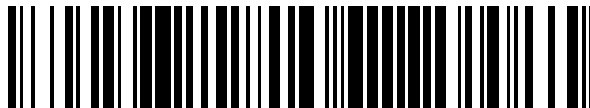


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 717 627**

51 Int. Cl.:

C12Q 1/68 (2008.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **28.07.2006 PCT/US2006/029475**

87 Fecha y número de publicación internacional: **08.02.2007 WO07016367**

96 Fecha de presentación y número de la solicitud europea: **28.07.2006 E 06813244 (8)**

97 Fecha y número de publicación de la concesión europea: **26.12.2018 EP 1920071**

54 Título: **Métodos, kits, sistemas y bases de datos relacionados con enfermedades neoplásicas**

30 Prioridad:

29.07.2005 US 703681 P

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

24.06.2019

73 Titular/es:

**SIEMENS HEALTHCARE DIAGNOSTICS INC.
(100.0%)
511 Benedict Avenue
Tarrytown, NY 10591, US**

72 Inventor/es:

**WIRTZ, RALPH, M.;
AVERDICK, MANUELA;
BRÜCKL, WOLFGANG;
WEIN, AXEL y
THIEL, ROBERT, P.**

74 Agente/Representante:

CARVAJAL Y URQUIJO, Isabel

ES 2 717 627 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).

DESCRIPCION

Métodos, kits, sistemas y bases de datos relacionados con enfermedades neoplásicas

Campo de la invención

5 En una realización, la invención proporciona métodos para predecir un resultado clínico relacionado con un paciente que padece o tiene riesgo de desarrollar una enfermedad neoplásica, que comprende: (a) determinar un valor predictivo mediante el uso de valores de pacientes para (1) al menos un marcador seleccionado del grupo que consiste en marcadores tumorales, marcadores inmunes y marcadores de fase aguda, y (2) al menos un marcador que es (i) un marcador de matriz extracelular (ECM), (ii) un marcador que es indicativo de la síntesis de matriz extracelular (fibrogénesis), o (iii) un marcador que es indicativo de la degradación de la matriz extracelular (fibrólisis); y (b) predecir el resultado clínico de la enfermedad neoplásica mediante la evaluación del valor predictivo.

Antecedentes de la invención

El cáncer colorrectal (CCR) es el segundo tipo de cáncer más prevalente y es la segunda causa de muerte relacionada con el cáncer en los países industrializados de Occidente. Aproximadamente 50000 nuevos casos de CCR se diagnostican anualmente solo en Alemania.

15 Alrededor del 75% de los pacientes diagnosticados de CCR se someten a un tratamiento curativo. La supervivencia a largo plazo de los pacientes con CCR depende del estadio del tumor y del desarrollo potencial de metástasis a distancia sincronas o metacrónicas. La tasa de supervivencia a los 5 años de los pacientes con CCR supera el 90% en el estadio I de la UICC (invasión limitada sin metástasis a los ganglios linfáticos regionales), pero disminuye a menos del 20% en el estadio IV de la UICC (presencia de metástasis a distancia). Se recomienda la radioquimioterapia neoadyuvante en el cáncer rectal UICC estadio II y III y la quimioterapia adyuvante en el estadio III UICC que se agrega para prevenir las recurrencias locorregionales en el cáncer rectal y las recidivas distantes en el cáncer de colon. Sin embargo, estas estrategias son menos efectivas para prevenir la recurrencia a distancia en el cáncer rectal y no se recomienda la quimioterapia adyuvante (fuera de los estudios clínicos) en el cáncer colorrectal resecado con R0 que se presenta en el estadio IV de la UICC en el momento del diagnóstico. La quimioterapia puede conducir a una remisión parcial de metástasis a distancia, y puede permitir cirugías curativas secundarias y, por lo tanto, dar como resultado una supervivencia a largo plazo (cinco años de supervivencia general) de alrededor del 30%. Aproximadamente 25.000 pacientes con cáncer colorrectal metastásico reciben quimioterapia paliativa en Alemania cada año. Se han logrado tasas de respuesta de hasta el 50% mediante la aplicación de regímenes de quimioterapia modernos como 5-Fluorouracil (5-FU), ácido folínico (FA), irinotecan y oxaliplatino. Para hasta el 15% de los pacientes con metástasis no resecables antes de la quimioterapia, es posible una resección secundaria de metástasis hepáticas o pulmonares R0 y conduce a la supervivencia a largo plazo. Las decisiones clínicas sobre el procedimiento terapéutico y la extensión del tratamiento de resección en el carcinoma colorrectal se basan actualmente en las imágenes y en las características histopatológicas convencionales. La precisión diagnóstica de estos enfoques es limitada, lo que conduce a intervenciones quirúrgicas que a menudo son más radicales de lo necesario, o al tratamiento quimioterapéutico de pacientes que no se benefician de este régimen riguroso.

A medida que el CCR avanza, puede hacer metástasis en el hígado y disminuir las posibilidades de supervivencia del paciente. De hecho, las metástasis hepáticas son una causa importante de mortalidad en pacientes con cáncer colorrectal. Sin embargo, hasta la fecha, no se ha realizado un análisis detallado de cómo las células tumorales invaden el hígado y de la interacción de las células tumorales diseminadas en el hígado con el tejido hepático no neoplásico circundante.

La evaluación de la gravedad y la progresión de la enfermedad cancerosa es difícil, y la mayoría de las veces implica la realización de una biopsia. La biopsia implica posibles complicaciones clínicas y dificultades tecnológicas. Además, el muestreo en serie para evaluar la efectividad temprana del tratamiento y elaborar tecnologías de imágenes (por ejemplo, tomografía computarizada), clínicamente no es factible para el uso de rutina. En consecuencia, el desarrollo de métodos menos invasivos y costosos, que identifican regímenes efectivos antes o poco después del primer tratamiento, es de alto valor clínico. El análisis de los factores predictivos llevaría a una terapia individualizada adaptada al tumor con un aumento en la respuesta a la quimioterapia y la supervivencia y una disminución de la toxicidad y los valores económicos.

50 Hanke, et al., British Journal of Cancer (2003) 88, 1248-50 ("Hanke"), describe que los niveles de prueba de los niveles séricos de colágeno (IV) y (VI), tenascina-C, MMP-2, el complejo MMP-9/TIMP-1 y el TIMP-1 libre tomados de pacientes que padecen cáncer colorrectal metastásico en el hígado. Hanke concluye que la MMP-2 en suero parece reflejar la reabsorción del tumor, mientras que la TIMP-1 en suero puede reflejar la expansión del tumor.

El Documento de Solicitud de Patente de los Estados Unidos No. 20030219842 describe un método para controlar la progresión de la enfermedad o la efectividad del tratamiento del cáncer en un paciente con cáncer midiendo el nivel del dominio extracelular (ECD) del receptor del factor de crecimiento epidérmico (EGFR) en una muestra tomada del paciente con cáncer, preferiblemente antes del tratamiento, al inicio del tratamiento y en varios intervalos de tiempo durante el tratamiento, en donde una disminución en el nivel de la ECD del EGFR en el paciente con cáncer en

comparación con el nivel de la ECD del EGFR en individuos de control normal sirve como un indicador de avance o progresión del cáncer y/o falta de efectividad del tratamiento para el paciente.

5 El Documento de Solicitud de Patente de los Estados Unidos No. 20030180819 describe un método para monitorizar la progresión de la enfermedad, o la eficacia del tratamiento del cáncer, en un paciente con cáncer midiendo los niveles de uno o más analitos del sistema activador del plasminógeno (uPA), a saber, uPA, PAI-1 y el complejo de uPA:PAI-1, en una muestra tomada del paciente con cáncer, preferiblemente, antes del tratamiento, al inicio del tratamiento y en diversos intervalos de tiempo durante el tratamiento.

10 El Documento de Solicitud de Patente de los Estados Unidos No. 20040157278 describe un método para detectar la presencia de cáncer colorrectal en un individuo, en donde: el cáncer colorrectal se detecta al detectar la presencia de moléculas de ácido nucleico o aminoácidos Reg1 α o TIMP1 en una muestra clínica obtenida del paciente y la expresión de Reg1 α o TIMP1 es indicativa de la presencia de cáncer colorrectal.

15 El Documento de Solicitud de Patente de los Estados Unidos No. 20040146921 describe un método para proporcionar un diagnóstico de paciente para cáncer de colon, que comprende las etapas de: (a) determinar el nivel de expresión de uno o más genes o productos génicos en una primera muestra biológica tomada del paciente; (b) determinar el nivel de expresión de uno o más genes o productos génicos en al menos una segunda muestra biológica tomada de una muestra de paciente normal; y (c) comparar el nivel de expresión de uno o más genes o productos génicos en la primera muestra biológica con el nivel de expresión de uno o más genes o productos génicos en la segunda muestra biológica; en donde un cambio en el nivel de expresión de uno o más genes o productos génicos en la primera muestra biológica en comparación con el nivel de expresión de uno o más genes o productos génicos en la segunda muestra biológica es un diagnóstico de la enfermedad.

20

25 El Documento de Solicitud de Patente de los Estados Unidos No. 20040146879 describe secuencias de ácido nucleico y proteínas codificadas de este modo, así como sondas derivadas de las secuencias de ácido nucleico, anticuerpos dirigidos a las proteínas codificadas y métodos de diagnóstico y pronóstico para detectar y controlar el cáncer, especialmente cáncer de colon. Se ha encontrado que las secuencias descritas en el Documento de Solicitud de Patente de los Estados Unidos No. 20040146879 se expresan diferencialmente en muestras obtenidas a partir de líneas celulares de cáncer de colon y/o tejido de cáncer de colon.

La Patente de los Estados Unidos No. 6,262,333 describe secuencias de ácido nucleico y proteínas codificadas de este modo, así como sondas derivadas de las secuencias de ácido nucleico, anticuerpos dirigidos a las proteínas codificadas y métodos de diagnóstico para detectar células cancerosas, especialmente células de cáncer de colon.

30 El documento US 2004/0018525 A1 describe un método que puede detectar múltiples marcadores de diagnóstico y pronóstico para enfermedades neoplásicas simultáneamente. En el método de predicción, diagnóstico o pronóstico de neoplasia maligna, se detectan al menos dos marcadores. Los marcadores son genes y fragmentos de los mismos o secuencias de ácido nucleico genómico que están ubicadas en una región cromosómica que está alterada en la neoplasia maligna. Alternativamente, los marcadores son a) genes y b) receptor y ligando, o miembros de la misma vía de transducción, o miembros de la vía de transducción de señales sinérgica, o miembros de vías de transducción de señales antagonistas, o sitio de unión del factor de transcripción y factor de transcripción.

35

40 El documento US 2003/0082652 A1 describe un método para monitorizar pacientes con cáncer colorrectal para la persistencia de la enfermedad residual mínima (MRD) o para la recurrencia de su enfermedad de cáncer. En el método, la cantidad de TIMP-1 total, complejado y/o libre en las muestras de pacientes se determina como un primer parámetro. En una realización preferida, se determina al menos un segundo parámetro diferente de cualquier forma de TIMP-1. A partir del primer y segundo parámetro, se calcula un parámetro combinado que indica la probabilidad de que una persona tenga MRD y/o cáncer recurrente. El segundo parámetro representa preferiblemente la concentración de un marcador tumoral que se selecciona del grupo que consiste en CEA, u-PAR soluble, catepsina B, catepsina H, HER2-neu, CA 15-3 y YKL 40, CA 19-9 y CA 242.

45 En el artículo de N. Yukawa et al., Plasma concentration of tissue inhibitor of matrix metalloproteinase 1 in patients with colorrectal carcinoma, en: Brit. J. of Surgery 88 [2001] 1596-1601, se describe un método para determinar el nivel plasmático del inhibidor tisular de la metaloproteinasas de matriz (TIMP-1) en muestras de pacientes con carcinoma colorrectal que se sometieron a resección del tumor primario. Se encontró que el nivel de TIMP-1 en plasma se asoció significativamente con la profundidad de la invasión y la metástasis en los ganglios linfáticos y el hígado. También se determinó la concentración sérica de antígeno carcinoembrionario (CEA) y de antígeno de carbohidrato 19-9 (CA 19-9). CEA y CA 19-9 son ambos marcadores tumorales.

50

55 A pesar de los métodos de diagnóstico, predicción y pronóstico descritos anteriormente, sigue existiendo la necesidad de contar con métodos predictivos mejorados que faciliten una evaluación precisa y asequible acerca de si un paciente responderá de manera positiva a un régimen de tratamiento contra el cáncer en particular. Los pacientes con cáncer no pueden permitirse el tiempo y los efectos adversos asociados con la selección actual de la terapia de prueba y error y las biopsias imprecisas y de riesgo.

Los marcadores predictivos confiables para una respuesta de quimioterapia conducirían a una terapia individualmente adaptada, y aumentarían el resultado beneficioso (por ejemplo, la mediana general o el tiempo de supervivencia libre

de progresión) y la tasa de resección metastásica curativa secundaria. Sin embargo, hasta la fecha, no se han validado suficientemente dichos marcadores predictivos en el entorno paliativo.

Resumen de la invención

5 En una realización, la invención proporciona métodos implementados por ordenador para predecir un resultado clínico relacionado con un paciente que padece una enfermedad neoplásica, a saber, cáncer colorrectal, que comprende: (a) determinar un valor predictivo algorítmicamente utilizando valores de muestra del paciente para (1) al menos un marcador seleccionado del grupo que consiste en marcadores tumorales, marcadores inmunes y marcadores de fase aguda, y (2) al menos un marcador que es (i) un marcador de matriz extracelular (ECM) (ii) un marcador que es indicativo de síntesis de matriz extracelular (fibrogénesis), o (iii) un marcador que es indicativo de degradación de matriz extracelular (fibrólisis); y (b) predecir el resultado clínico del cáncer colorrectal mediante la evaluación del valor del predictor, en donde el marcador de (1) y el marcador de (2) no son los mismos, en donde el valor del predictor se (1) determina usando un algoritmo derivado por análisis de regresión de Cox y (2) se utiliza para evaluar la probabilidad de que el paciente responda favorablemente a un régimen de tratamiento antineoplásico. Cada uno de los marcadores mencionados anteriormente se define a continuación.

15 "Predecir un resultado clínico relacionado con un paciente que padece o tiene riesgo de desarrollar una enfermedad neoplásica" significa predecir: (1) si un paciente que padece una enfermedad neoplásica responderá a uno o más regímenes de tratamiento de la enfermedad neoplásica; (2) la probabilidad y duración de la supervivencia de un paciente que padece una enfermedad neoplásica; y (3) predecir la probabilidad de que el paciente desarrolle una enfermedad neoplásica y la probable progresión de esa enfermedad neoplásica.

20 "Responder a uno o más regímenes de enfermedad neoplásica" significa que el régimen de tratamiento de la enfermedad es eficaz en el tratamiento de una enfermedad neoplásica. La respuesta se define según la OMS como remisión completa (RC), remisión parcial (RP), no respuesta como enfermedad estable (SD) o enfermedad progresiva (EP) según el tamaño de una lesión indicadora, medida en dos dimensiones.

25 Los valores del predictor se determinan algorítmicamente mediante el análisis de regresión de Cox o mediante el uso de algoritmos de función lineal o no lineal.

30 En otra realización, la divulgación proporciona un método para evaluar el pronóstico de un paciente que padece o está en riesgo de desarrollar una enfermedad neoplásica que comprende evaluar los valores de predicción determinados en uno o más puntos temporales, en donde: (a) los valores de predicción se determinan algorítmicamente utilizando valores de muestra de pacientes para (1) al menos un marcador seleccionado del grupo que consiste en marcadores tumorales, marcadores inmunes y marcadores de fase aguda, y (2) al menos un marcador que es (i) un marcador de matriz extracelular (ECM) (ii) un marcador que es indicativo de la síntesis de la matriz extracelular (fibrogénesis), o (iii) un marcador que es indicativo de la degradación de la matriz extracelular (fibrólisis); y (b) el pronóstico del paciente se evalúa mediante la evaluación de los valores predictores.

35 Los valores de predicción y la evaluación de los valores de muestra del paciente que se determinan de acuerdo con los métodos de la invención: (1) se correlacionan con al menos el control del tumor o una respuesta clínica primaria a un régimen de tratamiento antineoplásico de la enfermedad, tiempo hasta la progresión de la enfermedad neoplásica y supervivencia global; y (2) son aplicables a los cánceres metastásicos y no metastásicos.

