rtPCR se puede realizar ya sea por el uso de un intercalador de ácido nucleico, tal como por ejemplo bromuro de etidio u otro colorante fluorescente de intercalación, tal como, por ejemplo, CyberGreen, que interactúa con los productos bicatenarios que se generan en el proceso resultando en un aumento en la fluorescencia durante la amplificación, o por el uso de sondas marcadas que reaccionan específicamente con el producto bicatenario generado del gen de interés. Los métodos de detección alternativos que se pueden usar se proporcionan mediante la amplificación de la señal del dendrímero, amplificación de la señal de hibridación, y balizas moleculares.

15 Un técnico con experiencia conoce diferentes métodos de amplificación y pueden emplearse para la qPCR, incluyendo, pero sin limitarse a la PCR clásica que emplea una transcriptasa inversa y una ADN polimerasa, la amplificación por círculo rodante, la amplificación basada en secuencia de ácido nucleico, la amplificación mediada por la transcripción, y la amplificación de ARN lineal.

20 Se prefiere que múltiples productos de expresión génica de ácido nucleicos se pueden detectar simultáneamente. Por lo tanto, se prefieren los métodos tales como la amplificación dependiente de la unión múltiple a la transcriptasa inversa, que cuantifica con precisión múltiples transcritos de interés en un ensayo de un tubo (Eldering y otros, Nucleic Acids Res 2003; 31: e153), y la amplificación por rtPCR en un formato de múltiples pocillos.

25 Se prefiere además que el conjunto de genes que comprende al menos dos de los genes enumerados en la Tabla 1, más preferidos al menos tres de los genes enumerados en la Tabla 1, más preferidos al menos cuatro de los genes enumerados en la Tabla 1, más preferidos al menos cinco de los genes enumerados en la Tabla 1, más preferidos al menos seis de los genes enumerados en la Tabla 1, más preferidos al menos siete de los genes enumerados en la Tabla 1, más preferidos al menos ocho de los genes enumerados en la Tabla 1, más preferidos al menos nueve de los genes enumerados en la Tabla 1, más preferidos al menos diez de los genes enumerados en la Tabla 1, más preferidos al menos once de los genes enumerados en la Tabla 1, más preferidos al menos doce de los genes enumerados en la Tabla 1, más preferidos al menos trece de los genes enumerados en la Tabla 1, más preferidos de al menos catorce de los genes enumerados en la Tabla 1, más preferidos al menos quince de los genes enumerados en la Tabla 1, más preferidos al menos dieciséis de los genes enumerados en la Tabla 1, más preferidos al menos diecisiete de los genes enumerados en la Tabla 1, más preferidos al menos dieciocho de los genes enumerados en la Tabla 1, más preferidos al menos diecinueve de los genes enumerados en la Tabla 1, más preferidos al menos veinte de los genes listado en la Tabla 1, más preferidos al menos el veinticinco de los genes enumerados en la Tabla 1, más preferidos al menos treinta de los genes enumerados en la Tabla 1, más preferidos todos los genes enumerados en la Tabla 1.

35 Se prefiere además, que un conjunto de genes que comprende al menos dos de los genes enumerados en la Tabla 1 se seleccione aleatoriamente de los genes enumerados en la Tabla 1.

40 Se prefiere más aún, que un conjunto de genes de acuerdo con la invención que comprende al menos dos de los genes enumerados en la Tabla 1 comprende los genes que están clasificados con el número 1 y 2, como se indica en la Tabla 1, cuyos genes se identifican por la sec. con núm. de ident.: 3 y sec. con núm. de ident.: 14.

45 La validación cruzada (VC) se usó para clasificar el orden de los genes de acuerdo con su desempeño en la VC. El orden de clasificación de un gen, por lo tanto, como se proporciona en la Tabla 1, es una medida de la correlación con el TCP patológicamente determinado a través de las muestras tumorales de un gen. Un gen con la más alta correlación se puso en la posición 1 (posiciones ordenadas por clasificación se proporcionan en la columna titulada "Clasificación" de la Tabla 1).

50 Más preferido es un conjunto de genes que comprende al menos trece de los genes enumerados en la Tabla 1 que se seleccionan aleatoriamente de los genes que enumerados en la Tabla 1. Más preferido es un conjunto de al menos trece genes que se clasifican 1-13 y comprende todos los genes identificados por la sec. con núm. de ident.: 1, sec. con núm. de ident.: 2, sec. con núm. de ident.: 3, sec. con núm. de ident.: 9, sec. con núm. de ident.: 13, sec. con núm. de ident.: 14, sec. con núm. de ident.: 15, sec. con núm. de ident.: 20, sec. con núm. de ident.: 23, sec. con núm. de ident.: 24, sec. con núm. de ident.: 25, sec. con núm. de ident.: 28, y sec. con núm. de ident.: 31, como se indica en la Tabla 1.

60 Aún más preferido es un conjunto de al menos trece genes que comprenden genes identificados por la sec. con núm. de ident.: 1; sec. con núm. de ident.: 2, sec. con núm. de ident.: 3, sec. con núm. de ident.: 5, sec. con núm. de ident.: 9, sec. con núm. de ident.: 14, sec. con núm. de ident.: 16, sec. con núm. de ident.: 18, sec. con núm. de ident.: 24, sec. con núm. de ident.: 25, sec. con núm. de ident.: 30, sec. con núm. de ident.: 32, y sec. con núm. de ident.: 33.

El nivel de expresión de todos los genes en dicho conjunto de trece genes establemente expresados son todos dependientes del porcentaje de células tumorales en las muestras, y se incrementa en una muestra con un porcentaje superior de células tumorales. Este clasificador gen-13 constituye un perfil de expresión de porcentaje de células

tumorales robusto con un alto desempeño que resulta en una clasificación exacta de una muestra con un porcentaje bajo de células tumorales (TCP <30) y una clasificación exacta de una muestra con un alto porcentaje de células tumorales (TCP ≥30).

5 El nivel de expresión de un gen en una muestra se normaliza preferentemente para corregir la diferencia entre diferentes muestras y para corregir el sesgo sistémico del ejemplo. La normalización se puede realizar mediante la corrección del nivel de expresión de un gen en una muestra para el nivel de expresión de uno o más genes de control en la muestra. Los genes de control se seleccionan sobre la base de que su nivel de expresión apenas difiera entre las muestras de diferentes individuos. Los genes de control que se usan frecuentemente para normalizar los datos de la expresión de ARN son genes constitutivos tales como, por ejemplo, ARN ribosomal, gliceraldehído 3-fosfato deshidrogenasa, y beta-actina. Los genes constitutivos conocidos más aun incluyen la proteína ribosomal ácida humana, ciclofilina, fosfoglicerocinasa, β2-microglobulina, 6-glucuronidasa, fosforribosiltransferasa de hipoxantina, proteína de unión del factor de transcripción TATA IID, y el receptor de la transferrina.

15 Por lo tanto, un método preferido de la invención comprende normalizar los niveles de ARN determinados del conjunto de genes en la muestra.

Un conjunto de genes de control que se pueden usar para la normalización de los datos de la expresión, tales como los datos de la expresión basada en micromatrices y que son especialmente adecuados para la muestra de mama se publicó recientemente (WO08039071A1).

20 Un conjunto preferido de genes de control para la normalización de los datos de la expresión del ARN comprende la subunidad 5 gamma, dependiente del voltaje y canal de calcio (CACNG5; NM_014404). Más preferido es un conjunto de genes de control que comprende la caja 1 de CACNG5 y CAAX (CXX1, NM003928). Más preferido es un conjunto de genes de control que comprenden el CACNG5, CXX1 y el receptor tipo inmunoglobulina de células citolíticas, tres dominios, cola citoplasmática 1 larga (KIR3DL1, NM_013289). Más preferido es un conjunto de genes de control que comprende CACNG5, CXX1, KIR3DL1, y fosfolipasa específica a glicosilfosfatidilinositol D1 (GPLD1, NM_001503). Más preferido es un conjunto de genes de control que comprende CACNG5, CXX1, KIR3DL1, GPLD1, y anhidrasa carbónica (CA1, NM_001738).