40 En una realización, los métodos de la invención predicen al menos un control del tumor o una respuesta clínica a un régimen de tratamiento dirigido contra el CCR avanzado, el CCR menos avanzado y las lesiones neoplásicas de diferentes orígenes (tales como mama, ovario, vejiga, colon), pancreático, pulmonar, mamario, gástrico, cabeza y cuello, o cáncer de próstata).

45 Los marcadores relacionados con la enfermedad neoplásica usados en los métodos de la invención incluyen nucleótidos o aminoácidos detectados en muestras de biopsia, fluidos corporales, muestras de sangre completa, y más preferiblemente en muestras de suero o plasma. Dichos marcadores incluyen genes y productos génicos (por ejemplo, péptidos, fragmentos de proteínas, proteínas precursoras o proteínas maduras y/o modificadas postraducción) que son expresadas por células malignas y/o células estroma no neoplásicas circundantes. En los métodos de la invención, estos productos génicos pueden detectarse en fluidos corporales antes, durante o después de la intervención terapéutica.

50 Si bien se desea estar limitados por ninguna teoría, se ha descubierto que ciertos procesos fibróticos son indicativos de la progresión del cáncer. Para las etapas avanzadas del cáncer, estos procesos fibróticos pueden ir acompañados de reacciones de fase aguda del tejido hepático (es decir, reacciones tisulares asociadas al cáncer). Se ha encontrado que los genes de la ECM, los genes asociados con la remodelación tisular o los productos de expresión de dichos genes son muy informativos con respecto a la respuesta clínica y la evaluación de la supervivencia general en oncología, particularmente si se combinan con marcadores relacionados con el tumor o el sistema inmunitario. Por lo tanto, en los métodos de la invención, se puede usar una combinación de marcadores moleculares que indiquen cambios patológicos del hígado y marcadores relacionados con tumores para evaluar el resultado clínico de la enfermedad cancerosa.

- Además, se ha determinado que la detección de genes ECM, genes asociados con la remodelación tisular o productos de expresión de dichos genes en muestras de pretratamiento es indicativa de la progresión de la enfermedad y del tejido maligno y se puede usar para el pronóstico y la predicción de la respuesta del tumor al tratamiento. La detección de dichos genes o productos genéticos en muestras obtenidas en serie, como muestras de suero o plasma, es indicativa de la presencia de tejido maligno y/o regresión y recurrencia de la enfermedad.
- De nuevo, aunque no se desea estar limitados por ninguna teoría, se ha concluido que la "respuesta de lesión" del tejido hepático, tal como se detecta al medir los procesos fibróticos, es un indicador sustituto de las neoplasias. Este problema es de gran relevancia clínica con respecto a las decisiones terapéuticas tomadas en las primeras etapas del desarrollo del tumor (por ejemplo, el tratamiento de la terapia en pacientes con cáncer colorrectal UICC I-III (Dukes A a C)), donde no se puede detectar metástasis a distancia. Por ejemplo, para el cáncer colorrectal, una parte sustancial de los pacientes desarrollan metástasis a distancia en el hígado sin presentarse como ganglios linfáticos positivos durante la resección quirúrgica. Concluimos que la evidencia de procesos fibróticos en el hígado es un indicador para pacientes de alto riesgo que necesitan un tratamiento más radical a pesar de un pronóstico positivo basado, por ejemplo, en indicadores de ganglios linfáticos negativos que se han determinado quirúrgicamente.
- Los métodos de la invención permiten a un proveedor de atención médica: (1) predecir, antes de la terapia, cómo responderá un paciente que padece una enfermedad neoplásica a un régimen de tratamiento antineoplásico; (2) evaluar el estado o progreso de una enfermedad neoplásica; (3) evaluar la probabilidad y duración de la supervivencia de un paciente que padece una enfermedad neoplásica; (4) evaluar el tiempo hasta la progresión (TTP) de una enfermedad neoplásica; (5) evaluar la toxicidad y los efectos secundarios de una quimioterapia aplicada; (6) evaluar la remodelación tisular implicada en la aparición de una enfermedad neoplásica; (7) determinar los regímenes de tratamiento óptimos para los pacientes que están predispuestos o padecen una enfermedad neoplásica; (8) diseñar programas clínicos útiles para monitorear el estado o progreso de una enfermedad neoplásica en uno o más pacientes; (9) facilitar el punto de atención o los diagnósticos remotos de enfermedades neoplásicas y monitorear el estado o progreso de una enfermedad neoplásica en uno o más puntos temporales.
- De acuerdo con la invención, en función de los valores de predicción y las evaluaciones de los valores de las muestras de pacientes, un proveedor de atención médica puede, por ejemplo, seleccionar terapias dirigidas combinadas, como los inhibidores de moléculas pequeñas que se dirigen al dominio de la quinasa (por ejemplo, Iressa®, Tarceva®, Vatalanib), un régimen de anticuerpos (por ejemplo, bevacizumab, trastuzumab o cetuximab), o un régimen de quimioterapia (como un régimen basado en 5'FU) o regímenes de quimioterapia combinada que incluyen al menos uno de los fármacos mencionados anteriormente y oxaliplatino, irinotecán, mitomicina o gemcitabina.
- En una realización preferida, los valores predictores se determinan mediante el análisis de regresión de Cox de valores de marcadores discretos y combinados correspondientes a los niveles de umbral de TIMP-1, gastrina, tenascina, colágeno VI y uPA en una muestra de suero de un paciente con cáncer colorrectal, y los valores de predicción se utilizan en un análisis de ROC para determinar la probabilidad de que el paciente responda favorablemente a un tratamiento dado.
- En otra realización preferida, los valores de predicción se determinan mediante el análisis de regresión de Cox de valores discretos y combinados correspondientes a los niveles de umbral de TIMP-1, gastrina, tenascina, colágeno VI y uPA en una muestra de suero de un paciente con cáncer colorrectal, y el predictor los valores se bifurcan y se utilizan para generar curvas de Kaplan Meier que reflejan la probabilidad de supervivencia del paciente.
- En otra realización preferida, los valores de predicción se determinan mediante análisis algorítmico de valores discretos y combinados correspondientes a niveles de umbral de Her-2/neu, EGFr y VEGF165 en una muestra de suero de paciente con cáncer colorrectal, y los valores de predicción se usan para predecir la probabilidad de supervivencia del paciente. Los niveles individuales elevados o reducidos de uno o más de estos marcadores, cuando se analizan de acuerdo con la invención, se correlacionan con una menor probabilidad de supervivencia del paciente.
- En otra realización preferida, los valores de predicción de enfermedad neoplásica se determinan usando valores discretos y combinados correspondientes a niveles de umbral de marcadores que incluyen MMP-2, Colágeno VI, Tenascin y VEGF. Los niveles individuales elevados o reducidos de cualquiera de estos marcadores, cuando se analizan de acuerdo con la invención, se correlacionan con una menor probabilidad de supervivencia del paciente.
- Se ha descubierto que los genes que se relacionan con los marcadores mencionados anteriormente representan motivos biológicos que afectan la organización general del tejido y que muestran características del tejido asociado a la enfermedad, particularmente en células neoplásicas. Los métodos de la invención pueden detectar estos fenómenos asociados a enfermedades neoplásicas en un nivel de ADN, ARN y proteínas.
- En otra realización preferida de los métodos de la invención, los valores de predicción se determinan en dos o más puntos de tiempo y la respuesta del paciente al régimen de tratamiento antineoplásico se evalúa comparando los valores de predicción determinados en cada punto de tiempo.
- En otra realización preferida, los valores del predictor de enfermedad neoplásica se determinan usando valores discretos y combinados correspondientes a niveles de umbral de marcadores que incluyen MMP-2, Gastrina, TIMP-1,

CA-19-9, o EGFr. Los niveles individuales elevados o reducidos de cualquiera de estos marcadores, cuando se analizan de acuerdo con la invención, se correlacionan con una menor probabilidad de supervivencia del paciente.

5 En otra realización preferida, los valores del predictor de enfermedad neoplásica se determinan usando valores discretos y combinados correspondientes a niveles de umbral de marcadores en un panel de marcadores que incluye al menos una matriz extracelular y un marcador de metaloproteinasa de matriz y VEGF. Una disminución en el nivel individual del marcador de matriz extracelular y un aumento en el nivel individual del marcador de metaloproteinasa de matriz, en ausencia de VEGF, se correlacionan con una mayor probabilidad de supervivencia del paciente. Por el contrario, una disminución en el nivel individual del marcador de matriz extracelular y un aumento en el nivel individual del marcador de metaloproteinasa de matriz, cuando se combinan con la detección de VEGF, se correlacionan con una menor probabilidad de supervivencia del paciente.

10 Los algoritmos de función lineal o no lineal utilizados para generar valores predictivos en relación con los métodos de la invención pueden derivarse correlacionando datos de marcadores relacionados con enfermedades neoplásicas de referencia utilizando, por ejemplo, análisis de función discriminante o análisis de regresión no paramétrica. Por ejemplo, los algoritmos de función lineal o no lineal utilizados en la invención pueden derivarse de:

15 (a) compilación de un conjunto de datos que comprende datos de marcadores relacionados con enfermedades neoplásicas para un primer grupo de sujetos, en donde los datos de marcadores incluyen datos relacionados con (1) al menos un marcador tumoral o al menos un marcador inmunitario o al menos un marcador agudo marcador de fase, y (2) al menos un marcador que es (i) un marcador de matriz extracelular (ECM) (ii) un marcador que es indicativo de síntesis de matriz extracelular (fibrogénesis), o (iii) un marcador que es indicativo de degradación de la matriz extracelular (fibrólisis);

20 (b) derivar un algoritmo de función lineal o no lineal del conjunto de datos compilados a través de la aplicación de al menos una metodología analítica seleccionada del grupo que consiste en análisis de función discriminante, análisis de regresión no paramétrica, árboles de clasificación, máquinas de vectores de soporte, vecino K más cercano y centroides y redes neuronales reducidas;

25 (c) calcular los valores del predictor de validación para un segundo grupo de sujetos ingresando datos que comprenden datos de marcadores relacionados con enfermedades neoplásicas para el segundo grupo de sujetos en el algoritmo derivado en el paso (b);

(d) comparar los valores de predicción de validación calculados en el paso (c) con las puntuaciones relacionadas con la enfermedad neoplásica para el segundo grupo de sujetos; y

30 (e) si los valores del predictor de validación determinados en el paso (c) no se correlacionan dentro de un nivel de tolerancia clínicamente aceptable con los valores del predictor de validación para el segundo grupo de sujetos, realizar las siguientes operaciones (i)-(iii) hasta que la tolerancia se cumple: (i) modificar el algoritmo sobre una base o bases que comprenden (1) revisar el conjunto de datos para el primer grupo de sujetos, y (2) revisar o cambiar la metodología analítica (ii) calcular valores de predictor de validación para el segundo grupo de sujetos ingresando datos que comprenden datos de marcadores relacionados con enfermedades neoplásicas para el segundo grupo de sujetos en el algoritmo modificado (iii) que evalúa si los valores del predictor de validación calculados utilizando el algoritmo modificado se correlacionan con los valores del predictor para el segundo grupo de sujetos dentro del nivel de tolerancia clínicamente aceptable. Las metodologías analíticas utilizadas en la derivación mencionada pueden incluir análisis de función discriminante y análisis de regresión no paramétrica, así como técnicas como árboles de clasificación, redes neuronales, máquinas de vectores de soporte, vecinos más cercanos a K y centroides encodigos.

40 La invención también proporciona una estructura de datos almacenada en un medio legible por ordenador que puede ser leído por un microprocesador y que comprende al menos un código que identifica de forma única un algoritmo que genera un valor de predictor de la manera descrita en este documento.

45 En otra realización, la invención proporciona un kit que comprende uno o más inmunoensayos que comprenden anticuerpos monoclonales y/o policlonales o cualquiera de sus fragmentos inmunorreactivos que se unen específicamente al marcador, en donde los inmunoensayos detectan y determinan los niveles de (1) al menos un marcador tumoral o al menos un marcador inmunitario o al menos un marcador de fase aguda; y (2) al menos un marcador que es (i) un marcador de matriz extracelular (ECM) (ii) un marcador que es indicativo de la síntesis de matriz extracelular (fibrogénesis); o (iii) un marcador que sea indicativo de la degradación de la matriz extracelular (fibrólisis).

50 En otra realización, la invención proporciona un kit que comprende:

(a) una estructura de datos que comprende un campo de datos que contiene datos que representan un algoritmo que genera valores predictivos que pueden usarse para predecir la respuesta de un paciente a un régimen de tratamiento antineoplásico, almacenándose dicha estructura de datos en un medio legible por ordenador que puede ser leído por un microprocesador y que comprende al menos un código que identifica de forma única un algoritmo que genera un valor de predictor de la manera descrita en este documento; y

- (b) uno o más inmunoensayos que detectan y determinan los niveles de (1) al menos un marcador tumoral o al menos un marcador inmunitario; y (2) al menos un marcador que es (i) un marcador de matriz extracelular (ECM) (ii) un marcador que es indicativo de la síntesis de matriz extracelular (fibrogénesis); o (iii) un marcador que sea indicativo de la degradación de la matriz extracelular (fibrólisis). En otra realización más, la invención proporciona métodos y sistemas implementables por ordenador para determinar si una composición es útil en el tratamiento de una enfermedad neoplásica.
- 5 En otra realización más, la invención proporciona métodos y sistemas implementables por ordenador útiles para tomar una decisión sobre gastos médicos relacionados con el tratamiento de una enfermedad neoplásica.
- Estas y otras realizaciones de la invención se describen adicionalmente en la descripción detallada de la invención.
- 10 Breve descripción de las figuras.
- La figura 1 ilustra las curvas de supervivencia de Kaplan-Meier que se generaron en relación con el experimento del Ejemplo 2 del presente documento.
- Las figuras 2 a 8, los resultados del análisis de Kaplan Meier de un solo parámetro mediante el uso de los valores de corte para cada uno de los marcadores seleccionados como se muestra en la figura 1.
- 15 La figura 2 ilustra un análisis Kaplan Meier de la gastrina.
- La figura 3 ilustra un análisis de Kaplan Meier de CA 19-9.
- La figura 4 ilustra un análisis de Kaplan Meier de TIMP-1.
- La figura 5 ilustra un análisis de Kaplan Meier de MMP-2.
- La figura 6 ilustra un análisis de Kaplan Meier de EGFr.
- 20 La figura 7 ilustra un análisis de Kaplan Meier de VEGF.
- La figura 8 ilustra un análisis de Kaplan Meier de CEA.
- La figura 9 ilustra un análisis de Kaplan Meier de los respectivos valores del algoritmo "MCT-V".
- La figura 10 ilustra la partición inicial en dos grupos cuando se usan los 17 parámetros identificados en las Tablas 4A y 4B.
- 25 La figura 10A ilustra la partición inicial en dos grupos cuando se usan los 17 parámetros identificados en las Tablas 4A y 4B. Los "supervivientes" se muestran como bolas verdes y los "no supervivientes" se muestran como bolas rojas.
- La figura 11 ilustra la partición mejorada en dos grupos mediante el análisis de componentes principales (PCA) cuando se utilizan los 5 principales parámetros discriminatorios (es decir, CEA, tamaño inicial del tumor, colágeno VI, MMP-2 y gastrina) descritos en las Tablas 4A y 4B.
- 30 La figura 11A ilustra la partición mejorada en dos grupos mediante el Análisis de componentes principales (PCA) cuando se usan los 5 principales parámetros discriminatorios (es decir, CEA, tamaño inicial del tumor, colágeno VI, MMP-2 y gastrina) descritos en las Tablas 4A y 4B. Los "supervivientes" se muestran como bolas verdes y los "no supervivientes" se muestran como bolas rojas.
- La figura 12 ilustra la expresión relativa de los miembros de la familia tirosina quinasa del receptor ERB en tejidos de FFPE de reseccionados de tumores primarios de pacientes como se describe en el Ejemplo 1 y como se determina por el perfil de qRT-PCR.
- 35 La figura 12A ilustra la expresión relativa de los miembros de la familia tirosina quinasa del receptor ERB en tejidos de FFPE de reseccionados de tumores primarios de pacientes como se describe en el Ejemplo 1 y se determina por el perfil qRT-PCR.
- 40 La figura 13 ilustra las curvas de supervivencia de Kaplan-Meier del análisis combinado de los niveles séricos de TIMP-1 y EGFr.
- Las figuras 14 y 14A ilustran la expresión relativa de la fase aguda, los marcadores inmunitarios y los marcadores corregulados en muestras de tumores frescos de pacientes como se describe en el Ejemplo 1 y según lo determinado por el análisis Affymetrix GeneChip.
- 45 Las figuras 15 y 15A ilustran la expresión relativa de la fase aguda, los marcadores inmunitarios y los marcadores corregulados en muestras de tumores frescos de pacientes como se describe en el Ejemplo 1 y según lo determinado por el análisis Affymetrix GeneChip.

Las figuras 16A y 16B ilustran mediciones en serie de muestras de suero de varios pacientes que revelaron un aumento en los niveles séricos de CRP [mg/l] en pacientes que sufrieron progresión de la enfermedad metastásica lateron como lo muestran los cambios en el tamaño del tumor [cm²].

Breve descripción de las tablas.

5 La Tabla 1 enumera los anticuerpos utilizados para detectar la ECM, la fibrosis y el marcador de fibrogénesis que se usaron dentro de esta invención.

La Tabla 2 enumera secuencias de nucleótidos representativas que pueden expresarse para producir marcadores que son útiles en los métodos de la invención y que se han usado para derivar algoritmos descritos en esta solicitud de patente.

10 La Tabla 3 enumera los tamaños de tumores ajustados por tomografía computarizada en cada ciclo de terapia para evaluar la respuesta del tumor al tratamiento.

Las tablas 4A y 4B muestran datos experimentales según lo determinan las mediciones duplicadas o por triplicado para cada uno de los 17 marcadores indicados en la muestra de suero pretratamiento.

15 La Tabla 5 presenta los resultados de un análisis de regresión cox utilizando todas las variables, incluidos los datos imputados.

Las tablas 6A y 6B enumeran las estimaciones del parámetro de regresión de Cox y las coordenadas ROC que se determinaron de acuerdo con el experimento del ejemplo 2.

La Tabla 7 enumera la evaluación de los valores del algoritmo MCT-V.

La Tabla 8 enumera una comparación de curvas de supervivencia ilustradas en las figuras 2 a 8.

20 La Tabla 9 muestra los resultados de múltiples pruebas estadísticas para discriminar a los pacientes con CCR metastásico que sobrevivieron durante más de 40 meses o menos de 18 meses desde el tratamiento primario mediante la evaluación de los parámetros séricos.

25 La Tabla 10 muestra los resultados de múltiples pruebas estadísticas para discriminar a los pacientes con CCR metastásico cuyas lesiones metastásicas responden al régimen basado en 5FU (Respuesta parcial) o no responden (enfermedad estable y enfermedad progresiva) mediante la determinación del ARN de un miembro de la familia EGFR en muestras de tejido FFPE.

La Tabla 11 muestra los datos experimentales según lo determinado por mediciones duplicadas o por triplicado para TIMP-1 y EGFR en la muestra de suero pretratamiento y el análisis combinado de los mismos.

30 La Tabla 12 enumera secuencias de nucleótidos representativas de la fase aguda, marcadores inmunes y marcadores que pueden expresarse para producir marcadores que son útiles en los métodos de la invención.

La Tabla 13 muestra los niveles de expresión de la fase aguda y los marcadores inmunitarios que discriminan entre los tumores que responden y los que no responden según lo determinado por el perfil de expresión génica utilizando Affymetrix GeneChip HG U133A.

Descripción detallada de la invención

35 Como se usa en el presente documento, los siguientes términos tienen los siguientes significados respectivos.

40 Los "marcadores de fase aguda" incluyen, entre otros, PCR, coeruloplasmina, fibrinógeno, haptoglobina, ferritina, proteína de unión a lipopolisacáridos (LBP), procalcitonina, bradiquinina, histamina, serotonina, leucotrienos (por ejemplo, LTB₄), interleucinas factor de necrosis tumoral alfa y prealbúmina. Los marcadores de fase aguda indican enfermedades inflamatorias de diverso origen. Se han descrito niveles elevados de proteínas de fase aguda para pacientes colorrectales. Glojnaric et al. (2001) Clin. Chem. Lab. Med. 2001; Feb. 39(2) 129-133, mostró que el carcinoma colorrectal causó un aumento en los niveles séricos de reactantes de fase aguda múltiples. En su estudio, la proteína amiloide A sérica mostró la reacción más potente en el estadio de la enfermedad preoperatoria, con un valor medio de 330 mg/l (rango 7-2506 mg/l) en comparación con los valores normales de menos de 1.2 mg/l obtenidos en 30 adultos sanos. Glojnaric describe la proteína amiloide A sérica como la mejor especificidad para el carcinoma colorrectal de todas las proteínas de fase aguda estudiadas (83-100%), y también indica que tiene una sensibilidad del 100%. Una lista no exclusiva de marcadores de fase aguda de ejemplo se enumera en la Tabla 12.

"Marcadores de pronóstico", como se usa en el presente documento, se refiere a factores que proporcionan información sobre el resultado clínico de pacientes con o sin tratamiento. La información proporcionada por los marcadores de pronóstico no se ve afectada por la interferencia terapéutica.

50 "Marcadores predictivos", como se usa en el presente documento, se refiere a factores que proporcionan información sobre la posible respuesta de un tumor a un agente o régimen terapéutico distinto.

El término "marcador" o "biomarcador" se refiere a una molécula biológica, por ejemplo, un ácido nucleico, péptido, hormona, etc., cuya presencia o concentración se puede detectar y correlacionar con una condición conocida, como un estado de enfermedad.

5 La estadificación es un método para describir qué tan avanzado está un cáncer. La estadificación del cáncer colorrectal toma en cuenta la profundidad de la invasión en la pared del colon y se disemina a los ganglios linfáticos y otros órganos. Etapa 0 (carcinoma in situ): el cáncer en etapa 0 también se denomina carcinoma in situ. Esta es una condición precancerosa, que generalmente se encuentra en un pólipo. Etapa I (Dukes A): El cáncer se diseminó a través del revestimiento interno del colon hasta la segunda y tercera capas de la pared del colon. No se ha diseminado fuera del colon. Etapa II (Dukes B): El cáncer se diseminó a través de la pared del colon fuera del colon hasta los tejidos cercanos. Etapa III (Dukes C): El cáncer se ha diseminado a los ganglios linfáticos cercanos, pero no a otras partes del cuerpo. Etapa IV: El cáncer se ha diseminado a otras partes del cuerpo, por ejemplo, metástasis al hígado o pulmones. Según la UICC, las etapas se subdividen aun más según T y N.

15 "Anticuerpo" incluye anticuerpos policlonales o monoclonales o cualquier fragmento de los mismos. Los anticuerpos monoclonales y/o policlonales se pueden usar en los métodos y sistemas de la invención. "Anticuerpo" u otro término similar como se usa en el presente documento incluye una inmunoglobulina completa que es monoclonal o policlonal, así como también fragmentos inmunorreactivos que se unen específicamente al marcador, incluidos Fab, Fab', F(ab')₂ y F(v). El término "Anticuerpo" también incluye proteínas de unión. Los anticuerpos marcadores de suero preferidos se describen en lo sucesivo.

20 Las muestras de fluido humano usadas en los ensayos de la invención pueden ser cualquier muestra que contenga marcadores de pacientes, por ejemplo, sangre, suero, plasma, orina, esputo o bronco alveolar (BAL) o cualquier otro líquido o heces corporales. Normalmente se emplea una muestra de suero o plasma.

25 Los anticuerpos utilizados en la invención se pueden preparar mediante técnicas generalmente conocidas en la técnica, y se generan típicamente para una muestra de los marcadores, ya sea como una proteína aislada y de origen natural, como una proteína expresada de forma recombinante o un péptido sintético que representa una porción antigénica de la proteína natural. El segundo anticuerpo se conjuga con un grupo detector, por ejemplo, fosfatasa alcalina, peroxidasa de rábano picante, un colorante fluorescente o cualquier otro resto de marcado generalmente útil para detectar biomoléculas en ensayos. Los conjugados se preparan mediante técnicas generalmente conocidas en la técnica.

30 Los "inmunoensayos" determinan la presencia de un marcador de paciente en una muestra biológica haciendo reaccionar la muestra con un anticuerpo que se une al marcador de suero, la reacción se lleva a cabo durante un tiempo y en condiciones que permiten la formación de un inmunocomplejo entre los anticuerpos y los marcadores séricos. Luego se realiza la determinación cuantitativa de tal inmunocomplejo.

35 En una versión, el anticuerpo usado es un anticuerpo generado al administrar a un mamífero (por ejemplo, un conejo, una cabra, un ratón, un cerdo, etc.) un inmunógeno que es un marcador de suero, un fragmento inmunogénico de un marcador de suero, o un anticuerpo idiotípico de unión a marcador anti-suero. Otros inmunoensayos útiles caracterizan el uso de anticuerpos de unión a marcadores séricos en general (independientemente de si se derivan a uno de los inmunógenos descritos anteriormente). Se puede emplear un formato de inmunoensayo de tipo sándwich que usa un segundo anticuerpo que también se une a un marcador de suero, uno de los dos anticuerpos está inmovilizado y el otro está marcado.

40 Los inmunoensayos preferidos detectan un complejo inmovilizado entre un marcador de suero y un anticuerpo de unión a marcador de suero usando un segundo anticuerpo que está marcado y se une al primer anticuerpo. Alternativamente, la primera versión presenta un formato de sándwich en donde el segundo anticuerpo también se une a un marcador de suero. En los procedimientos de inmunoensayo en sándwich, un anticuerpo de unión a marcador de suero puede ser un anticuerpo de captura unido a un material insoluble y el segundo un anticuerpo de unión a marcador de suero puede ser un anticuerpo de marcaje. Los procedimientos de inmunoensayo en sándwich descritos anteriormente se pueden usar con los anticuerpos descritos a continuación.

50 Los ensayos utilizados en la invención se pueden usar para determinar un marcador sanguíneo, por ejemplo, un marcador plasmático o sérico en muestras que incluyen orina, plasma, suero, líquido peritoneal o líquido linfático. Los kits de inmunoensayo para detectar un marcador sérico también se pueden usar en la invención, y comprenden un anticuerpo de unión a marcador sérico y los medios para determinar la unión del anticuerpo a un marcador sérico en una muestra biológica. En realizaciones preferidas, el kit incluye uno de los segundos anticuerpos o los antígenos competidores descritos anteriormente.

55 "Datos de enfermedad neoplásica de referencia y marcador de sangre" y "datos de enfermedad neoplásica" incluyen, entre otros, datos de suero o plasma indicativos del estado de la enfermedad, pero también se refieren a datos de expresión de tejidos o biopsias y el análisis de expresión respectivo de dicha muestras. Estos datos comprenden datos de proteínas, péptidos, ARN y ADN. Los datos de referencia de la enfermedad neoplásica se refieren a la cohorte de pacientes con estado clínico y resultados bien caracterizados. Esto permite el análisis comparativo.

Los "valores de predicción de validación" pueden calcularse ingresando datos que comprenden datos de marcadores relacionados con enfermedades neoplásicas para un grupo de sujetos en el algoritmo en caso de determinaciones de marcadores incompletas.

5 "Análisis de función discriminante" es una técnica utilizada para determinar qué variables discriminan entre dos o más grupos mutuamente excluyentes que ocurren naturalmente. La idea básica que subyace en el análisis de la función discriminante es determinar si los grupos difieren con respecto a un conjunto de variables predictoras que pueden o no ser independientes entre sí, y luego usar esas variables para predecir la pertenencia a un grupo (por ejemplo, de nuevos casos).

10 El análisis de la función discriminante comienza con una variable de resultado que es categórica (dos o más niveles mutuamente excluyentes). El modelo asume que estos niveles pueden ser discriminados por un conjunto de variables predictoras que, como ANOVA (análisis de varianza), pueden ser continuas o categóricas (pero preferiblemente son continuas) y, como ANOVA, asume que las funciones discriminantes subyacentes son lineales. El análisis discriminante no hace "partición de variación". Busca las correlaciones canónicas entre el conjunto de variables predictoras y las utiliza para construir funciones propias que explican los porcentajes de la variación total de todas las variables predictoras en todos los niveles de la variable de resultado.

15 El resultado del análisis es un conjunto de funciones discriminantes lineales (funciones propias) que utilizan combinaciones de las variables predictoras para generar una "puntuación discriminante" independientemente del nivel de la variable de resultado. El porcentaje de variación total se presenta para cada función. Además, para cada función propia, se desarrolla un conjunto de funciones discriminantes de Fisher que producen una puntuación discriminante basada en combinaciones de las variables predictoras dentro de cada nivel de la variable de resultado.

20 En general, se incluyen varias variables en un estudio para ver qué variable contribuye a la discriminación entre grupos. En ese caso, se genera una matriz de varianzas totales y covarianzas. De manera similar, se puede generar una matriz de desviaciones y covarianzas dentro del grupo. Se realiza una comparación de esas dos matrices a través de pruebas F multivariantes para determinar si hay diferencias significativas (con respecto a todas las variables) entre los grupos. Este procedimiento es idéntico al análisis multivariado de varianza o MANOVA. Al igual que en MANOVA, se podría primero realizar la prueba multivariable y, si es estadísticamente significativo, proceder a ver cuál de las variables tiene medios significativamente diferentes entre los grupos.

30 Para un conjunto de observaciones que contienen una o más variables cuantitativas y una variable de clasificación que define grupos de observaciones, el procedimiento de discriminación desarrolla un criterio discriminante para clasificar cada observación en uno de los grupos. Para tener una idea de qué tan bien "cumple" un criterio discriminante, es necesario clasificar (a priori) los diferentes casos, es decir, los casos que no se utilizaron para estimar el criterio discriminante. Sólo la clasificación de nuevos casos permite una evaluación de la validez predictiva del criterio discriminante.

35 Para validar el criterio derivado, la clasificación puede aplicarse a otros conjuntos de datos. El conjunto de datos utilizado para derivar el criterio discriminante se denomina conjunto de datos de entrenamiento o calibración o cohorte de entrenamiento del paciente. El conjunto de datos utilizado para validar el desempeño de los criterios discriminantes se denomina conjunto de datos de validación o cohorte de validación.

40 El criterio discriminante (función(es) o algoritmo), determina una medida de la distancia al cuadrado generalizada. Estas distancias se basan en la matriz de covarianza agrupada. Se puede usar la distancia de Mahalanobis o Euclidiana para determinar la proximidad. Estas distancias se pueden usar para identificar agrupaciones de los niveles de resultados y así determinar una posible reducción de niveles para la variable.

Una "matriz de covarianza combinada" es una matriz numérica formada al sumar los componentes de la matriz de covarianza para cada subpoblación en un análisis.

45 Un "predictor" es cualquier variable que se puede aplicar a una función para generar una variable dependiente o de respuesta o un "valor de predictor". En una realización de la presente invención, un valor de predictor puede ser una puntuación discriminante determinada a través del análisis de la función discriminante de dos o más marcadores sanguíneos del paciente (por ejemplo, marcadores de plasma o suero). Por ejemplo, un modelo lineal especifica la relación (lineal) entre una variable Y dependiente (o respuesta) y un conjunto de variables predictoras, las X, de modo que $\hat{Y} = \rho^0 + \rho^1 X^1 + \rho^2 X^2 + \dots + \rho^k X^k$ En esta ecuación, ρ^0 es el coeficiente de regresión para la intersección y los valores ρ^i son los coeficientes de regresión (para las variables 1 a k) calculados a partir de los datos.

50 Los "árboles de clasificación" se utilizan para predecir la pertenencia de casos u objetos en las clases de una variable dependiente categórica a partir de sus mediciones en una o más variables predictoras. El análisis del árbol de clasificación es una de las principales técnicas utilizadas en la llamada Minería de Datos. El objetivo de los árboles de clasificación es predecir o explicar las respuestas en una variable dependiente categórica y, como tal, las técnicas disponibles tienen mucho en común con las técnicas utilizadas en los métodos más tradicionales de análisis discriminante, análisis de conglomerados, estadísticas no paramétricas y estimación no lineal.