30 Además, se describe un conjunto de sondas para la determinación del porcentaje de células tumorales en una muestra de un individuo, por lo cual dicho conjunto de sondas comprende sondas específicas para al menos uno de los genes enumerados en la Tabla 1.

35 Una sonda comprende preferentemente una secuencia de entre 18 y 100 nucleótidos que pueden parearse con un producto de la expresión del ARN o derivado de este de un gen enumerado en la Tabla 1. Una sonda más preferida comprende entre 25 y 80 nucleótidos, más preferido aproximadamente 60 nucleótidos. Se entenderá por un experto que el término "nucleótido" incluye un residuo de ácido ribonucleico, un residuo de ácido desoxirribonucleico, un residuo de ácido nucleico modificado y/o químicamente alterado tal como un ácido nucleico bloqueado y un ácido nucleico peptídico, y cualquier combinación de éstos.

Un conjunto preferido de sondas comprende sondas que son específicas para genes identificados por la sec. con núm. de ident.: 3 y sec. con núm. de ident.: 14.

45 Los genes identificados por sec. con núm. de ident.: 3 y sec. con núm. de ident.: 14 tienen el orden de clasificación más alto y por lo tanto, son especialmente adecuados para la determinación del porcentaje de células tumorales en una muestra.

Otro conjunto de sondas preferido comprende sondas que son específicas para al menos trece de los genes que se enumeran en la Tabla 1, en donde se prefiere que el conjunto de sondas comprenda sondas específicas para los genes identificados por la sec. con núm. de ident.: 1; sec. con núm. de ident.: 2, sec. con núm. de ident.: 3, sec. con núm. de ident.: 5, sec. con núm. de ident.: 9, sec. con núm. de ident.: 14, sec. con núm. de ident.: 16, sec. con núm. de ident.: 18, sec. con núm. de ident.: 24, sec. con núm. de ident.: 25, sec. con núm. de ident.: 30, sec. con núm. de ident.: 32, y sec. con núm. de ident.: 33.

55 Además, se prefiere que dicho conjunto de sondas esté presente en una matriz, preferentemente una micromatriz. Como se describe, dicha micromatriz comprende además un conjunto de genes de control que puede usarse para normalizar los datos de expresión. Además, se prefiere que dicha micromatriz comprenda no más de 10.000 sondas, más preferido no más de 5000 sondas, más preferido no más de 1000 sondas, tales como, por ejemplo, 100 sondas o 500 sondas.

60 Se describe además el uso de un conjunto de sondas para la determinación del porcentaje de células tumorales en una muestra de un individuo, por lo cual se prefiere más aun, que el individuo esté padeciendo de un tumor de mama o se sospecha que padece de éste.

Se describe además un conjunto de pares de iniciadores para la determinación del porcentaje de células tumorales en una muestra de un individuo, por lo cual dicho conjunto de pares de iniciadores comprende pares de iniciadores específicos para al menos uno de los genes enumerados en la Tabla 1.

5

Un iniciador comprende una extensión de residuos de nucleótidos corta que pueden aparearse con un producto de expresión de ARN o derivado de un gen enumerado en la Tabla 1. El término par de iniciadores se refiere a dos iniciadores que pueden aparearse con un producto de la expresión de ARN o derivado de este de un gen. Se prefiere que los dos iniciadores de un par de iniciadores hibriden en diferentes posiciones sobre un producto de la expresión del ARN o derivado de éste para formar secuencias bicatenarias que se pueden usar como un sitio de inicio para una polimerasa sintetizar un producto bicatenario que es complementario al menos a una parte del producto de expresión de ARN o derivado de éste. Repetir el procedimiento dará como resultado la amplificación del producto bicatenario. La extensión de residuos de nucleótidos corta de un iniciador determina la especificidad del apareamiento de bases entre un iniciador y el producto de la expresión del ARN complementario o derivado de éste.

15

Los factores que influyen en el apareamiento de bases comprenden el número de residuos de nucleótidos, el contenido de GC, los residuos de nucleótidos de los últimos 5 residuos en los extremos, y la posibilidad de formación de la horquilla del iniciador. Un iniciador preferido comprende entre 15 y 50 residuos de nucleótidos, más preferido entre 16 y 30 residuos, más preferido entre 18 y 25 residuos, tales como 18 residuos, 19 residuos, 20 residuos, 21 residuos, 22 residuos, 23 residuos, 24 residuos, o 25 residuos. El número de residuos de nucleótidos más preferido será determinado por los residuos de nucleótidos individuales.

20

Un par de iniciadores preferentemente se puede usar para obtener una indicación del nivel de expresión de ARN del conjunto de genes en la muestra por amplificación cuantitativa. La cantidad de producto bicatenario que se forma durante la amplificación cuantitativa puede determinarse por colorantes fluorescentes que se pueden intercalar en los productos bicatenarios. Alternativamente, se puede emplear una baliza molecular que comprende un fluoróforo en un extremo y un colorante de desactivación en el extremo opuesto, que se une al objetivo amplificado que resulta en una señal fluorescente que es proporcional a la cantidad de producto bicatenario.

25

Un conjunto preferido de pares de iniciadores comprende pares de iniciadores específicos para al menos dos de los genes enumerados en la Tabla 1.

30

Un conjunto más preferido de pares de iniciadores comprende pares de iniciadores que son específicos para genes identificados por la sec. con núm. de ident.: 3 y sec. con núm. de ident.: 14.

35

Un conjunto más aún preferido de pares de iniciadores comprende pares de iniciadores que son específicos para genes identificados por la sec. con núm. de ident.: 1; sec. con núm. de ident.: 2, sec. con núm. de ident.: 3, sec. con núm. de ident.: 5, sec. con núm. de ident.: 9, sec. con núm. de ident.: 14, sec. con núm. de ident.: 16, sec. con núm. de ident.: 18, sec. con núm. de ident.: 24, sec. con núm. de ident.: 25, sec. con núm. de ident.: 30, sec. con núm. de ident.: 32, y sec. con núm. de ident.: 33.

40

La descripción además contempla el uso de un conjunto de pares de iniciadores para la determinación del porcentaje de células tumorales en una muestra de un individuo, mediante la cual se prefiere además, que el individuo esté padeciendo de un tumor de mama o se sospecha que padece de éste.

45

La descripción contempla además, un kit para determinar el porcentaje de células tumorales en una muestra de un individuo, el kit comprende métodos y medios para determinar el nivel de expresión de al menos dos genes en la muestra. Dichos métodos y medios comprenden una polimerasa, tal como una ADN-polimerasa dependiente de ADN estable a la temperatura, y un tampón de reacción que proporciona condiciones preferidas tales como la concentración de sal y pH para la reacción de amplificación. Los métodos y medios pueden comprender además un colorante fluorescente de intercalación o una baliza molecular que es específica para la detección del producto del conjunto de al menos dos genes.

50

Un estuche preferido comprende pares de iniciadores que permiten la amplificación cuantitativa del conjunto de al menos dos genes de los genes enumerados en la Tabla 1 en la muestra.

55

La descripción contempla además el uso de un kit para determinar el porcentaje de células tumorales en una muestra de un individuo, mediante la cual se prefiere que el individuo esté padeciendo de un tumor de mama o se sospecha que padece de éste.