La flexibilidad de los árboles de clasificación los convierte en una opción de análisis muy atractiva, pero esto no quiere decir que se recomiende su uso con exclusión de los métodos más tradicionales. De hecho, cuando se cumplen los supuestos teóricos y distributivos típicamente más estrictos de los métodos más tradicionales, los métodos tradicionales pueden ser preferibles. Pero como técnica de exploración, o como técnica de último recurso cuando los métodos tradicionales fallan, los árboles de clasificación son, en opinión de muchos investigadores, insuperables. Los árboles de clasificación se utilizan ampliamente en campos aplicados tan diversos como la medicina (diagnóstico), la informática (estructuras de datos), la botánica (clasificación) y la psicología (teoría de la decisión). Los árboles de clasificación se prestan fácilmente para ser mostrados gráficamente, ayudando a que sean más fáciles de interpretar de lo que serían si solo fuera posible una interpretación numérica estricta.

5 Las "redes neuronales" son técnicas analíticas modeladas a partir de los procesos (hipotetizados) de aprendizaje en el sistema cognitivo y las funciones neurológicas del cerebro y capaces de predecir nuevas observaciones (sobre variables específicas) a partir de otras observaciones (sobre la misma u otras variables) después de ejecutar un proceso de aprendizaje a partir de datos existentes. Redes neuronales es una de las técnicas de minería de datos. El primer paso es diseñar una arquitectura de red específica (que incluya un número específico de "capas", cada una de las cuales consiste en un cierto número de "neuronas"). El tamaño y la estructura de la red deben coincidir con la naturaleza (por ejemplo, la complejidad formal) del fenómeno investigado. Debido a que este último obviamente no se conoce muy bien en esta etapa temprana, esta tarea no es fácil y con frecuencia involucra múltiples "pruebas y errores".

10 La red neuronal es luego sometida al proceso de "entrenamiento". En esa fase, la memoria del ordenador actúa como neuronas que aplican un proceso iterativo al número de entradas (variables) para ajustar los pesos de la red a fin de predecir de manera óptima los datos de muestra en los que se realiza el "entrenamiento". Después de la fase de aprendizaje de un conjunto de datos existente, la nueva red está lista y puede utilizarse para generar predicciones.

En una realización de la invención, las redes neuronales pueden comprender memorias de uno o más ordenadores personales o mainframe o dispositivos informatizados de punto de atención.

15 El "análisis de regresión de Cox" es una técnica estadística en la que la regresión de riesgos proporcionales de Cox se utiliza para analizar el efecto de varios factores de riesgo en la supervivencia. La probabilidad del punto final (muerte o cualquier otro evento de interés, por ejemplo, recurrencia de la enfermedad) se denomina peligro. El peligro se modela como:

$$H(t) = H_0(t) \times \exp(b_1 X_1 + b_2 X_2 + b_3 X_3 + \dots + b_k X_k)$$

30 donde $X_1 \dots X_k$ es una colección de variables predictoras y $H_0(t)$ es el riesgo de referencia en el tiempo t , que representa el peligro para una persona con el valor 0 para todas las variables predictoras. Al dividir ambos lados de la ecuación anterior por $H_0(t)$ y al tomar logaritmos, se obtiene:

$$\ln\left(\frac{H(t)}{H_0(t)}\right) = b_1 X_1 + b_2 X_2 + b_3 X_3 + \dots + b_k X_k$$

35 $H(t)/H_0(t)$ es la relación de riesgo. Los coeficientes $b_1 \dots b_k$ se estiman mediante la regresión de Cox y se pueden interpretar de manera similar a la de la regresión logística múltiple.

40 Si la covariable (factor de riesgo) es dicotómica y está codificada 1 si está presente y 0 si está ausente, entonces la cantidad $\exp(b_i)$ puede interpretarse como el riesgo relativo instantáneo de un evento, en cualquier momento, para un individuo con el factor de riesgo presente en comparación con un individuo con el factor de riesgo ausente, dado que ambos individuos son iguales en todas las demás covariables. Si la covariable es continua, entonces la cantidad $\exp(b_i)$ es el riesgo relativo instantáneo de un evento, en cualquier momento, para un individuo con un aumento de 1 en el valor del covariable en comparación con otro individuo, dado que ambos individuos son iguales en todas las demás covariables.

45 Las "curvas de Kaplan Meier" son una técnica no paramétrica (actuarial) para estimar eventos relacionados con el tiempo (la función de supervivencia). Normalmente, las curvas de Kaplan Meier se utilizan para analizar la muerte como un resultado. Puede usarse de manera efectiva para analizar el tiempo hasta un punto final, como la remisión. Las curvas de Kaplan Meier son un análisis univariado, una técnica de inicio apropiada, y estimar la probabilidad de la proporción de individuos en remisión en un momento determinado, a partir del inicio de la fecha activa (tiempo cero), es especialmente aplicable cuando la duración del seguimiento varía de un paciente a otro, y toma en cuenta a aquellos pacientes perdidos durante el seguimiento o que aún no se encuentran en remisión al final de un estudio clínico (por ejemplo, pacientes censurados, donde la censura no es informativa). Por lo tanto, Kaplan Meier es útil para evaluar las remisiones después de perder a un paciente. Dado que la distribución de supervivencia estimada para el estudio de cohorte tiene cierto grado de incertidumbre, se pueden calcular intervalos de confianza del 95% para cada probabilidad de supervivencia en la curva "estimada".

Se puede usar una variedad de pruebas (log-rank, Wilcoxon and Gehan) para comparar dos o más "curvas" de Kaplan-Meier en ciertas circunstancias bien definidas. La mediana del tiempo de remisión (el tiempo en que el 50% de la cohorte ha alcanzado la remisión), así como las cantidades de probabilidad de remisión de tres, cinco y diez años, también se pueden generar a partir del análisis de Kaplan-Meier, siempre que haya habido suficiente seguimiento de pacientes.

El análisis de regresión de Kaplan-Meier y Cox se puede realizar mediante el uso de paquetes de software disponibles comercialmente, por ejemplo, Graph Pad Prism™ y SPSS versión 11.

"Ordenador" se refiere a una combinación de un sistema de hardware de ordenador particular y un sistema operativo de software particular. Un ordenador o sistema computarizado de la invención puede comprender una calculadora de mano. Ejemplos de sistemas de hardware útiles incluyen aquellos con cualquier tipo de procesador de datos adecuado. El término "ordenador" también incluye, pero no se limita a, ordenadores personales (PC) que tienen un sistema operativo como DOS, Windows®, OS/2® o Linux®; Ordenadores Macintosh®; ordenadores que tienen JAVA®-OS como sistema operativo; y estaciones de trabajo gráficas, como los ordenadores de Sun Microsystems® y Silicon Graphics®, y otros ordenadores que tengan alguna versión del sistema operativo UNIX como AIX® o SOLARIS® de Sun Microsystems®; ordenadores incorporados que ejecuten un programador de control como una versión delgada de un sistema operativo, un dispositivo de mano; cualquier otro dispositivo con sistema operativo conocido y disponible; así como cualquier tipo de dispositivo que tenga un procesador de datos de algún tipo con una memoria asociada.

Si bien la invención se describirá en el contexto general de las instrucciones ejecutables por ordenador de un programa de ordenador que se ejecuta en un ordenador personal, los expertos en la técnica reconocerán que la invención también puede implementarse en combinación con otros módulos de programa. En general, los módulos de programa incluyen rutinas, programas, componentes y estructuras de datos que realizan tareas particulares o implementan tipos de datos abstractos particulares. Además, los expertos en la materia apreciarán que la invención puede ponerse en práctica con otras configuraciones de sistemas informáticos, incluidos dispositivos portátiles, sistemas multiprocesador, electrónica de consumo programable o basada en microprocesador, miniordenadores, ordenadores centrales y similares. La invención también puede ponerse en práctica en entornos informáticos distribuidos donde las tareas se realizan mediante dispositivos de procesamiento remoto que están vinculados a través de una red de comunicaciones. En un entorno informático distribuido, los módulos de programa pueden estar ubicados en dispositivos de almacenamiento de memoria local y remota.

Un sistema puramente ilustrativo para implementar la invención incluye un ordenador personal convencional, que incluye una unidad de procesamiento, una memoria del sistema y un bus de sistema que acopla varios componentes del sistema, incluida la memoria del sistema, a la unidad de procesamiento. El bus del sistema puede ser cualquiera de los varios tipos de estructura de bus que incluye un bus de memoria o un controlador de memoria, un bus periférico y un bus local que utiliza cualquiera de una variedad de arquitecturas de bus convencionales como PCI, VESA, Microchannel, ISA y EISA, para nombrar unos pocos. La memoria del sistema incluye una memoria de solo lectura (ROM) y una memoria de acceso aleatorio (RAM). Un sistema básico de entrada/salida (BIOS), que contiene las rutinas básicas que ayudan a transferir información entre elementos dentro del ordenador personal, como durante el inicio, se almacena en la ROM.

El ordenador personal incluye además una unidad de disco duro, una unidad de disco magnético, por ejemplo, para leer o escribir en un disco extraíble, y una unidad de disco óptico, por ejemplo, para leer un disco CD-ROM o para leer o leer escribir a otros medios ópticos. La unidad de disco duro, la unidad de disco magnético y la unidad de disco óptico están conectadas al bus del sistema mediante una interfaz de unidad de disco duro, una interfaz de unidad de disco magnético y una interfaz de unidad óptica, respectivamente. Las unidades y sus medios asociados legibles por ordenador proporcionan almacenamiento no volátil de datos, estructura de datos, instrucciones ejecutables por ordenador, etc. para el ordenador personal. Aunque la descripción de medios legibles por ordenador arriba se refiere a un disco duro, un disco magnético extraíble y un CD, los expertos en la técnica deben apreciar que otros tipos de medios que son legibles por ordenador, como los casetes magnéticos, tarjeta de memoria flash, los discos de video digital, los cartuchos Bernoulli y similares también pueden usarse en el entorno operativo ejemplar.

Se pueden almacenar varios módulos de programa en la memoria RAM del disco, incluido un sistema operativo, uno o más programas de aplicación, otros módulos de programa y datos del programa. Un usuario puede ingresar comandos e información en el ordenador personal a través de un teclado y un dispositivo señalador, como un mouse. Otros dispositivos de entrada pueden incluir un micrófono, palanca de mando, consola de juegos, antena parabólica, escáner o similares. Estos y otros dispositivos de entrada a menudo se conectan a la unidad de procesamiento a través de una interfaz de puerto serie que está acoplada al bus del sistema, pero pueden estar conectadas por otras interfaces, como un puerto paralelo, un puerto de juegos o un bus serie universal (USB). Un monitor u otro tipo de dispositivo de visualización también está conectado al bus del sistema a través de una interfaz, como un adaptador de video. Además del monitor, los ordenadores personales generalmente incluyen otros dispositivos de salida periféricos (no mostrados), como altavoces e impresoras.

El ordenador personal puede operar en un entorno de red usando conexiones lógicas a uno o más ordenadores remotos, como un ordenador remoto. El ordenador remoto puede ser un servidor, un enrutador, un dispositivo de par u otro nodo de red común, y típicamente incluye muchos o todos los elementos descritos en relación con el ordenador

personal. Las conexiones lógicas incluyen una red de área local (LAN) y una red de área amplia (WAN). Dichos entornos de red son comunes en oficinas, redes de ordenadores de toda la empresa (como los ordenadores de hospitales), intranets e Internet.

5 Cuando se usa en un entorno de red LAN, el ordenador personal se puede conectar a la red local a través de una interfaz de red o un adaptador. Cuando se usa en un entorno de red WAN, el ordenador personal generalmente incluye un módem u otros medios para establecer comunicaciones a través de la red de área amplia, como Internet. El módem, que puede ser interno o externo, se conecta al bus del sistema a través de la interfaz del puerto serie. En un entorno de red, los módulos de programa descritos en relación con el ordenador personal, o partes de los mismos, pueden almacenarse en el dispositivo de almacenamiento de memoria remota. Se apreciará que las conexiones de red
10 mostradas son de ejemplo y se pueden utilizar otros medios para establecer un enlace de comunicaciones entre los ordenadores.

Una plataforma de implementación puramente ilustrativa de la presente invención es un sistema implementado en un ordenador personal compatible con IBM que tiene al menos ocho megabytes de memoria principal y un disco duro de gigabytes, con Microsoft Windows como interfaz de usuario y cualquier variedad de software de administración de bases de datos, incluido Paradox. El software de la aplicación que implementa funciones predictivas se puede escribir en cualquier variedad de lenguajes, incluidos, entre otros, C++, y se almacena en medios legibles por ordenador como se define a continuación. Un usuario ingresa comandos e información que reflejan los marcadores de los pacientes en el ordenador personal a través de un teclado y un dispositivo señalador, como un ratón.

20 En una realización preferida, la invención proporciona una estructura de datos almacenada en un medio legible por ordenador, para ser leída por un microprocesador que comprende al menos un código que identifica de manera única las funciones de predictor y los valores derivados como se describe a continuación. Ejemplos de medios de uso de ordenador preferidos incluyen: medios de tipo no volátiles, codificados de manera rígida, tales como memorias de solo lectura (ROM) o memorias de solo lectura programables eléctricamente (EEPROM), medios de grabación grabables como disquetes, discos duros y CD-ROM y medios de transmisión, como los enlaces de comunicación digitales y analógicos.

Una "estructura de datos" puede incluir una colección de elementos de datos relacionados, junto con un conjunto de operaciones que reflejan las relaciones entre los elementos. Se puede considerar que una estructura de datos refleja la organización de los datos y su asignación de almacenamiento dentro de un dispositivo como un ordenador.

30 Por lo tanto, una estructura de datos puede comprender una organización de información, generalmente en la memoria, para una mejor eficiencia de algoritmo, como cola, pila, lista enlazada, montón, diccionario y árbol, o unidad conceptual, como el nombre y la dirección de una persona. Puede incluir información redundante, como la longitud de la lista o el número de nodos en un subárbol. Una estructura de datos puede ser una estructura de datos externa, que es eficiente incluso cuando el acceso a la mayoría de los datos es muy lento, como en un disco. Una estructura de datos puede ser una estructura de datos pasiva que solo se modifica mediante procesos o subprocesos externos, en contraste con una estructura de datos activa. Una estructura de datos activa o funcional tiene un hilo o proceso asociado que realiza operaciones internas para proporcionar el comportamiento externo de otra estructura de datos, a menudo más general. Una estructura de datos también puede ser una estructura de datos persistente que conserva sus versiones anteriores, es decir, las versiones anteriores pueden consultarse además de la última versión. Una estructura de datos puede ser una estructura de datos recursiva que está parcialmente compuesta por instancias más pequeñas o más simples de la misma estructura de datos. Una estructura de datos también puede ser un tipo de datos abstracto, es decir, un conjunto de valores de datos y operaciones asociadas que se especifican con precisión independientemente de cualquier implementación en particular.

Estos ejemplos de estructuras de datos, como en todas las formas de realización ejemplificadas en el presente documento, son solo ilustrativos y de ninguna manera son limitantes.

45 Un sistema de la invención puede comprender un dispositivo de mano útil en aplicaciones de punto de atención o puede ser un sistema que funciona de forma remota desde el punto de atención del paciente. En cualquier caso, el sistema puede incluir software complementario programado en cualquier lenguaje útil para implementar los métodos de la invención de acuerdo con algoritmos u otras técnicas analíticas descritas en este documento.

50 "Prueba de punto de atención" se refiere a la prueba predictiva en tiempo real que se puede realizar en un marco de tiempo rápido para que la prueba resultante se realice más rápido que las pruebas comparables que no emplean este sistema. Las pruebas en los puntos de atención se pueden realizar rápidamente y en el sitio, como en el consultorio de un médico, junto a la cama, en un laboratorio de estadísticas, en una sala de emergencias u otros lugares similares, especialmente cuando se requieren resultados rápidos y precisos. El paciente puede estar presente, pero tal presencia no es necesaria. El punto de atención incluye, pero no se limita a: salas de emergencia, quirófanos, laboratorios hospitalarios y otros laboratorios clínicos, consultorios médicos, en el campo o en cualquier situación en la que se desee un resultado rápido y preciso.

55 El término "paciente" se refiere a un animal, preferiblemente un mamífero, y lo más preferiblemente un humano.

Un "proveedor de atención médica" o un "tomador de decisiones de atención médica" comprende a cualquier persona autorizada para diagnosticar o tratar a un paciente, o para ayudar en el diagnóstico o tratamiento de un paciente. En el contexto de la identificación de nuevos medicamentos útiles para tratar la enfermedad hepática, un proveedor de atención médica puede ser una persona que no está autorizada para diagnosticar o tratar a un paciente, o para ayudar en el diagnóstico o tratamiento de un paciente.