60

La invención proporciona además, un método para determinar un porcentaje de las células tumorales de mama en una muestra de un individuo que padece de un tumor de mama, o se sospecha que padece de un tumor de mama, el método que comprende clasificar una muestra como derivada de un individuo que tiene un porcentaje de células tumorales más abajo o por encima de un umbral definido por un método que comprende determinar un nivel de ARN

- 5 para un conjunto de genes que comprenden AK094860, identificado por sec. con núm. de ident.: 3, y ENST00000300458, identificado por sec. con núm. de ident.: 14 en una muestra de dicho individuo, determinar un valor de similitud para el nivel de ARN en la muestra y un nivel de ARN para el conjunto de genes en un paciente que no tiene células tumorales, y clasificar dicha muestra como teniendo un porcentaje de células tumorales más abajo del umbral si dicho valor de similitud está más abajo de un primer valor umbral de similitud, y clasificar dicha muestra como que tiene un porcentaje de células tumorales más abajo del umbral si dicho valor de similitud excede dicho primer valor umbral de similitud.
- 10 Dicho umbral definido es una medida para el porcentaje de células tumorales en una muestra de células con relación a la cantidad total de células en dicha muestra celular. Dicho umbral definido es preferentemente 80%, más preferido 70%, más preferido 60%, más preferido 50%, más preferido 40%, más preferido 30%, más preferido 20%, más preferido 10%. Un umbral más preferido es 30%, mediante el cual 30% de la cantidad total de las células comprende células tumorales.
- 15 En un método preferido de la invención, un conjunto adicional de genes comprende la sec. con núm. de ident.: 1; sec. con núm. de ident.: 2, sec. con núm. de ident.: 3, sec. con núm. de ident.: 5, sec. con núm. de ident.: 9, sec. con núm. de ident.: 14, sec. con núm. de ident.: 16, sec. con núm. de ident.: 18, sec. con núm. de ident.: 24, sec. con núm. de ident.: 25, sec. con núm. de ident.: 30, sec. con núm. de ident.: 32, y sec. con núm. de ident.: 33.
- 20 Un método para determinar un porcentaje de las células tumorales en una muestra de células de un individuo que padece de un tumor, o se sospecha que padece de un tumor, se puede usar en un método para clasificar dicha muestra de células, por ejemplo como proporcionado por WO02103320. El porcentaje de células tumorales se usa para la subtipificación de dicha muestra de células en base al porcentaje de células tumorales determinado. Las muestras celulares se pueden subtipar como que comprenden, por ejemplo, porcentajes de células tumorales de entre 10 y 30%,
 25 de entre 30 y 50%, de entre 50 y 70%, y por encima de 70%. Un conjunto específico de al menos 5 marcadores de clasificación para cada subtipo se selecciona posteriormente de los marcadores identificados en cualquiera de las Tablas 1-6 de WO02103320, y se usa para clasificar dicha muestra celular.
- 30 De acuerdo con la descripción, al menos uno de los genes enumerados en la Tabla 1 se usa como un marcador tumoral. Un gen preferido se selecciona de sec. con núm. de ident.: 7 (AQP1 acuaporina 1); sec. con núm. de ident.: 17 (JAM2; molécula de adhesión y unión 2); sec. con núm. de ident.: 20 (LY75; antígeno de linfocitos 75); sec. con núm. de ident.: 23 (NTRK2 receptor neurotrófico tirosina quinasa, tipo 2); sec. con núm. de ident.: 28 (canal de sodio SCN4B, dependiente de voltaje, tipo IV, beta); y sec. con núm. de ident.: 29 (proteína 1 asociada con la SFRP1 frizzled secretada).
- 35 La descripción contempla además un anticuerpo contra uno de los genes seleccionados de la sec. con núm. de ident.: 7 (AQP1 acuaporina 1); sec. con núm. de ident.: 17 (JAM2; molécula de adhesión y unión 2); sec. con núm. de ident.: 20 (LY75; antígeno de linfocitos 75); sec. con núm. de ident.: 23 (NTRK2 receptor neurotrófico tirosina quinasa, tipo 2); sec. con núm. de ident.: 28 (canal de sodio SCN4B, dependiente de voltaje, tipo IV, beta); y sec. con núm. de ident.: 29 (proteína 1 asociada con la SFRP1 frizzled secretada). Los métodos para generar un anticuerpo monoclonal dirigido
 40 contra el producto proteico de uno o más de los genes son conocidos para un experto. Alternativamente, oligoclonos o anticuerpos policlonales se pueden generar usando un método conocido en la técnica. La generación de oligoclonos se describe en, por ejemplo, WO2004009618, WO2004106375, y WO2005068622.
- 45 Un anticuerpo contra uno o más de los genes enumerados en la Tabla 1 se puede usar como un marcador para diagnosticar la presencia de una célula tumoral de mama en una muestra de mama. Además, un anticuerpo contra uno o más de los genes enumerados en la Tabla 1 se puede usar para la preparación de un medicamento para el tratamiento de cáncer, especialmente el cáncer de mama. La combinación del tratamiento mediado por anticuerpos con
 50 otras modalidades de la inmunoterapia del cáncer se describen en, por ejemplo Baxevanis (2008) Expert Opinion on Drug Discovery 3: 441-452.

ES 2 559 757 T3

Tabla 1 Genes usados para preparar un perfil asociado al porcentaje de célula tumoral

SEC. ID	Secuencia.	Nombre del Gen	RANGO	13set	
5	Sec. con núm. de ident.: 1	CCTGTTCTTCCCAACTTGGCTTTCCTTTTCTTTTGGTCATGGGCTCTCAGAGTCTGGG	AI732974	7	Sí
	Sec. con núm. ident.: 2	GCCACCTGTACATCTTTTGTTCCTTATACTGCCTATTTTGTAGAAAGTAGCTAT	AK025430	3	Sí
10	Sec. con núm. ident.: 3	ATGACATCCCCTCTTGTTTTGCCTCTCTTCTCCTGATGCAATGGCCAAATGCTGGAA	AK094860	1	Sí
	Sec. con núm. ident.: 4	ACATTCTGAAATCATTCTCTGTAATGGTTGGATTTCAATTCACCCTTAAAGGGATGC	AK125361	26	No
15	Sec. con núm. ident.: 5	GGCTGTGGTGAAAAAGCAGAACTACTTAGCAGAGAACAGAAATATGAAGATGGAATTGC	ANAPC7	32	Sí
	Sec. con núm. ident.: 6	GACCTGAGAACCTGGACCCTGGAATCAGCATGGATGAGACCAGAGGAGGTATGGATGGAA	ANKRD35	15	No
20	Sec. con núm. ident.: 7	_CCTCTGCATATATGTCTCTTGGAGTTGGAATTCATTATATGTTAAGAAAATAAGGAA	AQP1	14	No
	Sec. con núm. ident.: 8	AGAGGTCTTAACCTAATGCGCATAGAGAAATTGTTCTATTGTAACATACCCTGTCTCT	A358804	17	No
25	Sec. con núm. ident.: 9	ACCTACCACATTTCTCCCAACAAAATAATAATAATAAAGGCTACACAGTATCTGGA	BC031974	8	Sí
	Sec. con núm. ident.: 10	TGGGTGTTTTTGGTTTTTGGTTTCTGGTTTACAATCTCGTCATTCAACAAAGATGGGAG	BC045756	16	No
30	Sec. con núm. ident.: 11	TCTCAGCACAAGAGCGCTICCITIGCACAGAATGAGCITCGAGCIIIGTICAGACTAAAT	CBX7	19	No
	Sec. con núm. ident.: 12	ACTATATCAAGGACATGAAGGAGAACTGAGTGACCCAGAAGGGGTGGCGAAGGCACAGCT	CCL15	34	No
35	Sec. con núm. ident.: 13	AGCGAAATTTGAAGATGACATCACCTATTGGCTTAACAGAGATCGAAATGGACATGAATA	ECRG4	9	No
	Sec. con núm. ident.: 14	TGCAGCCTCCTTTCCCTATCTATAAAATAAAATGACCCTGCTCTATCTCACTGGGcIG	ENST00000300458	2	Sí
40	Sec. con núm. ident.: 16	GTATTGGGGTTCTTGAGCTTGTTAAATTTGCTGCTCCAATCCAGGGTTATTAGGCCA	ENST00000371497	5	No
	Sec. con núm. ident.: 16	TCCAGCCGGTCTTTAAATGAAGATACGTAAGAAGGGAGAGGTAAGTATAGCACAGATC	HEATR3	32	Sí
45	Sec. con núm. ident.: 17	GCCTTGGTGTATGCTATGCTCAGAGGAAAGGCTACTTTTCAAAGAAACCTCCTCCAGA	JAM2	32	No
	Sec. con núm. ident.: 18	ATACAAAGGTTATTGAAATCTGGGCCTTAAAGATGAAGAAGATGGAATCCAGGAAATC	LOC147804	18	Sí
50	Sec. con núm. ident.: 19	TCGCTGCCGACAGAAGTCACTGCCTACCTCAGGGTCCCCTTACCTGGGTGGGAATAAAT	LOC338328	22	No
	Sec. con núm. ident.: 20	TGTGGTTATCACTTTAAGTTTTGACACCTAGATTATAGTCTTAGTAATAGCATCCACTGG	LY75	13	No
55	Sec. con núm. ident.: 21	TATGGAAGCTGTGAAAATCATCACAAGTGCCTCTGAAAGCGAGTGTAGGTTGGTTAGAG	METTL7A	25	No
	Sec. con núm. ident.: 22	GGTAAATGAGAACACIACAACCTGTAGTCAGCTCACAATTTTAAATAAAGGATACCACAG	MYLK	23	No
60	Sec. con núm. ident.: 23	TTCTATACTCTAATCAGCACTGAATTCAGAGGGTTGACTTTTTTCATCTATAACACAGTG	NTRK2	6	No
	Sec. con núm. ident.: 24	ATACCCTCACCTTAAATTTCTCCTGTAACCACTAACAAAATCAAGCCTGATTCAAAA	OGT	11	Sí
	Sec. con núm. ident.: 25	GTGGACAAGAAGATGCAGAAGAAAATGAAGAAAGCTCATAAAAAGATGCACAAGCACCAA	PRR13	10	Sí
	Sec. con núm. de ident. .26	AAGGAGACAACTCCACCTGGAATGTTCTGTGAAGGTCAAGTGTGCAAACAGACATTCC	RBP7	35	No