Se hace referencia aquí colectivamente a "marcadores tumorales", "marcadores inmunitarios", "marcadores de fase aguda", "marcadores de matriz extracelular (ECM)", "marcadores que son indicativos de síntesis de matriz extracelular (fibrogénesis)" y "marcadores que son indicativos de degradación de matriz extracelular (fibrólisis)", como "marcadores", "marcadores relacionados con enfermedades neoplásicas" y "marcadores asociados con el cáncer". Estos marcadores: (1) incluyen, por ejemplo, un ácido nucleico, péptido, proteína o fragmento de gen que puede detectarse y correlacionarse con una condición conocida (como un estado de enfermedad); y (2) "marcadores sanguíneos" y "marcadores sanguíneos, por ejemplo, marcadores de plasma y suero". Tal como se usa en el presente documento, los marcadores incluyen ácidos nucleicos, péptidos, proteínas, fragmentos de polipéptidos o secuencias de ácidos nucleicos que muestran una expresión excesiva o insuficiente en un sujeto que padece cáncer de al menos alrededor del 10% en células cancerosas, en células estroma no cancerosas, en tejidos o en suero obtenido de un individuo que padece cáncer, en comparación con niveles de marcadores comparables obtenidos de un sujeto que no padece cáncer o que padece un cáncer más o menos avanzado.

Un ejemplo de un panel marcador usado en los métodos de la invención incluye:

(1) al menos un marcador seleccionado del grupo que consiste en marcadores tumorales, marcadores inmunes y marcadores de fase aguda, incluidos, entre otros, CEA, CA15-3, CA19-9, miembros de la superfamilia EGFR (por ejemplo, EGFr, HER-2/neu, HER-3 y HER-4), ERBB3, ERBB4, c-Kit, KDR, FLT4, FLT3, c-Met, miembros de la superfamilia FGFR (FGFR1, FGFR2, FGFR3, FGFR4), miembros de la familia de ligandos FGFR (por ejemplo, FGF-1, FGF-2, FGF-3, FGF-4, FGF-5, FGF-6, FGF-7 y FGF-9 y variantes de empalme relacionadas), miembros de la familia del factor de crecimiento (como VEGF y VEGF alfa), miembros de la superfamilia VEGFR, por ejemplo, KDR, FLT4, FLT3, miembros de la familia de ligandos VEGFR, incluidos VEGFA, VEGFB, VEGFC y VEGFD, dominios integrados de miembros de factores de crecimiento (incluidos miembros de la familia como VEGF-A, VEGF-B, VEGF-C (preferiblemente isoformas alfa de VEGF como VEGF189, VEGF165, VEGF121, etc.) y VEGFC, hormonas (como Gastrina), receptores de interleucina (como IL2R), interleuquinas (tales como IL6), factores del complemento, proteína de fase aguda (como CRP; ORM1, ORM2, amiloide sérico A2, componente amiloide P); y

(2) al menos un marcador que es:

(i) un marcador de matriz extracelular (ECM) seleccionado del grupo que consiste en colágenos, proteínas de adhesión basales (fibronectinas, lamininas), entactina, proteoglicanos y glicosaminoglicanos como PIIINP, miembros de la superfamilia de colágeno, por ejemplo, Colágeno I, Colágeno II, Colágeno III, Colágeno IV, Colágeno V, Colágeno VI, Colágeno VII, Colágeno VIII, Colágeno IX, Colágeno X, Colágeno XI, Colágeno XII, y Tenascina, Lamininas, HA; o

(ii) un marcador que es indicativo de la síntesis de matriz extracelular (fibrogénesis) seleccionada del grupo que consiste en preformas de colágenos, proteínas de adhesión basales (fibronectinas, lamininas), entactina, proteoglicanos y glicosaminoglicanos o prepropéptidos de los mismos, como PIIINP Colágeno IV, Colágeno VI, Tenascina, Laminina, Hialurón (HA); o

(iii) un marcador que sea indicativo de la degradación de la matriz extracelular (fibrólisis) seleccionado del grupo que consiste en la superfamilia de MMP (incluyendo MMP-1, MMP-2, MMP-3, MMP-7, MMP-8, MMP-9, MMP-12, MMP-13, MMP-14, MMP-15, MMP-16, MMP-17, MMP-19, MMP-20, MMP-24 y MMP-26, preferiblemente MMP-2, MMP-3, MMP-7, MMP-9, MMP-12, MMP-24 y MMP-26); complejo MMP-9/TIMP-1, o inhibidores asociados del mismo, tales como TIMP-1, TIMP-2, TIMP-3 y TIMP-4.

Un ejemplo de un panel marcador usado en los métodos de la invención incluye la combinación de:

(1) al menos un marcador seleccionado del grupo que consiste en marcadores tumorales, incluidos, entre otros, CEA, CA15-3, CA19-9, miembros de la superfamilia EGFR (por ejemplo, EGFr, HER-2/neu, HER-3 y HER-4), ERBB3, ERBB4, c-Kit, KDR, FLT4, FLT3, c-Met, miembros de la superfamilia FGFR (FGFR1, FGFR2, FGFR3, FGFR4), miembros de la familia de ligandos FGFR (por ejemplo, FGF-1, FGF-2, FGF-3, FGF-4, FGF-5, FGF-6, FGF-7 y FGF-9 y variantes de empalme relacionadas), miembros de la superfamilia VEGFR, por ejemplo, KDR, FLT4, FLT3, miembros de la familia de ligandos VEGFR que incluyen VEGFA, VEGFB, VEGFC y VEGFD, dominios integrados de miembros de factores de crecimiento (incluidos miembros de la familia como VEGF-A, VEGF-B), VEGF-C (preferiblemente isoformas alfa de VEGF, tales como VEGF189, VEGF165, VEGF121, etc.), y VEGFC, hormonas (como la gastrina), y/o

(2) al menos un marcador seleccionado del grupo que consiste en marcadores inmunes que incluyen pero no se limitan a los receptores de interleucina (como IL2R), interleucinas (como IL6), factores del complemento; y/o

(3) al menos un marcador seleccionado del grupo que consiste en marcadores de fase aguda que incluyen, entre otros, proteínas de fase aguda (como CRP; ORM1, ORM2, suero amiloide A2, componente amiloide P) y genes corregidos (APOB, APOC1, APOE, C1QA, C1QB, C3, C4A, CRP, F2, F5, FGA, FGB, FGG, ITIH3, ITIH4, TF, ARL7, BBOX1, C4B, C4BPA, C8B, CAST, CPB2, FBP17, FGL1, FLJ11560, FSTL3, GC, HXB, IGF1, ITIH2, KMO, MAGP2, MGC4638, NNMT, PBX3, PCDH17, PLOD, PPP3R1, PRKCDBP, SERPINA1, SERPINE1, SERPING1, TEGT, TUBB, UGT2B4); y/o

(4) al menos un marcador que es:

(i) un marcador de matriz extracelular (ECM) seleccionado del grupo que consiste en colágenos, proteínas de adhesión basales (fibronectinas, lamininas), entactina, proteoglicanos y glicosaminoglicanos como PIIINP, miembros de la superfamilia de colágeno, por ejemplo, Colágeno I, Colágeno II, Colágeno III, Colágeno IV, Colágeno V, Colágeno VI, Colágeno VII, Colágeno VIII, Colágeno IX, Colágeno X, Colágeno XI, Colágeno XII, y Tenascina, Laminina, HA; o

(ii) un marcador que es indicativo de la síntesis de matriz extracelular (fibrogénesis) seleccionada del grupo que consiste en preformas de colágenos, proteínas de adhesión basales (fibronectinas, lamininas), entactina, proteoglicanos y glicosaminoglicanos o prepropéptidos de los mismos, como PIIINP Colágeno IV, Colágeno VI, Tenascina, Laminina, Hialurón (HA); o

(iii) un marcador que sea indicativo de la degradación de la matriz extracelular (fibrólisis) seleccionado del grupo que consiste en la superfamilia de MMP (incluyendo MMP-1, MMP-2, MMP-3, MMP-7, MMP-8, MMP-9, MMP-12, MMP-13, MMP-14, MMP-15, MMP-16, MMP-17, MMP-19, MMP-20, MMP-24 y MMP-26, preferiblemente MMP-2, MMP-3, MMP-7, MMP-9, MMP-12, MMP-24 y MMP-26); Complejo MMP-9/TIMP-1, o inhibidores asociados del mismo, tales como TIMP-1, TIMP-2, TIMP-3 y TIMP-4.

Preferiblemente, el panel incluye al menos dos marcadores, y más preferiblemente tres marcadores, siendo cada marcador de un conjunto diferente y diferentes entre sí.

Los paneles marcadores preferidos usados en los métodos de la invención incluyen:

(1) al menos un marcador seleccionado del grupo que consiste en marcadores tumorales séricos, marcadores inmunitarios séricos y marcadores de fase aguda que incluyen, entre otros: CEA, CA15-3, CA19-9, miembros EGFR, HER-2/neu, VEGF alfa, gastrina, IL2R, IL6, CRP, ORM1, ORM2, suero amiloide A2 (SAA2), componente amiloide P, C4A, C1QB, C1QA, APOC1, F2, APOB, C3, TF, F5, FGA, FGB, FGG, APOE, ITIH3, ITIH4; y

(2) al menos un marcador que es (i) un marcador ECM hepático seleccionado del grupo que consiste en PIIINP, Colágeno IV, Colágeno VI, Tenascina, Laminina, HA (ii) un marcador que es indicativo de fibrogénesis hepática seleccionada de el grupo que consiste en prepropéptidos de los mismos, tales como PIIINP, Colágeno IV, Colágeno VI, Tenascina, Laminina, HA o (iii) un marcador que es indicativo de fibrólisis hepática seleccionado del grupo que consiste en MMP-2, MMP-3, MMP-7, MMP-9, MMP-12, MMP-24, MMP-9/TIMP-1 y uPA.

La expresión de MMP-7 y MMP-12 se pronuncia en el cáncer colorrectal y, si se determina a nivel de ARN, se correlaciona con un resultado negativo.

Un "conjunto de datos comparativos" puede comprender cualquier dato que refleje cualquier indicio cualitativo o cuantitativo de una enfermedad neoplásica. En una realización, el conjunto de datos comparativos puede comprender uno o más valores numéricos, o rango de valores numéricos, asociados con disminuciones y elevaciones en los niveles (1) de al menos un marcador tumoral o al menos un marcador inmunitario, y (2) en al menos un marcador que es (i) un marcador de matriz extracelular (ECM) (ii) un marcador que es indicativo de la síntesis de matriz extracelular (fibrogénesis), o (iii) un marcador que es indicativo de la degradación de la matriz extracelular (fibrólisis).

Los niveles de marcadores de conjuntos de datos comparativos se determinan típicamente por comparación con niveles normales (sanos) o umbrales de marcadores en sujetos que comprenden cohortes de referencia.

Por ejemplo, el rango normal de TIMP-1 en los sueros está entre aproximadamente 424 y aproximadamente 1037 ng/ml. El rango normal de Colágeno VI en los sueros está entre aproximadamente 1.2 y aproximadamente 7.2 ng/ml. El rango normal de HA en los sueros está entre aproximadamente 5.4 y aproximadamente 34.7 ng/ml. El rango normal de Laminina en los sueros está entre aproximadamente 6.3 y aproximadamente 3.7 ng/ml. El rango normal de MMP-2 en sueros de todas las edades está entre aproximadamente 388 y aproximadamente 1051 ng/ml (media 668 ng/ml; mediana de 647 ng/l). El rango normal de MMP-9 en sueros de todas las edades es de aproximadamente 201.6 a aproximadamente 1545 ng/ml (media de 719 ng/ml; mediana de 683 ng/l). El rango normal de PIIINP en sueros de todas las edades es de aproximadamente 0.9 a aproximadamente 25.6 ng/ml (media 5.84 ng/ml). El rango normal de Tenascina en sueros de todas las edades es de aproximadamente 206.9 a aproximadamente 1083.2 ng/ml (media de 455 ng/ml). El rango normal de Colágeno IV en sueros de todas las edades está entre aproximadamente 66 y aproximadamente 315 ng/ml (media 183 ng/ml). El rango normal de HER-2/neu en los sueros es inferior a aproximadamente 15 ng/ml.

Los niveles normales (saludables) o umbrales son menores que alrededor de 163 pg/ml de VEGF165, (95% de caída por debajo), menos de alrededor de 5 ng/ml para CEA, menos de alrededor de 20 U/ml para CA 15-3, menos que alrededor de 28-115 µE/ml para Gastrina, menos de alrededor de 15 ng/ml para Her-2/neu y más de 45 ng/ml para EGFR.

5 Un nivel disminuido de EGFR es uno que es menor que el rango normal o umbral de EGFR, es decir, alrededor de 45-78 ng/ml. De manera similar, un nivel de TIMP-1 aumentado es uno que es mayor que los niveles de TIMP-1 normales de menos de alrededor de 1037 ng/ml (formato Immuno1). Un nivel incrementado de HER-2/neu es uno que es mayor que los niveles normales de HER-2/neu de menos de alrededor de 15 ng/ml. De manera similar, un nivel de CEA aumentado es uno que es mayor que el nivel de CEA ajustado a la enfermedad de alrededor de 499 ng/ml, mientras que el nivel de CEA normal es de alrededor de 5 ng/ml.

10 En particular, se encuentra un tiempo de progresión más corto y una supervivencia global más corta en pacientes con cáncer colorrectal metastásico que tienen niveles de EGFR que son menores que el rango de control de aproximadamente 45-78 ng/ml, niveles bajos de Tenascina por debajo del rango de corte de aproximadamente 1083 ng/ml y/o niveles bajos de Colágeno VI por debajo de un rango de corte de aproximadamente 7.2 ng/ml combinados con niveles elevados de HER-2/neu, en donde elevados se refieren a niveles que son mayores que el valor de control de aproximadamente menos de aproximadamente 15 ng/ml, los niveles de TIMP-1 por encima del rango de corte de aproximadamente 1037 ng/ml (Formato inmunológico) o por encima de aproximadamente 250 ng/ml (Formato ELISA), niveles elevados de VEGF165 por encima de un rango de corte de aproximadamente 221 pg/ml y/o niveles de gastrina por encima de alrededor de 25.4 pg/ml.

20 Un conjunto de datos comparativos que se relaciona con niveles séricos alterados de marcadores tumorales indicativos de enfermedad cancerosa puede incluir o identificar una combinación de niveles séricos elevados de HER-2/neu (por ejemplo, mayor que el nivel normal de menos de aproximadamente 15 ng/ml) y/o niveles de ECD de EGFR disminuidos (por ejemplo, menos del rango normal de alrededor de 45-78 ng/ml) y/o niveles altos de VEGF (por ejemplo, para la isoforma 165 de VEGFA, mayor que el nivel normal de menos de alrededor 221 pg/ml), como valores indicativos de un tiempo de progresión más corto y un tiempo de supervivencia general más corto.

25 Los "marcadores complementarios" incluyen, pero no se limitan a, datos de peso, sexo, edad y perfil de expresión del paciente de tejido tumoral fresco y fijo.

• Preferiblemente, los marcadores se obtienen de una muestra de fluido corporal o una muestra de tejido. Los fluidos corporales adecuados incluyen, entre otros, muestras de líquido pleural, muestras de líquido de lavado pulmonar o bronquial, muestras de líquido sinovial, muestras de líquido peritoneal, heces, muestras de aspirado de médula ósea, linfa, líquido cefalorraquídeo, muestras de líquido ascítico, muestras de líquido amniótico, muestras de esputo, lavados de vejiga, semen, orina, saliva, lágrimas, sangre y componentes sanguíneos, suero y plasma, y similares. El suero es una muestra de fluido corporal preferida. Las muestras de tejido adecuadas también incluyen varios tipos de tumor o tejido canceroso, o tejido orgánico, como los que se toman en la biopsia.

35 Uno o más valores numéricos, o rango de valores numéricos que están asociados con una enfermedad neoplásica.

La "predicción de un resultado clínico relacionado con un paciente que padece o tiene riesgo de desarrollar una enfermedad neoplásica" se ha definido previamente.

"Responder a uno o más regímenes de tratamiento de enfermedades neoplásicas" se ha definido previamente.

40 "Tomar una decisión sobre gastos médicos relacionada con el tratamiento de una enfermedad neoplásica" incluye, entre otros, una decisión de un asegurador relacionada con el reembolso de un régimen de tratamiento de enfermedades neoplásicas o una evaluación de las tasas de seguro u otros cargos o pagos.