5	Sec. con ident.: 27	núm.	CAATATGCTTGCCACTCCTTAAATGTCTAATGATGAGAACTCTCTTTCTGACCAATTG	SCARA5	30	No
	Sec. con ident.: 28	núm.	ATGCATGGTTTATTCTCTGGCTTGGATGACAACAATACCCATAGTCAATTTCTATGT	SCN4B	4	No
10	Sec. con ident.: 29	núm.	CGTTTCTCTAGTTTCTCTGTAGTACTCTCTTTTAGATCCTAAGTCTTTACAAAAG	SFRP1	27	No
	Sec. con ident.: 30	núm.	TCCCAGCCCAACAACAGATTGAAAAGAGCTATAAGTACAAGTAAAGAACAGGAAGCAAAG	SNAP29	21	Si
	Sec. con ident.: 31	núm.	GAGGCATTGAGAAAAGGGGAAAGCGGGTATTTTTAAAAGCCAAAGATTGACCCAGTTACT	SNF1LK	12	No
15	Sec. con ident.: 32	núm.	GACCTCATCTCTAAAAAGCGAATAAAAACAGAAAAACACCAAAACAACCAGGATGTCT	TNRC5	28	Si
	Sec. con ident.: 33	núm.	GAGACAGGTATGAAGGTTTTGAAAACCGGCCTTAAAGATGAAGAAAAGATAGAACTC	TPM3	20	Si
20	Sec. con ident.: 34	núm.	CCCACAAGAGCGTATGCAAATCTCTAAGTTTACGGGACTCTCAATGACCACTATCAGTCA	TSHZ2	29	No
	Sec. con ident.: 35	núm.	AGTCTTTTTGGTGAATAGTGGGATGTCTGCTTAGTTGGCAGGGGTCAGTCCAAATGGA	YRDC	24	No

25 Leyendas de las figuras

30 Figura 1. Variación en la puntuación histopatológica del porcentaje de tumor. Dos secciones tumorales coloreadas con HE se clasificaron varias veces por cinco patólogos diferentes (líneas de color) durante un período de 6 meses. Los gráficos de puntos representan la variación patológica global de la puntuación de TCP para cada una de las dos muestras tumorales.

35 Figura 2. Identificación de un perfil transcripcional del porcentaje de tumor. (A) Desempeño del clasificador (área bajo la curva ROC) usando diferentes tamaños de conjunto de genes de mejor clasificación. El desempeño se da para la clasificación de muestras de TCP bajo (<30%) y alto (≥50%). (B) Expresión génica diferencial de los 35 mejores genes clasificados entre las muestras de TCP bajo y alto. Las barras representan la variación de expresión de cada gen a través de 70 muestras de TCP alto. Trece genes más estables diferencialmente expresados para TCP alto se indican en azul. (C) Mapa del calor del perfil de porcentaje de tumor de 13 genes para 95 muestras de entrenamiento. (D) Resultado del perfil para las 165 muestras de entrenamiento. Las muestras tumorales se agrupan de acuerdo con su puntuación patológica y se ordenan de acuerdo con el perfil. La línea horizontal discontinua representa el umbral de clasificación óptima.

45 Figura 3. Figura 3 . Validación del perfil de porcentaje de tumor de 13-genes en 239 muestras independientes. (A) Resultado del perfil para 239 muestras de validación independiente. Las muestras tumorales se agrupan de acuerdo con su puntuación patológica y se ordenan de acuerdo con el índice de perfil. La línea horizontal representa el umbral de clasificación que se determinó sobre la cohorte de entrenamiento. (B) Gráficos de puntos y las diferencias estadísticas en los índices de perfiles entre los grupos de TCP patológico bajo, medio y alto. Colores similares a los de A.

50 Figura 4. El desempeño (AUC) del clasificador medio más cercano a TCP usando diferentes tamaños de conjunto de genes clasificados. Las líneas negras y rojas indican el desempeño clasificador (AUC, eje y) basado en la selección de genes usando 95 muestras (cohorte de entrenamiento 1), y el uso de 95 más 77 muestras (cohorte de entrenamiento 1+2), incluyendo así un criterio de selección adicional. Las líneas horizontales representan el número de genes óptimo seleccionado: 35 basados en el entrenamiento 1, y 13 basados en el entrenamiento 1+2 (ver el ejemplo 1 para más detalles con respecto a las cohortes de entrenamiento).

55 Figura 5. Desempeño (AUC) del clasificador medio más cercano a TCP usando el conjunto de genes aleatorios de diferentes tamaños. El desempeño clasificador de TCP de los subconjuntos aleatorios de diferente tamaño del conjunto total de 35 genes. Para los diferentes tamaños de subconjuntos (2, 3, 4, ...- 35) se calcularon el valor medio y el intervalo de confianza del 95% para el desempeño AUC. Las líneas negras y rojas indican el desempeño clasificador (AUC, eje y) basado en el análisis usando 95 muestras (cohorte de entrenamiento 1), y el uso de 95, más 77 muestras (cohortes de entrenamiento 1+2), incluyendo así un criterio de selección adicional (ver el ejemplo 1 para más detalles con respecto a las cohortes de entrenamiento).

EJEMPLOS

EJEMPLO 1

5 **Materiales y Métodos**

Se realizó un análisis de expresión génica de 95 muestras de tumor de mama usando micromatrices de alta densidad Agilent 44K hechas a la medida. Las muestras incluyeron muestras con un porcentaje de células tumorales (TCP) bajo (<30%, 27 muestras), medio (30-49, 19 muestras) y alto (≥50%, 49 muestras).

10 Siguiendo los datos de pre-procesamiento y la normalización, un enfoque de aprendizaje supervisado se aplicó para diseñar un perfil de expresión de genes para la identificación de muestras con TCP bajo y alto (cohorte de entrenamiento 1). La incertidumbre patológica en la puntuación de células tumorales (+/- 10%, Figura 1) se tuvo en cuenta en el modelo de aprendizaje mediante el ajuste aleatorizado del TCP de entrenamiento durante cada iteración con el porcentaje de -10, 0 o 10.

15 Un método de validación cruzada (CV) 3 veces se usó para identificar un perfil de expresión de genes (clasificador de medio más cercano) que mostró una fuerte asociación con el TCP patológico. Este procedimiento se repitió quinientas veces (enfoque de muestreo múltiple) y los genes se clasificaron de acuerdo con su desempeño de CV.

20 Para identificar el número óptimo de genes que se usarán en el modelo clasificador, el conjunto de genes de mejor rango se amplió gradualmente y el desempeño clasificador (área bajo la curva ROC (AUC), omisión de un CV) se determinó para cada tamaño de conjunto de genes .

25 A continuación, dentro del conjunto óptimo de 35 genes de alta clasificación, se eliminaron los genes que mostraron una gran variación entre un conjunto adicional de 70 muestras tumorales con TCP alta (≥50%) (cohorte de entrenamiento 2). El conjunto restante de 13 genes se usó para construir un clasificación-medio más cercano y ajuste del umbral óptimo y desempeño de todas las muestras de entrenamiento (165 en total) se determinó usando una omisión de CV. El porcentaje del clasificador tumoral fue validado en 2239 muestras de tumores de mama independientes

30 **Resultados**

Se identificó un perfil de expresión de genes del porcentaje de tumor basado en 95 muestras de tumor de mama que se encontraban en el contenido de células tumorales de cero a 85 por ciento. El desempeño de 3 veces la validación cruzada (CV) del muestreo múltiple indicó una fuerte asociación del modelo clasificador medio más cercano seleccionado con el TCP patológico (R2=0.42, Wilcoxon P<0.001 [Wilcoxon, 1945. Biometrics Bulletin 1: 80-83; Bauer, 1972. Journal of the American Statistical Association 67, 687-690]). El mayor desempeño del clasificador se alcanzó bajo la inclusión de los 35 genes mejor clasificados (Figura 2A), aunque el desempeño se mantuvo relativamente estable para los tamaños de conjunto de genes más pequeños.

40 A continuación, se usó un conjunto adicional de 70 muestras tumorales para seleccionar dentro del conjunto de 35 genes que mostraron una expresión estable en muestras con un alto porcentaje de tumor. Como esperado, los genes con una mayor expresión de TCP alto (Figura 2B, lado derecho) mostraron una expresión más estable a través de las 70 muestras adicionales de TCP alto (Figura 2B, barras azules) Los trece genes más establemente expresados, que se regularon negativamente en TCP alto vs bajo (Figura 2B), se seleccionaron para construir un perfil de expresión robusto del porcentaje de tumor (Figura 2C). El clasificador de 13 genes mostró un alto desempeño (AUC 0,92) en la identificación de las muestras que son adecuados para, por ejemplo, pronosticar la muestra de tumor.

50 El empleo del umbral de clasificación óptima para la máxima sensibilidad y especificidad resultó en una clasificación exacta de 85 por ciento para las muestras de TCP bajo y el 93 por ciento de las muestras de TCP alto (Figura 2D).

El perfil del porcentaje de tumor desarrollado se validó en un conjunto independiente de 239 muestras tumorales de mama. La distribución de TCP a través de esta cohorte de validación fue representativa para la recolección estándar de muestras de diagnóstico. El resultado del índice del perfil de 13 genes se asoció significativamente con la puntuación patológica de las muestras de validación, (R2=0.48, Wilcoxon P<0.001, AUC 0.90). El clasificador molecular asignó correctamente 78, 71 y 95 por ciento de las muestras de TCP bajo, medio y alto de acuerdo con la puntuación patológica (Figura 3A) con una precisión global del 91 por ciento y una puntuación Kappa de 0.70 (95% CI 0.54-0.81).

60 El análisis estadístico de los índices del perfil indicó una diferencia significativa en los resultados entre las muestras con un TCP bajo y medio (P = 0.01, prueba T de student) y entre las muestras de TCP bajo y alto (P<0.0001) (Figura 3B). La diferencia entre muestras de TCP medio y alto no es significativa (p=0.02) ya que la mayoría de las muestras de TCP medio se clasificaron correctamente como TCP alto por el perfil molecular. Los resultados de la validación confirman que el perfil molecular puede discriminar con precisión las muestras con un porcentaje de tumor suficiente para pronóstico de la muestra basado en el análisis de transcripción de los 13 genes identificados.

Aunque los perfiles de expresión de genes tienden a ser más robustos después de la inclusión de un conjunto más amplio de genes, se decide limitar el perfil desarrollado a un número relativamente pequeño de genes con un desempeño óptimo. El razonamiento de esta estrategia es que la evaluación transcripcional de TCP es preferentemente realizada antes del análisis de micromatrices y esta evaluación indicará si una muestra es elegible para el perfil de expresión génica.

Este enfoque puede probar ser más rápido y más objetivo que la puntuación del porcentaje de tumor patológico.

EJEMPLO 2

Para determinar el número mínimo de los genes distintivos que se necesitan para una distinción precisa, todos los genes asociados al porcentaje de células tumorales (TCP) se clasificaron de acuerdo con su inclusión distintiva y su correlación con el TCP patológico a través de 95 muestras de entrenamiento (cohorte de entrenamiento 1). Los mejores 35 genes representan el perfil identificado como se describió.

Para investigar un número mínimo de genes de alta clasificación con un desempeño significativo (medido como el área bajo la curva ROC (AUC), que puede encontrarse en el intervalo entre 0.50 (al azar) a 1.00 (100% de precisión)), se reduce posteriormente el conjunto de genes clasificados a 2 y se calcula el AUC para cada tamaño de conjunto de genes (Figura 4, línea negra) El análisis además se realizó en conjuntos de genes clasificados para los que el procedimiento de selección incluyó un criterio adicional basado en un segundo conjunto de 77 muestras (cohorte de entrenamiento 2). Los resultados de este segundo análisis del conjunto de genes, más robusto, se muestran en rojo (Figura 4).

Ambos análisis resultaron en un alto desempeño para diferentes tamaños del conjunto de genes con el óptimo a 35 y 13 genes. La reducción del conjunto de genes resultó en una ligera disminución de la precisión, pero se mantuvo muy por encima del 80 por ciento.

El AUC también se determinó para genes individuales. Se encontró que al menos la sec. con núm. de ident.: 1(0.88), sec. con núm. de ident.: 2 (0.86), sec. con núm. de ident.: 3 (0.83), sec. con núm. de ident.: 5 (0.81), sec. con núm. de ident.: 9 (0.81), sec. con núm. de ident.: 14 (0.84), sec. con núm. de ident.: 16 (0.87), sec. con núm. de ident.: 18 (0.82), sec. con núm. de ident.: 24 (0.8), sec. con núm. de ident.: 25 (0.80), sec. con núm. de ident.: 30 (0.80), sec. con núm. de ident.: 31 (0.81), sec. con núm. de ident.: 32 (0.79), sec. con núm. de ident.: 33 (0.80), y sec. con núm. de ident.: 35 (0.79) pueden usarse como un único determinante para determinar el TCP, de manera que el AUC determinado para cada gen se proporciona dentro de paréntesis.

La inclusión de los 2 mejores genes resultó en una precisión del 83 por ciento y la inclusión de los 10 mejores alcanzó una AUC de 0.89 y 0.91 basado en la cohorte de entrenamiento 1 y 1+2, respectivamente.