45 La invención proporciona métodos y sistemas implementables por ordenador para determinar si una composición es útil en el tratamiento de una enfermedad neoplásica. Por ejemplo, uno o más compuestos se administran a uno o más sujetos (preferiblemente mamíferos, y lo más preferiblemente humanos) que padecen una enfermedad neoplásica y la respuesta del sujeto al régimen de tratamiento de la enfermedad neoplásica se usa para evaluar la eficacia del compuesto como un agente de la enfermedad antineoplásica.

50 El término "enfermedad neoplásica" se usa para describir el proceso patológico que resulta en la formación y el crecimiento de una neoplasia, es decir, un tejido anormal que crece por proliferación celular más rápidamente que el tejido normal y continúa creciendo después de los estímulos que iniciaron el nuevo cese el crecimiento. Las enfermedades neoplásicas muestran una falta parcial o completa de organización estructural y coordinación funcional con el tejido normal, y generalmente forman una masa distinta de tejido que puede ser benigno (tumor benigno) o maligno (carcinoma). El término "cáncer" se usa como un término general para describir cualquiera de los diversos tipos de enfermedades neoplásicas malignas, la mayoría de las cuales invaden los tejidos circundantes, pueden metastatizarse en varios sitios y es probable que se repitan después de un intento de extirpación y causen la muerte del paciente a menos que se trate adecuadamente. Tal como se usa en el presente documento, el término cáncer se incluye en el término enfermedad neoplásica.

- 5 Como se usa en el presente documento, "procesos fibróticos" o "fibrosis" se refiere a la formación de tejidos fibrosos como una reacción o como un proceso de reparación que puede ocurrir durante enfermedades de origen diverso (incluidas enfermedades cancerosas e inflamación) y/o tratamiento. La formación de tejido fibroso puede reemplazar a otro tejido y el "tejido cicatricial" resultante puede afectar la funcionalidad del órgano respectivo de manera detectable. Como parte de esta invención, estos procesos se pueden detectar en lesiones primarias y lesiones metastásicas de enfermedades cancerosas. Esto se refiere al hecho de que la remodelación de la ECM (por ejemplo, la destrucción de las membranas basales durante los primeros pasos de la invasión) encapsula las células tumorales y da como resultado la formación de un lecho tumoral.
- 10 Los avances científicos demuestran que los procesos patogénicos generales en el hígado, como los procesos fibróticos, implican la proliferación y activación de células estrelladas hepáticas (también llamadas lipocitos, almacenamiento de grasa o células Ito), que sintetizan y secretan proteínas de la matriz extracelular en exceso. Sin embargo, los procesos fibróticos no están restringidos al tejido hepático. La fibrosis se refiere a la formación de tejidos fibrosos como una reacción o como un proceso de reparación que puede ocurrir durante una enfermedad de origen diverso (incluida la inflamación) y/o tratamiento. La formación de tejido fibroso puede reemplazar a otro tejido y el
- 15 "tejido cicatricial" resultante puede afectar la funcionalidad del órgano respectivo de manera detectable. En el hígado, los cambios fibróticos son comunes para las enfermedades de múltiples etiologías, por ejemplo, la hepatitis viral crónica B y C, la enfermedad hepática alcohólica, así como las enfermedades hepáticas autoinmunes y genéticas. Todas estas enfermedades conducen a problemas clínicos a través de la vía final común de la fibrosis hepática progresiva y el desarrollo eventual de la cirrosis.
- 20 La fibrosis hepática es una acumulación reversible de matriz extracelular en respuesta a una lesión crónica en la que los nódulos aún no se han desarrollado, mientras que la cirrosis implica un proceso irreversible, en donde bandas gruesas de matriz rodean completamente el parénquima, formando nódulos. La evaluación de los procesos dinámicos en tejidos enfermos mediante la determinación en serie de los parámetros séricos permite el monitoreo efectivo del estado de la enfermedad y la respuesta al tratamiento.
- 25 Los métodos de la invención pueden evaluar cambios dentro de muestras tomadas de un paciente en diferentes puntos de tiempo antes, durante o después del tratamiento. Los valores de predicción determinados en base a dicho muestreo en serie se comparan con los valores de predicción calculados utilizando niveles asociados a la enfermedad normales o ajustados.
- 30 La actividad similar a la fibrosis en el hígado puede interrumpirse temporalmente debido a cambios en el tejido neoplásico causados por el tratamiento. Los procesos inflamatorios relacionados con el tratamiento también pueden inducirse debido a la muerte celular pronunciada de células cancerosas o no cancerosas y la invasión de células inmunitarias; la expresión de la proteína marcadora (EGFR, VEGFR, ligandos de VEGF, etc.) puede reducirse en respuesta al tratamiento tóxico o citostático de células de tumores o estroma. Por lo tanto, una evaluación de los cambios relacionados con el proceso fibrótico puede proporcionar información adicional en un solo punto de tiempo,
- 35 por ejemplo, muestras de pretratamiento.
- "Valores de puntuación de marcador de cohorte de validación" significa un puntaje numérico derivado de la combinación lineal de los pesos discriminantes obtenidos de la cohorte de entrenamiento y los valores de marcador para cada paciente en la cohorte de validación.
- 40 "Valores de corte de marcador de paciente" significa el valor de un marcador de combinación de marcadores en donde se alcanza una sensibilidad o especificidad predeterminada. "Valor predictivo positivo" ("PPV"): significa la probabilidad de tener una enfermedad dado que un valor de fabricante (o conjunto de valores de marcador) se eleva por encima de un límite definido.
- 45 "Curva característica del operador receptor" ("ROC"): es una representación gráfica de la relación funcional entre la distribución de la sensibilidad de un marcador y los valores de especificidad 1 en una cohorte de personas enfermas y en una cohorte de personas no enfermas.
- "Área bajo la curva" ("AUC") es un número que representa el área bajo una curva característica del operador receptor. Cuanto más cerca esté este número de uno, más discriminarán los valores de los marcadores entre cohortes enfermas y no enfermas.
- 50 "Prueba de Chi-cuadrado de McNemar" ("La prueba de McNemar X^2 ") es una prueba estadística que se utiliza para determinar si dos proporciones correlacionadas (proporciones que comparten un numerador común pero diferentes denominadores) son significativamente diferentes entre sí.
- 55 Un "análisis de regresión no paramétrico" es un conjunto de técnicas estadísticas que permite el ajuste de una línea para datos bivariados que hacen poca o ninguna suposición con respecto a la distribución de cada variable o el error en la estimación de cada variable. Ejemplos son: los estimadores de ubicación de Theil, la regresión de Passing-Bablok y la regresión de Deming.
- "Valores de corte" o "Valores de umbral" son valores numéricos de un marcador (o conjunto de marcadores) que definen una sensibilidad o especificidad especificada.

Se entiende que el término "equivalente", con respecto a una secuencia de nucleótidos, incluye secuencias de nucleótidos que codifican polipéptidos funcionalmente equivalentes. Las secuencias de nucleótidos equivalentes incluirán secuencias que difieren en una o más sustituciones, adiciones o eliminaciones de nucleótidos, como las variantes alélicas y, por lo tanto, incluyen secuencias que difieren debido a la degeneración del código genético.

5 "Equivalente" también se usa para referirse a secuencias de aminoácidos que son funcionalmente equivalentes a la secuencia de aminoácidos de un homólogo de mamíferos de una proteína marcadora de sangre (por ejemplo, sueros), pero que tienen diferentes secuencias de aminoácidos, por ejemplo, al menos una, pero menos de 30, 20, 10, 7, 5 o 3 diferencias, por ejemplo, sustituciones, adiciones o eliminaciones.

10 Como se usa en el presente documento, los términos "gen marcador de suero de enfermedad neoplásica" se refieren a un ácido nucleico que: (1) codifica proteínas marcadoras de sangre de enfermedad neoplásica (por ejemplo, suero), que incluyen proteínas marcadoras de suero de enfermedad neoplásica identificadas en el presente documento; y (2) que están asociados con un marco de lectura abierto, que incluye secuencias de exón y (opcionalmente) intrón. Un "gen marcador de suero de enfermedad neoplásica" puede comprender secuencias de exones, aunque puede incluir opcionalmente secuencias de intrones que se derivan de, por ejemplo, un gen cromosómico relacionado o no relacionado. El término "intrón" se refiere a una secuencia de ADN presente en un gen dado que no se traduce en proteínas y generalmente se encuentra entre exones. Un gen puede incluir además secuencias reguladoras, por ejemplo, un promotor, un potenciador, etc. "Enfermedad neoplásica del gen marcador de suero" incluye, pero no se limita a, secuencias de nucleótidos que son complementarias, equivalentes u homólogas a las SEQ ID NO: 1-42 de la Tabla 2.

20 "Homología", "homólogos de", "homólogo", o "identidad" o "similitud" se refiere a la similitud de secuencia entre dos polipéptidos o entre dos moléculas de ácido nucleico, siendo la identidad una comparación más estricta. Cada homología e identidad pueden determinarse comparando una posición en cada secuencia que puede alinearse para propósitos de comparación. Cuando una posición en la secuencia comparada está ocupada por la misma base o aminoácido, entonces las moléculas son idénticas en esa posición. Un grado de homología o similitud o identidad entre las secuencias de ácido nucleico es una función del número de nucleótidos idénticos o coincidentes en las posiciones compartidas por las secuencias de ácido nucleico.

El término "porcentaje idéntico" se refiere a la identidad de secuencia entre dos secuencias de aminoácidos o entre dos secuencias de nucleótidos. Cada identidad puede determinarse comparando una posición en cada secuencia que puede alinearse para fines de comparación.

30 Cuando una posición equivalente en las secuencias comparadas está ocupada por la misma base o aminoácido, entonces las moléculas son idénticas en esa posición; cuando el sitio equivalente está ocupado por el mismo o un residuo de aminoácido similar (por ejemplo, similar en naturaleza estérica y/o electrónica), entonces las moléculas pueden denominarse homólogas (similares) en esa posición. La expresión como un porcentaje de homología, similitud o identidad se refiere a una función del número de aminoácidos idénticos o similares en las posiciones compartidas por las secuencias comparadas. Se pueden usar diversos algoritmos de alineación y/o programas, incluidos FASTA, BLAST o ENTREZ. FASTA y BLAST están disponibles como parte del paquete de análisis de secuencia GCG (Universidad de Wisconsin, Madison, WI) y se pueden usar, por ejemplo, con la configuración predeterminada. ENTREZ está disponible a través del National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Md. En una realización, el porcentaje de identidad de dos secuencias puede determinarse mediante el programa GCG con un peso de brecha de 1, por ejemplo, cada brecha de aminoácidos se pesa como si se tratara de un solo aminoácido o un desajuste de nucleótidos entre las dos secuencias. Otras técnicas para determinar la identidad de secuencia son bien conocidas y se describen en la técnica.

45 Los ácidos nucleicos preferidos usados en la presente invención tienen una secuencia al menos 70%, y más preferiblemente 80% idénticas y más preferiblemente 90% e incluso más preferiblemente al menos 95% idénticas a, o complementarias a, una secuencia de ácido nucleico de un homólogo de mamífero de un gen que expresa un marcador como se definió anteriormente. Los ácidos nucleicos particularmente preferidos usados en la presente invención tienen una secuencia al menos 70%, y más preferiblemente 80% idéntica y más preferiblemente 90% e incluso más preferiblemente al menos 95% idéntica o complementaria a una secuencia de ácido nucleico de un mamífero homólogo de un gen que expresa un marcador como se definió anteriormente.

50 Inmunoensayos

Se pueden efectuar inmunoensayos en suero para detectar y medir los niveles de (1) al menos un marcador tumoral o al menos un marcador inmunitario o al menos un marcador de fase aguda, y (2) al menos un marcador que es (i) un marcador ECM (ii) a marcador que es indicativo de la síntesis de la matriz extracelular (fibrogénesis), o (iii) un marcador que es indicativo de la degradación de la matriz extracelular (fibrólisis) de acuerdo con los protocolos descritos a continuación. Los marcadores complementarios que incluyen el peso, el sexo y la edad, y los datos de perfiles de expresión de tejido tumoral fresco y fijo, también se pueden evaluar para determinar los valores de predictores de acuerdo con los métodos de la invención.

60 Los niveles de (1) al menos un marcador tumoral o al menos un marcador inmunitario o un marcador de fase aguda, y (2) al menos un marcador que es (i) un marcador ECM (ii) un marcador que es indicativo de la síntesis de la matriz extracelular (fibrogénesis), o (iii) un marcador que es indicativo de la degradación de la matriz extracelular (fibrólisis)

se puede medir mediante inmunoensayos en sándwich. Dos anticuerpos pueden reaccionar con muestras de fluidos humanos, en donde el anticuerpo de captura se une específicamente a un epítipo del marcador. El segundo anticuerpo de diferente especificidad de epítipo se usa para detectar este complejo. Preferiblemente, los anticuerpos son anticuerpos monoclonales, aunque también pueden emplearse anticuerpos policlonales. Ambos anticuerpos utilizados en los ensayos se unen específicamente a la proteína analito.

Por ejemplo, el ELISA de Her2/neu (Bayer) se puede usar para detectar el dominio extracelular de Her-2/neu en muestras de suero de pacientes con cáncer utilizando dos anticuerpos monoclonales de ratón dirigidos contra el dominio extracelular. EGFr ELISA (Bayer) puede usarse para detectar el dominio extracelular de EGFr en muestras de suero de pacientes con cáncer utilizando dos anticuerpos monoclonales de ratón dirigidos contra el dominio extracelular uPA ELISA (Bayer) puede usarse para detectar el uPA en muestras de suero de pacientes con cáncer utilizando dos anticuerpos monoclonales de ratón dirigidos contra la porción secretada de la proteína. CA 19-9 (Bayer) se puede usar para detectar CA 19-9 en muestras de suero de pacientes con cáncer utilizando dos anticuerpos monoclonales de ratón dirigidos contra la porción secretada de la proteína. CA 15-3® (Bayer) se puede usar para detectar la proteína Muc-1 en muestras de suero de pacientes con cáncer utilizando dos anticuerpos monoclonales de ratón dirigidos contra el producto del gen Muc-1.

Además, un ensayo para el colágeno IV puede usar un anticuerpo monoclonal de Fuji (IV-4H12) (No. de acceso FERM BP-2847) emparejado con un anticuerpo policlonal de Biodesign (T59106R) (No. de catálogo de Biodesign: T59106R). Los ensayos pueden ser inmunoensayos heterogéneos que emplean una técnica de separación de partículas magnéticas.

Un ensayo para PIIINP puede usar un anticuerpo monoclonal Bayer depositado en virtud del Tratado de Budapest el 24 de mayo de 2004 con la American Type Culture Collection, 10801 University Boulevard, Manassas, VA 20110-2209 (ATCC PTA-6013) emparejado con un monoclonal anticuerpo de Hoechst (No. de acceso ECCAC 87042308).

La Tabla 1 a continuación enumera los anticuerpos utilizados para detectar el marcador de ECM, fibrosis y fibrogénesis que se usaron dentro de esta invención.

Tabla 1

Anticuerpo útil para ECM, fibrosis y fibrogénesis.

Gen Marcador	Reactivo	Clon Ab	Proveedor/Desarrollador
Colágeno IV	R1	IV-4H12	ICN
Colágeno IV	R2	T59106R	Biodesign
PIIINP	R1	P3P 296/3/27	Dade Behring
PIIINP	R2	35J23	TSD
Colágeno VI	R1	34C6	TSD
Colágeno VI	R2	34F9	TSD
TIMP-1	R1	PRU-T9	Prof. Clark (UK)
TIMP-1	R2	11E7C6	Connex
Tenascina	R1	23G1	TSD
Tenascina	R2	23G2	TSD
Laminina	R1	67A23	TSD
Laminina	R2	67F8	TSD

ES 2 717 627 T3

Gen Marcador	Reactivo	Clon Ab	Proveedor/Desarrollador
MMP2	R1	85C1	TSD
MMP2	R2	VB31B4	Prof. Windsor (USA)
MMP-9/TIMP-1	R1	11E7C6	Connex
MMP-9/TIMP-1	R2	277.13	Bayer Pharmaceuticals
Ácido hialurónico	R1	HABP*	Bovino
Ácido hialurónico	R2	HABP*	Bovino
*Proteína de unión a ácido hialurónico aislada de cartílago nasal bovino			

La Tabla 2 a continuación enumera secuencias de nucleótidos representativas que pueden expresarse para producir marcadores que son útiles en los métodos de la invención.