EJEMPLO 3

Además, se analizó el desempeño de un subconjunto aleatorio de 2 o más genes seleccionados de un conjunto más amplio de genes asociados a TCP. Todos los genes asociados a TCP se clasificaron de acuerdo con su inclusión distintiva y su correlación con la TCP patológica a través de 95 muestras de entrenamiento (cohorte de entrenamiento 1). Además se analiza el desempeño de un subconjunto aleatorio de 2 o más genes seleccionados de un conjunto de 35 genes. Los subconjuntos se seleccionaron al azar con diferentes tamaños que van desde 2 genes hasta los 35 genes. En total, si es posible, se seleccionaron cientos de subconjuntos generados aleatoriamente por informática, y se calculó el AUC para cada tamaño de conjunto de genes diferentes y para cada subconjunto diferente. Posteriormente, se calcularon el desempeño medio y el intervalo de confianza del 95 por ciento para cada tamaño diferente de subconjunto. Los datos mostrados en la Figura 5 indican que los subconjuntos aleatorios de dos o más genes asociados a TCP mejores clasificados muestran sólo un descenso marginal de la clasificación del desempeño (Figura 5). Este resultado confirmó que el desempeño del perfil de TCP no cae sustancialmente, también en los casos en que sólo un pequeño número de genes se usan dentro de la distinción pronóstica. Sin embargo, el intervalo de confianza del 95% de los desempeños predictivos hse incrementa después del uso de conjuntos de perfiles más pequeños. Esto se explica por el hecho de que la selección aleatoria de un pequeño número de genes del conjunto total de 100 genes resultará en una variación mucho mayor en el resultado pronóstico que la selección de un subconjunto grande. A pesar de este aumento en la variación, el desempeño (AUC) del perfil de TCP se mantiene por encima de 0.80 usando al menos 2 genes y por encima de 0.85 usando un mínimo de 5 genes. Estos resultados indican que, aunque el desempeño más alto se logra usando el conjunto descrito de 13 genes, el uso de sólo 2 genes resulta ya en un perfil de TCP con un desempeño significativo AUC por encima de 0.80.

EJEMPLO 4

El ARN a partir de una muestra de tumor se hibrida contra un ARN de referencia de cáncer de mama usando una

micromatriz que contiene las sondas TCP y genes de normalización. La mezcla del ARN de referencia comprenden 105 muestras de cáncer de mama con un TCP patológico de >50%. Las relaciones del pre-procesamiento posterior de los datos (sustracción de fondo local, normalización dependiente de la intensidad usando genes de normalización) se calculan entre la muestra tumoral y de referencia. El índice TCP de una muestra tumoral se determina mediante el cálculo de la correlación coseno de las relaciones de la expresión génica de TCP con el molde de perfil de TCP alto y con el molde de perfil de TCP bajo. Si la correlación coseno de TCP alto supera la correlación del coseno de TCP bajo en un factor predefinido, la muestra será calificada como muestras de alto porcentaje de células tumorales. Si la correlación coseno de TCP bajo supera la correlación del coseno de TCP alto en un factor predefinido, será calificada como muestras de alto porcentaje de células tumorales. Los moldes de perfil de TCP alto y TCP bajo se evaluaron mediante el cálculo de la expresión génica media de una cohorte de muestras de tumores de mama con al menos 50% de células tumorales y basado en una cohorte con menos del 30% de células tumorales, respectivamente. Este procedimiento de clasificación también se conoce como un clasificador medio más cercano (o centroide).

REIVINDICACIONES

1. Un método para determinar el porcentaje de células de tumor de mama en una muestra de células de un individuo, el método comprende
- 5
- a. preparar el ARN a partir de una muestra de células de dicho individuo, dicha muestra de células que comprende células tumorales de mama o con sospecha de comprender células tumorales de mama;
- b. determinar los niveles de ARN para un conjunto de genes en dicha muestra de ARN; y
- 10
- c. determinar el porcentaje de células tumorales en dicha muestra de células basado en los niveles de ARN determinados para el conjunto de genes,
- 15
- por lo que el conjunto de genes comprende AK094860, identificado por sec. con núm. de ident.: 3, y ENST00000300458, identificado por sec. con núm. de ident.: 14.
2. El método de acuerdo con la reivindicación 1, que permite la discriminación entre una muestra de células que comprende al menos el treinta por ciento células tumorales (TCP ≥ 30), y una muestra de células que comprende menos del 30 por ciento de las células tumorales (TCP < 30).
- 20
3. El método de acuerdo con la reivindicación 1 o la reivindicación 2, en donde determinar el porcentaje de células tumorales en una muestra basado en los niveles de ARN determinados para un conjunto de genes que comprende comparar los niveles de ARN de dicho conjunto de genes que comprenden AK094860, identificado por sec. con núm. de ident.: 3, y ENST00000300458, identificado por sec. con núm. de ident.: 14 con los niveles de ARN de dicho conjunto de genes en una muestra de referencia.
- 25
4. El método de acuerdo con cualquiera de las reivindicaciones 1-3, en donde el conjunto de genes comprende los genes identificados por la sec. con núm. de ident.: 1; sec. con núm. de ident.: 2, sec. con núm. de ident.: 3, sec. con núm. de ident.: 5, sec. con núm. de ident.: 9, sec. con núm. de ident.: 14, sec. con núm. de ident.: 16, sec. con núm. de ident.: 18, sec. con núm. de ident.: 24, sec. con núm. de ident.: 25, sec. con núm. de ident.: 30, sec. con núm. de ident.: 32, y sec. con núm. de ident.: 33.
- 30
5. El método de acuerdo con cualquiera de las reivindicaciones anteriores, que comprende además normalizar los niveles de ARN determinados del conjunto de genes en la muestra.
- 35
6. El método para determinar un porcentaje de las células tumorales de mama en una muestra de un individuo que padece de un tumor de mama, o se sospecha que padece de un tumor de mama, el método comprende
- 40
- clasificar una muestra como derivada de un individuo que tiene un porcentaje de células tumorales por debajo o por encima de un umbral definido por un método que comprende
- 45
- determinar un nivel de ARN para un conjunto de genes que comprenden AK094860, identificado por la sec. con núm. de ident.: 3, y ENST00000300458, identificado por la sec. con núm. de ident.: 14 en una muestra de dicho individuo;
- 50
- determinar un valor de similitud para el nivel de ARN en la muestra y un nivel de ARN para el conjunto de genes en un paciente que no tiene células tumorales; y
- 55
- clasificar la muestra como que tiene un porcentaje de células tumorales por encima del umbral, si dicho valor de similitud está por debajo de un primer valor umbral de similitud, y clasificar dicha muestra como que tiene un porcentaje de células tumorales por debajo del umbral si dicho valor de similitud excede dicho primer valor umbral de similitud.
7. El método de acuerdo con la reivindicación 6, en donde el conjunto de genes comprende la sec. con núm. de ident.: 1; sec. con núm. de ident.: 2, sec. con núm. de ident.: 3, sec. con núm. de ident.: 5, sec. con núm. de ident.: 9, sec. con núm. de ident.: 14, sec. con núm. de ident.: 16, sec. con núm. de ident.: 18, sec. con núm. de ident.: 24, sec. con núm. de ident.: 25, sec. con núm. de ident.: 30, sec. con núm. de ident.: 32, y sec. con núm. de ident.: 33.