Tabla 2

Secuencias representativas de nucleótidos				
Símbolo genético	Descripción del gen	Ref. Secuencias	Unigene_ID	OMIM
MMP-2	preproteína de matriz de metaloproteínasa 2	NM_004530	Hs. 111301	120360
MMP3	preproteína de matriz de metaloproteínasa 3	NM_002422	Hs. 83326	185250
MMP7	preproteína de matriz de metaloproteínasa 7	NM_002423	Hs. 2256	178990
MMP9	preproteína de matriz de metaloproteínasa 9	NM_004994	Hs. 151738	120361
MMP12	preproteína de matriz de metaloproteínasa 12	NM_002426	Hs. 1695	601046
MMP24	matriz de metaloproteínasa 24 (insertada en la membrana)	NM_006690	Hs. 3743	604871
COL1A1	preproteína de colágeno tipo 1 alfa 1	NM_000088	Hs. 172928	120150
COL2A1	Isoforma 1 de colágeno tipo II alfa 1	NM_001844	Hs.81343	120140
COL3A1	colágeno alfa 1 tipo III	NM_000090	Hs. 119571	120180
COL4A1	preproteína de colágeno tipo IV alfa 1	NM_001845	Hs. 119129	120130
COL4A2	preproteína de colágeno tipo IV alfa 2	NM_001846	Hs.75617	120090
COL4A3	Colágeno alfa 3 tipo IV isoforma 1, precursor	NM_000091	Hs.530	120070
COL4A4	precursor de colágeno tipo IV alfa 4	NM_000092	Hs.180828	120131

ES 2 717 627 T3

Secuencias representativas de nucleótidos				
Símbolo genético	Descripción del gen	Ref. Secuencias	Unigene_ID	OMIM
COL4A5	Colágeno alfa 5 tipo IV isoforma 1, precursor	NM_000495	Hs. 169825	303630
COL4A6	Colágeno alfa 6 tipo IV isoforma A, precursor	NM_001847	Hs.408	303631
COL5A1	preproteína de colágeno tipo V alfa 1	NM_000093	Hs. 146428	120215
COL5A2	preproteína de colágeno tipo V alfa 2	NM_000393	Hs.82985	120190
COL5A3	preproteína de colágeno tipo V alfa 3	NM_015719	Hs.235368	120216
COL6A1	preproteína de colágeno tipo VI alfa 1	NM_001848.1	Hs.474053	120220
COL6A2	precursor Alfa 2 tipo VI colágeno isoforma 2C2	NM_001849	Hs. 159263	120240
COL6A3	precursor Alfa 3 tipo VI colágeno isoforma 1	NM_004369	Hs.80988	120250
COL7A1	Alfa 1 tipo VII colágeno precursor	NM_000094	Hs.1640	120120
COL8A1	Alfa 1 tipo VIII colágeno precursor	NM_001850	Hs.114599	120251
COL9A1	Alfa 1 tipo IX colágeno isoforma 1 precursor	NM_001851	Hs.154850	120210
COL9A2	colágeno Alfa 2 tipo IX	NM_001852	Hs.37165	120260
COL9A3	colágeno Alfa 3 tipo IX	NM_001853	Hs.53563	120270
COL10A 1	colágeno, tipo X, precursor alfa 1	NM_000493	Hs. 179729	120110
COL11A 1	preproteína alfa 1 tipo XI colágeno isoforma A	NM_001854	Hs.82772	120280
COL13A 1	alfa 1 tipo XIII colágeno isoforma 1	NM_005203	Hs.211933	120350
COL14A 1	alfa 1 tipo XIV colágeno, precursor	NM_021110	Hs.36131	120324
COL15A 1	alfa 1 tipo XV colágeno, precursor	NM_001855	Hs.83164	120325
COL16A 1	alfa 1 tipo XVI colágeno, precursor	NM_001856	Hs.26208	120326
COL17A 1	colágeno alfa 1 tipo XVII	NM_000494	Hs.117938	113811
COL18A 1	alfa 1 tipo XVIII colágeno, precursor	NM_016214	Hs.78409	120328
COL19A 1	alfa 1 tipo XIX colágeno, precursor	NM_001858	Hs.89457	120165
LAMA2	la subunidad alfa 2 de laminina, precursor	NM_000426	Hs.323511	156225
LAMA3	precursor de la subunidad alfa 3 de laminina	NM_000227	Hs.83450	600805

ES 2 717 627 T3

Secuencias representativas de nucleótidos				
Símbolo genético	Descripción del gen	Ref. Secuencias	Unigene_ID	OMIM
LAMA4	laminina, alfa 4, precursor	NM_002290	Hs.78672	600133
LAMA5	laminina, alfa 5	NM_005560	Hs.312953	601033
LAMB1	laminina, beta 1, precursor	NM_002291	Hs.82124	150240
LAMB2	lamina B2	NM_032737	Hs.76084	150341
LAMB2	laminina, beta 2, precursor	NM_002292	Hs.90291	150325
LAMB3	Laminina subunidad beta 3, precursor	NM_000228	Hs.75517	150310
LAMC1	laminina, gamma 1, precursor	NM_002293	Hs.214982	150290
LAMC2	laminina, gamma 2 isoforma a, precursor	NM_005562	Hs.54451	150292
LAMC3	laminina, gamma 3, precursor	NM_006059	Hs.69954	604349
HXB	tenascina C	NM_002160	Hs.289114	187380
TIMP-1	inhibidor tisular del metaloproteínasa 1, precursor	NM_003254	Hs.5831	305370
PLAU	activador del plasminógeno, uroquinasa	NM_002658	Hs.77274	191840
VEGF	factor de crecimiento endotelial vascular alfa	NM_003376	Hs.73793	192240
CEACA M1	Molécula 1 de adhesión celular relacionada con el antígeno carcinoembrionario (glicoproteína biliar)	NM_001712	Hs.50964	109770
MUC1	mucina 1, transmembrana	NM_002456	Hs.89603	158340
MUC1	mucina 1, transmembrana	NM_182741	Hs.89603	158340
IL2RA	receptor de interleucina 2, precursor de la cadena alfa	NM_000417	Hs.1724	147730
IL6	interleucina 6 (interferon, beta 2)	NM_000600	Hs.93913	147620
GAS	precursor de gastrina	NM_000805	Hs.2681	137250

5 Se pueden hacer anticuerpos para la detección de (1) al menos un marcador tumoral o al menos un marcador inmunitario o al menos un marcador de fase aguda, y (2) al menos un marcador que es (i) un marcador ECM (ii) un marcador que es indicativo de la síntesis de la matriz extracelular (fibrogénesis), o (iii) un marcador que es indicativo de la degradación de la matriz extracelular (fibrólisis), de acuerdo con el Protocolo de Expresión de Polinucleótidos y el Protocolo de Desarrollo de Hibridomas que se describe en detalle más abajo.

Expresión de polinucleótidos:

Para expresar los nucleótidos enumerados en la Tabla 2 y otros genes marcadores relacionados con enfermedades neoplásicas, los genes pueden insertarse en un vector de expresión que contiene los elementos necesarios para la

transcripción y traducción de la secuencia de codificación insertada. Los métodos que son bien conocidos por los expertos en la técnica pueden usarse para construir vectores de expresión que contienen secuencias que codifican polipéptidos marcadores relacionados con enfermedades neoplásicas y elementos de control de transcripción y traducción apropiados. Estos métodos incluyen técnicas de ADN recombinante *in vitro*, técnicas sintéticas y recombinación genética *in vivo*. Tales técnicas se describen, por ejemplo, en Sambrook et al., MOLECULAR CLONING: A LABORATORY MANUAL, 2d ed., (1989) y en Ausubel et al., CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, New York, N.Y. (1989).

Se puede utilizar una variedad de sistemas de vector de expresión/huésped para contener y expresar secuencias que codifican un polipéptido marcador relacionado con la enfermedad neoplásica. Estos incluyen, pero no se limitan a, microorganismos, tales como bacterias transformadas con vectores de expresión de ADN de bacteriófagos, plásmidos o cósmidos recombinantes; levadura transformada con vectores de expresión de levadura, sistemas de células de insecto infectados con vectores de expresión de virus (por ejemplo, baculovirus), sistemas de células vegetales transformados con vectores de expresión de virus (por ejemplo, virus del mosaico de la coliflor, CaMV; virus del mosaico del tabaco, TMV) o con vectores de expresión bacterianos (por ejemplo, plásmidos Ti o pBR322), o sistemas de células animales.

Los elementos de control o las secuencias reguladoras son aquellas regiones de los potenciadores del vector, promotores, regiones no traducidas 5' y 3' que interactúan con las proteínas celulares del huésped para llevar a cabo la transcripción y la traducción. Tales elementos pueden variar en su fuerza y especificidad. Dependiendo del sistema de vectores y del huésped utilizado, se puede usar cualquier número de elementos de transcripción y traducción adecuados, incluidos los promotores constitutivos e inducibles. Por ejemplo, cuando se clona en sistemas bacterianos, se pueden usar promotores inducibles como el promotor lacZ híbrido del fagémido BLUESCRIPT (Stratagene, LaJolla, California) o el plásmido pSPORT1 (Life Technologies) y similares. El promotor de la polihedrina del baculovirus se puede usar en células de insecto. Los promotores o potenciadores derivados de los genomas de células vegetales (por ejemplo, choque térmico, RUBISCO y genes de proteínas de almacenamiento) o de virus de plantas (por ejemplo, promotores virales o secuencias líderes) pueden clonarse en el vector. En sistemas celulares de mamíferos, son preferibles los promotores de genes de mamíferos o de virus de mamíferos. Si es necesario generar una línea celular que contenga múltiples copias de una secuencia de nucleótidos que codifique un polipéptido del "gen de la fibrosis hepática", se pueden usar vectores basados en SV40 o EBV con un marcador seleccionable apropiado.

Sistemas de expresión bacteriana y de levadura:

En sistemas bacterianos, se puede seleccionar cierto número de vectores de expresión dependiendo del uso destinado al polipéptido marcador relacionado con la enfermedad neoplásica. Por ejemplo, cuando se necesita una gran cantidad de polipéptido marcador relacionado con la enfermedad neoplásica para la inducción de anticuerpos, se pueden usar vectores que dirigen la expresión de alto nivel de proteínas de fusión que se purifican fácilmente. Tales vectores incluyen, pero no se limitan a, vectores de expresión y clonación de *E. coli* multifuncionales tales como BLUESCRIPT (Stratagene). En un vector BLUESCRIPT, una secuencia que codifica el polipéptido marcador relacionado con la enfermedad neoplásica se puede ligar en el vector en marco con secuencias para el terminal amino Met y los 7 residuos posteriores de β -galactosidasa para que se produzca una proteína híbrida. Vectores PIN [Van Heeke & Schuster, J. Biol. Chem. 264, 5503-5509, (1989)] o los vectores pGEX (Promega, Madison, WI) también pueden usarse para expresar polipéptidos extraños como proteínas de fusión con glutatión S-transferasa (GST). En general, tales proteínas de fusión son solubles y pueden purificarse fácilmente a partir de células lisadas por adsorción a glutatión glóbulos de agarosa seguido de elución en presencia de glutatión libre. Las proteínas elaboradas en tales sistemas pueden diseñarse para incluir sitios de escisión de heparina, trombina o factor Xa proteasa, de modo que el polipéptido de interés clonado pueda liberarse del resto GST a voluntad.

En la levadura *Saccharomyces cerevisiae*, se puede un cierto número de vectores que contienen promotores constitutivos o inducibles, como el factor alfa, la alcohol oxidasa y la PGH.

Sistemas de expresión en plantas e insectos:

Si se usan vectores de expresión en plantas, la expresión de secuencias que codifican polipéptidos marcadores relacionados con enfermedades neoplásicas puede ser dirigida por cualquiera de varios promotores. Por ejemplo, los promotores virales tales como los promotores 35S y 19S de CaMV pueden usarse solos o en combinación con la secuencia guía omega de TMV [Takamatsu, EMBO J. 6, 307-311, (1987)]. Alternativamente, se pueden usar promotores de plantas tales como la pequeña subunidad de RUBISCO o promotores de choque térmico [Coruzzi et al., EMBO J. 3, 1671-1680, (1984); Broglie et al., Science 224, 838-843, (1984); Winter et al., Results Probl. Cell Differ. 17, 85-105, (1991)]. Estas construcciones pueden introducirse en células vegetales mediante transformación directa de ADN o mediante transfección mediada por patógenos. Tales técnicas se describen en una serie de revisiones generalmente disponibles (por ejemplo, MCGRAW HILL YEARBOOK OF SCIENCE AND TECHNOLOGY, McGraw Hill, New York, N.Y., pp. 191-196, (1992)).

También se puede usar un sistema de insectos para expresar un polipéptido marcador relacionado con la enfermedad neoplásica. Por ejemplo, en uno de estos sistemas, el virus de la polihedrosis nuclear *Autographa californica* (AcNPV) se usa como un vector para expresar genes extraños en células de *Spodoptera frugiperda* o en larvas de *Trichoplusia*. Las secuencias que codifican los polipéptidos marcadores relacionados con la enfermedad neoplásica pueden

clonarse en una región no esencial del virus, tal como el gen de la polihedrina, y colocarse bajo el control del promotor de la polihedrina. La inserción exitosa del polipéptido marcador relacionado con la enfermedad neoplásica hará que el gen de la polihedrina se vuelva inactivo y produzca un virus recombinante que carezca de proteína de la cubierta. Los virus recombinantes se pueden usar para infectar células de *S. frugiperda* o larvas de *Trichoplusia* en las que se pueden expresar polipéptidos marcadores relacionados con enfermedades neoplásicas [Engelhard et al., Proc. Natl. Acad. Sci. 91, 3224 - 3227, (1994)].

Sistemas de expresión de mamíferos:

Se pueden usar varios sistemas de expresión basados en virus para expresar polipéptidos marcadores relacionados con enfermedades neoplásicas en células huésped de mamíferos. Por ejemplo, si se usa un adenovirus como un vector de expresión, las secuencias que codifican los polipéptidos marcadores relacionados con la enfermedad neoplásica se pueden ligar en un complejo de transcripción/traducción de adenovirus que comprende el promotor tardío y la secuencia guía tripartita. La inserción en una región E1 o E3 no esencial del genoma vírico se puede usar para obtener un virus viable que sea capaz de expresar polipéptidos marcadores relacionados con la enfermedad neoplásica en células huésped infectadas [Logan & Shenk, Proc. Natl. Acad. Sci. 81, 3655-3659, (1984)]. Si se desea, los potenciadores de la transcripción, como el potenciador del virus del sarcoma de Rous (RSV), se pueden utilizar para aumentar la expresión en células huésped de mamíferos.

Los cromosomas artificiales humanos (HAC) también se pueden usar para administrar fragmentos más grandes de ADN de los que se pueden contener y expresar en un plásmido. Los HAC de 6M a 10M se construyen y administran a las células mediante métodos de administración convencionales (por ejemplo, liposomas, polímeros de amino policatiónicos o vesículas).

También se pueden usar señales de iniciación específicas para lograr una traducción más eficiente de las secuencias que codifican polipéptidos marcadores relacionados con enfermedades neoplásicas. Tales señales incluyen el codón de iniciación ATG y las secuencias adyacentes. En los casos en que las secuencias que codifican un polipéptido marcador relacionado con la enfermedad neoplásica, su codón de iniciación y las secuencias anteriores se insertan en el vector de expresión apropiado, no se pueden necesitar señales de control transcripcionales o de traducción adicionales. Sin embargo, en los casos en que solo se inserta la secuencia de codificación, o un fragmento de la misma, se deben proporcionar señales de control de traducción exógenas (incluido el codón de iniciación ATG). El codón de inicio debe estar en el marco de lectura correcto para garantizar la traducción de todo el inserto. Los elementos de traducción exógenos y los codones de iniciación pueden ser de diversos orígenes, tanto naturales como sintéticos. La eficiencia de la expresión puede mejorarse mediante la inclusión de potenciadores que sean apropiados para el sistema celular particular que se usa [Scharf et al., Results Probl. Cell Differ. 20, 125-162, (1994)].