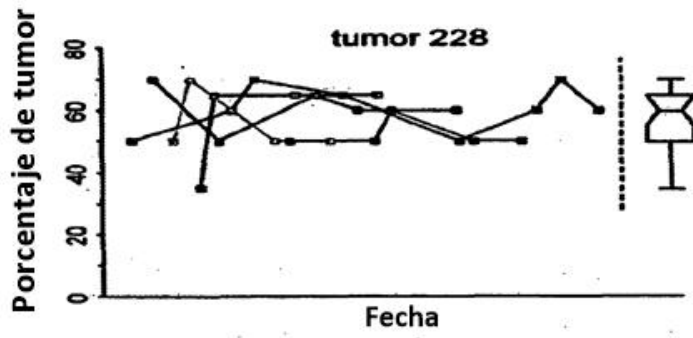


Figura 1A

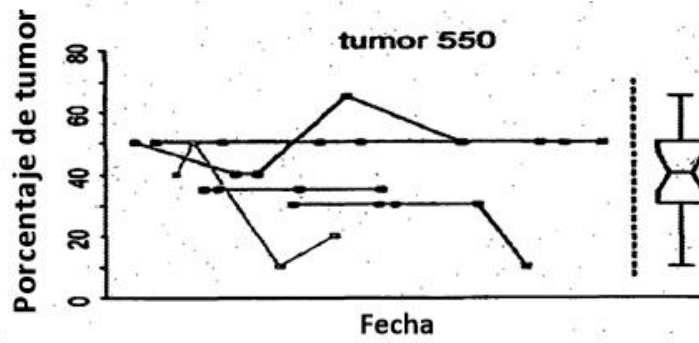


Figura 1B

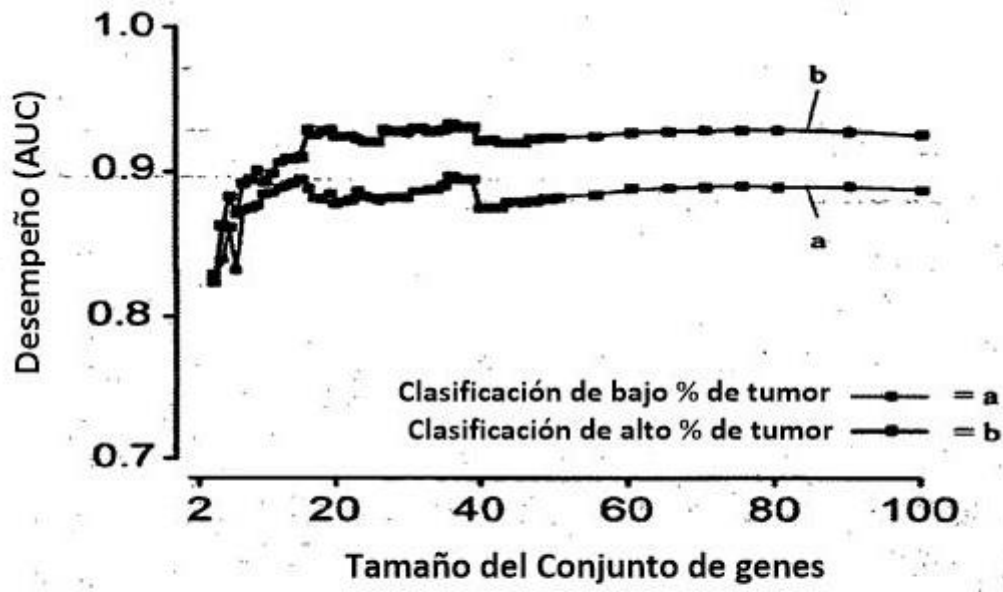


Figura 2A

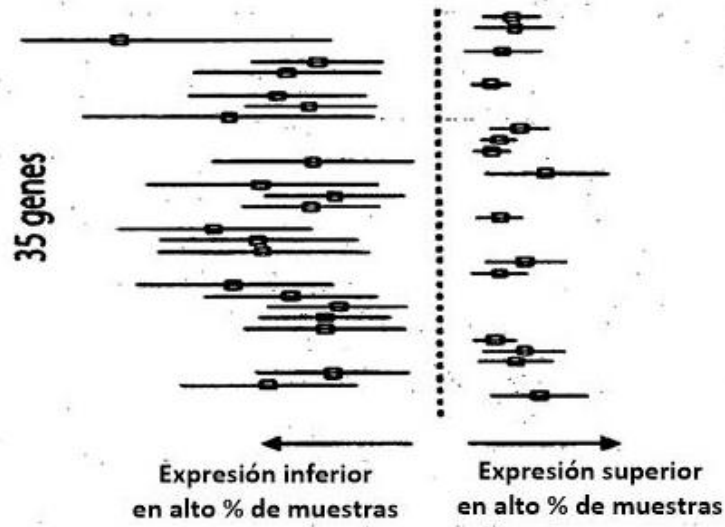


Figura 2B

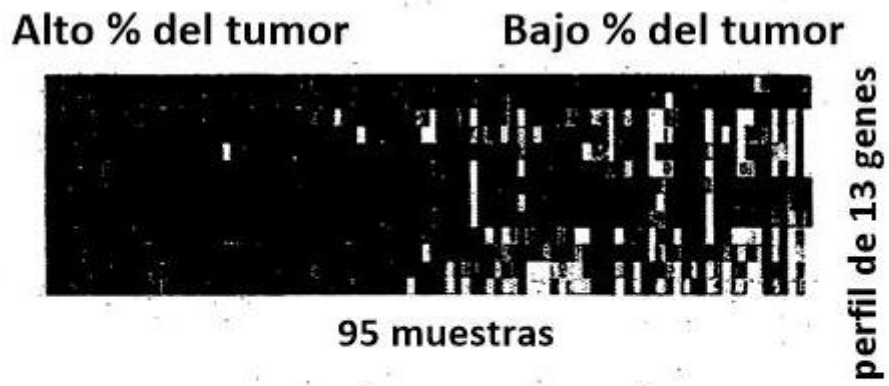


Figura 2C

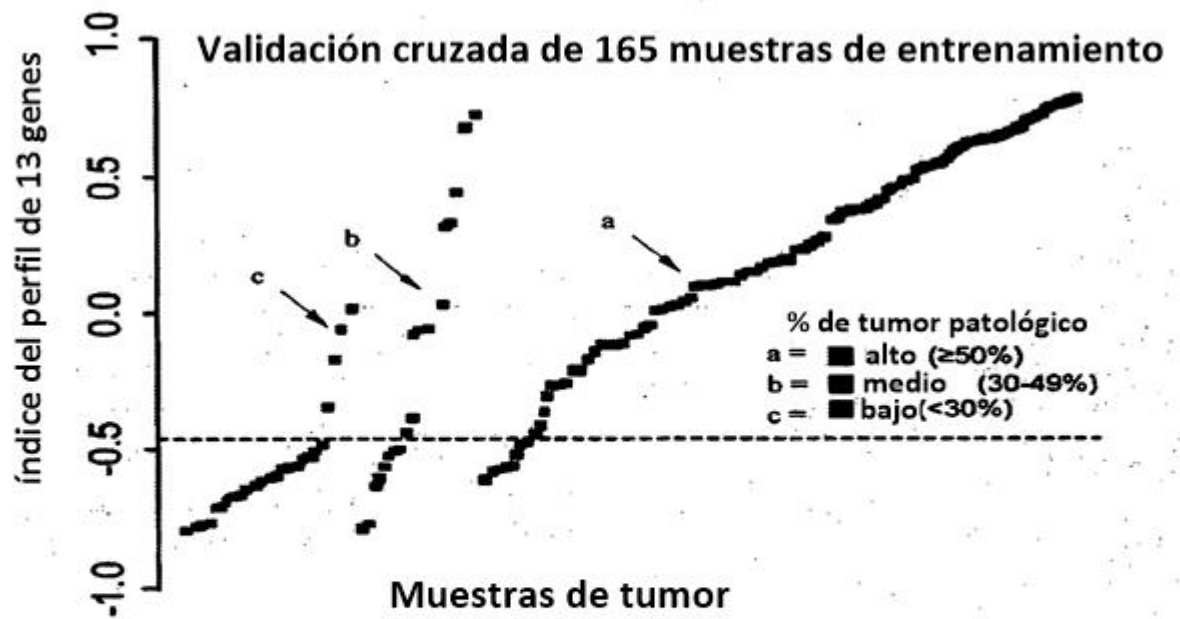


Fig. 2D

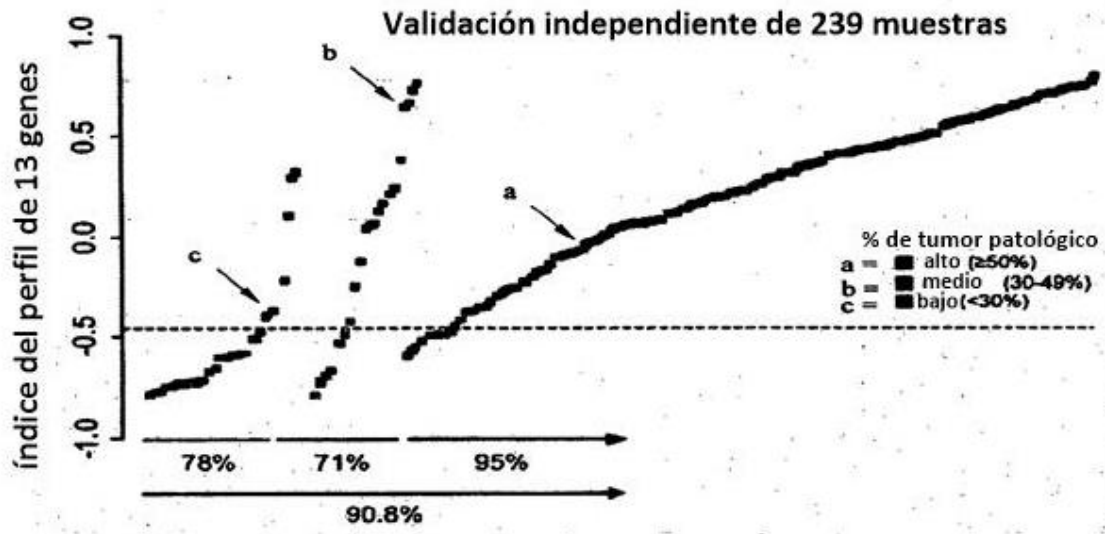


Figura 3A

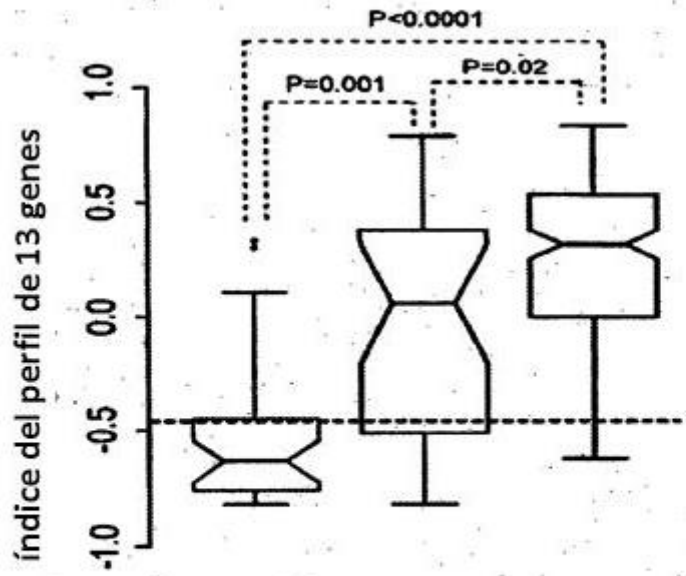


Figura 3B

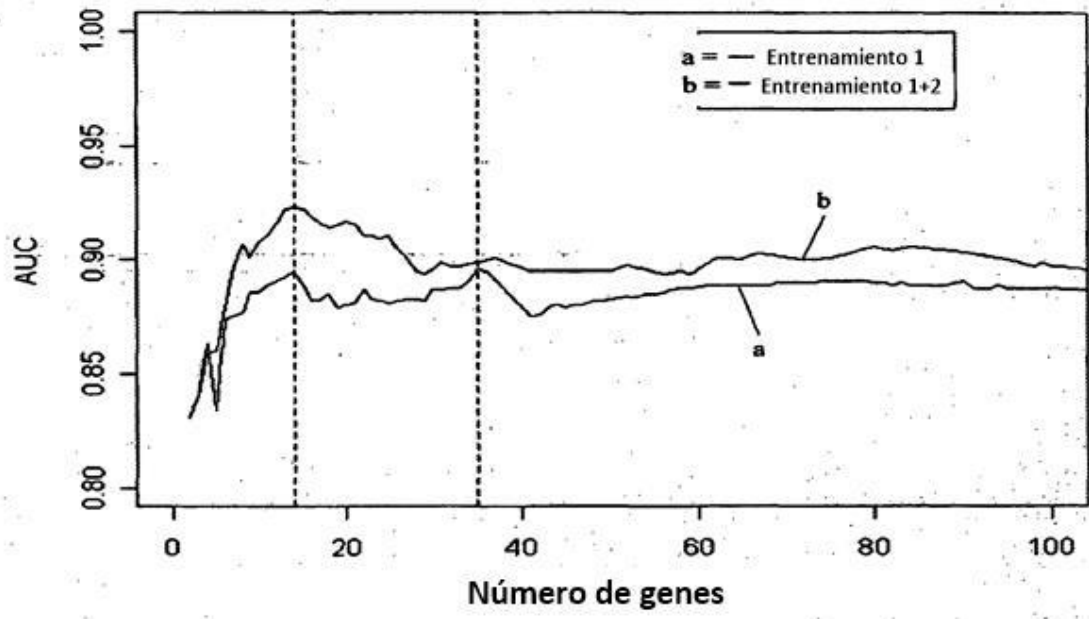


Figura 4

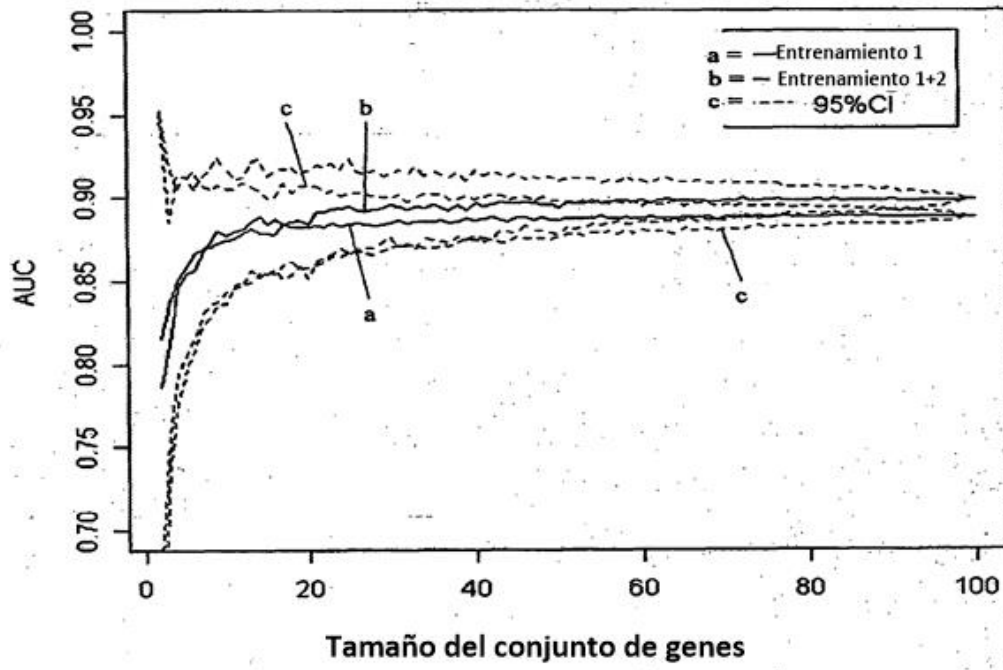


Figura 5