Células huésped:

Se puede elegir una cepa de células huésped por su capacidad para modular la expresión de las secuencias insertadas o para procesar el polipéptido marcador relacionado con la enfermedad neoplásica expresada de la manera deseada. Tales modificaciones del polipéptido incluyen, pero no se limitan a, acetilación, carboxilación, glicosilación, fosforilación, lipidación y acilación. El procesamiento postraducción que escinde una forma "prepro" del polipéptido también se puede usar para facilitar la inserción, el plegado y/o la función correctos. Diferentes células huésped que tienen una maquinaria celular específica y mecanismos característicos para actividades postraducción (por ejemplo, CHO, HeLa, MDCK, HEK293 y WI38) están disponibles en la American Type Culture Collection (ATCC; 10801 University Boulevard, Manassas, VA 20110-2209) y puede elegirse para garantizar la correcta modificación y procesamiento de la proteína extraña.

Se prefiere una expresión estable para la producción a largo plazo de proteínas recombinantes de alto rendimiento. Por ejemplo, las líneas celulares que expresan de forma estable los polipéptidos marcadores relacionados con la enfermedad neoplásica pueden transformarse utilizando vectores de expresión que pueden contener orígenes virales de replicación y/o elementos de expresión endógenos y un gen marcador seleccionable en el mismo vector o en un vector separado. Tras la introducción del vector, se puede permitir que las células crezcan durante 12 días en un medio enriquecido antes de cambiarlas a un medio selectivo. El propósito del marcador seleccionable es conferir resistencia a la selección, y su presencia permite el crecimiento y la recuperación de células que expresan con éxito las secuencias del gen del polipéptido marcador relacionado con la enfermedad neoplásica introducida. Los clones resistentes de células transformadas de manera estable pueden proliferar usando técnicas de cultivo de tejidos apropiadas para el tipo de célula. Véase, por ejemplo, Freshney R.I., ed., ANIMAL CELLURE CULTURE (1986).

Se puede usar cualquier número de sistemas de selección para recuperar líneas celulares transformadas. Estos incluyen, pero no se limitan a, la timidina quinasa del virus del herpes simple [Wigler et al., Cell 11, 223-232, (1977)] y adenina fosforibosiltransferasa [Lowy et al., Cell 22, 817-823, (1980)] genes que pueden emplearse en células tk^r o aprt^r, respectivamente. También se puede usar resistencia a antimetabolitos, antibióticos o herbicidas como base para la selección. Por ejemplo, dhfr confiere resistencia al metotrexato [Wigler et al., Proc. Natl. Acad. Sci. 77, 3567-3570, (1980)], npt confiere resistencia a los aminoglucósidos, neomicina y G418 [Colbere-Garapin et. Al., J. Mol. Biol. 150, 114, (1981)], y als y pat confieren resistencia al clorsulfuran y la fosfotricina acetiltransferasa, respectivamente. Se han descrito genes seleccionables adicionales. Por ejemplo, trpB permite que las células utilicen indol en lugar de triptófano, o hisD, lo que permite que las células utilicen histinol en lugar de histidina [Hartman & Mulligan, Proc. Natl.

Acad. Sci. 85, 8047-8051, (1988)]. Marcadores visibles como antocianinas, β -glucuronidasa y su sustrato GUS, y luciferasa y su sustrato luciferina, se pueden usar para identificar transformantes y para cuantificar la cantidad de expresión de proteína transitoria o estable atribuible a un sistema de vector específico [Rhodes et al., *Methods Mol. Biol.* 55, 121-131, (1995)].

5 Detección de expresión y productos genéticos:

Aunque la presencia de la expresión del gen marcador sugiere que también está presente un gen del polipéptido marcador relacionado con la enfermedad neoplásica, puede ser necesario confirmar la presencia y la expresión de ese gen. Por ejemplo, si una secuencia que codifica un polipéptido marcador relacionado con la enfermedad neoplásica se inserta dentro de una secuencia del gen marcador, las células transformadas que contienen secuencias que codifican un polipéptido marcador relacionado con la enfermedad neoplásica pueden identificarse por la ausencia de la función del gen marcador. Alternativamente, un gen marcador puede colocarse en tándem con una secuencia que codifica un polipéptido marcador relacionado con la enfermedad neoplásica bajo el control de un único promotor. La expresión del gen marcador en respuesta a la inducción o selección generalmente indica la expresión del polipéptido marcador relacionado con la enfermedad neoplásica.

15 Alternativamente, las células huésped que contienen polipéptidos marcadores relacionados con enfermedades neoplásicas y que expresan un polipéptido marcador relacionado con enfermedades neoplásicas pueden identificarse mediante una variedad de procedimientos conocidos por los expertos en la técnica. Estos procedimientos incluyen, pero no están limitados a, hibridación de ADN-ADN o ADN-ARN y bioensayo de proteínas o técnicas de inmunoensayo que incluyen tecnologías de membrana, solución o chip para la detección y/o cuantificación de ácido nucleico o proteína. Por ejemplo, la presencia de una secuencia polinucleotídica que codifica un polipéptido marcador relacionado con la enfermedad neoplásica puede detectarse mediante hibridación o amplificación de ADN-ADN o ADN-ARN utilizando sondas o fragmentos o fragmentos de polinucleótidos que codifican un polipéptido marcador relacionado con enfermedad neoplásica. Los ensayos basados en la amplificación de ácidos nucleicos implican el uso de oligonucleótidos seleccionados de secuencias que codifican un polipéptido marcador relacionado con la enfermedad neoplásica para detectar transformantes que contienen un polipéptido marcador relacionado con la enfermedad neoplásica.

Se conoce en la técnica una variedad de protocolos para detectar y medir la expresión de un polipéptido marcador relacionado con la enfermedad neoplásica, utilizando anticuerpos policlonales o monoclonales específicos para el polipéptido. Ejemplos incluyen el ensayo inmunoenzimático (ELISA), el radioinmunoensayo (RIA) y la clasificación de células activadas por fluorescencia (FACS). Se puede usar un inmunoensayo de dos sitios, basado en monoclonal, utilizando anticuerpos monoclonales reactivos a dos epítopos no interferentes en un polipéptido marcador relacionado con la enfermedad neoplásica, o se puede emplear un ensayo de unión competitiva. Estos y otros ensayos se describen en Hampton et al., *SEROLOGICAL METHODS: A LABORATORY MANUAL*, APS Press, St. Paul, Minn., (1990) y Maddox et al., *J. Exp. Med.* 158, 1211-1216, (1983).

Los expertos en la técnica conocen una amplia variedad de marcadores y técnicas de conjugación y se pueden usar en varios ensayos de ácidos nucleicos y aminoácidos. Los medios para producir hibridación marcada o sondas de PCR para detectar secuencias relacionadas con polinucleótidos que codifican polipéptidos marcadores relacionados con enfermedades neoplásicas incluyen marcaje con oligo, traducción de muescas, marcaje final o amplificación por PCR usando un nucleótido marcado. Alternativamente, las secuencias que codifican un polipéptido marcador relacionado con la enfermedad neoplásica pueden clonarse en un vector para la producción de una sonda de ARNm. Dichos vectores son conocidos en la técnica, están disponibles comercialmente y pueden usarse para sintetizar sondas de ARN *in vitro* mediante la adición de nucleótidos marcados y una ARN polimerasa apropiada tal como T7, T3 o SP6. Estos procedimientos se pueden realizar utilizando una variedad de kits disponibles comercialmente (Amersham Pharmacia Biotech, Promega y US Biochemical). Las moléculas o marcadores informadores adecuados que se pueden usar para facilitar la detección incluyen radionúclidos, enzimas y agentes fluorescentes, quimioluminiscentes o cromogénicos, así como sustratos, cofactores, inhibidores, partículas magnéticas y similares.

Expresión y purificación de polipéptidos:

Las células huésped transformadas con secuencias de nucleótidos que codifican un polipéptido marcador relacionado con la enfermedad neoplásica pueden cultivarse en condiciones adecuadas para la expresión y recuperación de la proteína del cultivo celular. El polipéptido producido por una célula transformada puede secretarse o almacenarse intracelularmente dependiendo de la secuencia y/o el vector utilizado. Como entenderán los expertos en la técnica, los vectores de expresión que contienen polinucleótidos que codifican polipéptidos marcadores relacionados con enfermedades neoplásicas pueden diseñarse para contener secuencias señal que dirigen la secreción de polipéptidos marcadores relacionados con enfermedades neoplásicas solubles a través de una membrana de célula procarionota o eucariota o que dirigen la inserción en la membrana de polipéptidos marcadores relacionados con enfermedades neoplásicas relacionadas con el gen de la fibrosis de membrana.

Como se discutió anteriormente, se pueden usar otras construcciones para unir una secuencia que codifica polipéptidos marcadores relacionados con enfermedades neoplásicas a una secuencia de nucleótidos que codifica un dominio polipeptídico que facilitará la purificación de proteínas solubles. Dichos dominios que facilitan la purificación incluyen, pero no se limitan a, péptidos quelantes de metales tales como módulos de histidina-triptófano que permiten

la purificación en metales inmovilizados, dominios de proteína A que permiten la purificación en inmunoglobulina inmovilizada, y el dominio utilizado en el sistema de purificación por extensión/afinidad FLAGS (Immunex Corp., Seattle, Washington). La inclusión de secuencias enlazadoras escindibles, tales como aquellas específicas para el Factor Xa o enteroquinasa (Invitrogen, San Diego, CA) entre el dominio de purificación y el polipéptido marcador relacionado con la enfermedad neoplásica, también se puede usar para facilitar la purificación. Un vector de expresión de este tipo proporciona la expresión de una proteína de fusión que contiene un polipéptido marcador relacionado con la enfermedad neoplásica y 6 residuos de histidina que preceden a una tiorredoxina o un sitio de escisión de enteroquinasa. Los residuos de histidina facilitan la purificación por IMAC (cromatografía de afinidad de iones metálicos inmovilizados, como se describe en Porath et al., *Prot. Exp. Purif.* 3, 263-281 (1992)), mientras que el sitio de escisión de enteroquinasa proporciona un medio para purificar el polipéptido del gen de la fibrosis hepática de la proteína de fusión. Los vectores que contienen proteínas de fusión se describen en Kroll et al., *DNA Cell Biol.* 12, 441-453, (1993).

Síntesis química:

Las secuencias que codifican un polipéptido marcador relacionado con la enfermedad neoplásica pueden sintetizarse, total o parcialmente, usando métodos químicos bien conocidos en la técnica (véase Caruthers et al., *Nucl. Acids Res. Symp. Ser.* 215-223, (1980) y Horn et al. *Nucl. Acids Res. Symp. Ser.* 225-232, (1980). Alternativamente, un polipéptido marcador relacionado con la enfermedad neoplásica puede producirse usando métodos químicos para sintetizar su secuencia de aminoácidos, como por síntesis directa de péptidos utilizando técnicas en fase sólida [Merrifield, *J. Am. Chem. Soc.* 85, 2149-2154, (1963) y Roberge et al., *Science* 269, 202-204, (1995)]. La síntesis de proteínas se puede realizar mediante técnicas manuales o mediante automatización. La síntesis automatizada se puede lograr, por ejemplo, utilizando el sintetizador de péptidos Applied Biosystems 431A (Perkin Elmer). Opcionalmente, los fragmentos de polipéptidos marcadores relacionados con enfermedades neoplásicas se pueden sintetizar por separado y combinar utilizando métodos químicos para producir una molécula de longitud completa.

El péptido recién sintetizado se puede purificar sustancialmente mediante cromatografía líquida preparativa de alto rendimiento [Creighton, *PROTEINS: STRUCTURES AND MOLECULAR PRINCIPLES*, WH y Co., New York, N.Y., (1983)]. La composición de un polipéptido marcador relacionado con la enfermedad neoplásica sintética se puede confirmar mediante análisis de aminoácidos o secuenciación (por ejemplo, el procedimiento de degradación de Edman; véase Creighton. Además, cualquier parte de la secuencia de aminoácidos del polipéptido marcador relacionado con la enfermedad neoplásica puede ser modificada durante la síntesis directa y/o se combinó usando métodos químicos con secuencias de otras proteínas para producir un polipéptido variante o una proteína de fusión.

Protocolo de desarrollo de hibridomas

Fase I: Inmunización.

Los ratones BALB/c y los ratones Swiss Webster (cinco por grupo) se inmunizan por vía intraperitoneal con uno de los marcadores relacionados con la enfermedad neoplásica identificados anteriormente (diferentes dosis) emulsionados con adyuvante completo de Freund (CFA) seguido de tres refuerzos (a intervalos de dos semanas) con inmunógeno emulsionado con adyuvante incompleto de Freund. Los ratones se sangran una semana después de cada refuerzo y los sueros se titulan contra el inmunógeno en ELISA. El ratón con el título más alto se selecciona para la fusión.

Fase II: Fusión celular y selección de hibridomas.

El ratón seleccionado para la fusión se refuerza con la misma dosis de antígeno utilizada en inmunizaciones previas. El refuerzo se administra cuatro días antes de la esplenectomía y la fusión celular. La preparación de antígeno se administra por vía intraperitoneal sin adyuvante.

En el día de la fusión, el ratón se sacrifica y el bazo se retira de forma aséptica. El bazo se tritura con unas pinzas y se filtra a través de un tamiz. Las células se lavan dos veces utilizando medios de Eagle modificados de Iscove (IMDM) y se cuentan utilizando un hemocitómetro.

La línea celular de mieloma de ratón P3x63Ag8.653 se retira de un cultivo estático, fase logarítmica, se lava con IMDM y se cuenta utilizando un hemocitómetro.

Las células de mieloma y bazo se mezclan en una proporción de 1:5 y se centrifugan. El sobrenadante se desecha. El sedimento celular se resuspende suavemente tocando el fondo del tubo. Se agrega gota a gota un mililitro de una solución al 50% de PEG (MW 1450) durante un período de 30 segundos. El sedimento se mezcla suavemente durante 30 segundos con una pipeta. La suspensión celular resultante se deja reposar durante otros 30 segundos. Se agregan cinco mililitros de IMDM durante un período de 90 segundos seguidos por otros 5 ml inmediatamente. La suspensión celular resultante se deja reposar durante 5 minutos. La suspensión celular se hace rotar y el sedimento se resuspende en medio HAT (IMDM que contiene FBS al 10%, L-glutamina 2 mM, 2-mercaptoetanol al 0.6% (solución al 0.04%), hipoxantina, aminopterina, timidina y 10% de crecimiento de origen factor). Las células se resuspenden a 5E5 células por mililitro. Las células se colocan en placas de 96 pozos. Se añaden 200 microlitros o células 2E5 a cada pozo.

Las placas se incuban a 37°C en una atmósfera de 7% de CO₂ con 100% de humedad. Siete días después de la fusión, los medios se eliminan y se reemplazan con IMDM que contiene 10% de FBS, 2 mM de L-glutamina, 0.6% de stock de 2-mercaptoetanol (0.04%), hipoxantina y timidina. Típicamente, las colonias en crecimiento de hibridomas se

observan microscópicamente unos siete días después de la fusión. Estas colonias se pueden ver a simple vista aproximadamente 10-14 días después de la fusión.

De diez a catorce días después de la fusión, el sobrenadante se toma de pozos con colonias de hibridomas en crecimiento. El volumen de sobrenadante es de aproximadamente 150-200 microlitros y contiene 10-100 microgramos de anticuerpo por mililitro. Este sobrenadante se analiza en busca de anticuerpos específicos utilizando el mismo análisis utilizado para detectar los sueros. Las colonias de hibridoma positivas se mueven de la placa de 96 pozos a una placa de 24 pozos. De tres a cinco días después, el sobrenadante de la placa de 24 pozos se analiza para confirmar la presencia de un anticuerpo específico. El volumen de sobrenadante de un pozo de una placa de 24 pozos es de aproximadamente 2 mL y contiene 10-100 microgramos/mL de anticuerpo. Las células de los pozos positivos se expanden en matraces T-25 y T-75. Las células se congelan a partir de matraces T-75. Las células de los pozos positivos también se clonan mediante dilución limitante. Las células de hibridoma se colocan en placas de 96 pozos a una densidad de 0.25 células por pozo o una célula en cada cuarto pozo. Las colonias en crecimiento se analizan 10-14 días más tarde, utilizando los mismos ensayos utilizados para seleccionar inicialmente los hibridomas. Los clones positivos son expandidos y congelados.

Fase III: Producción.

Las células de hibridoma se expandieron a matraces T-162 seguido de la transferencia de éstas a botellas rotativas para la producción de sobrenadante celular. Las células se cultivan en botellas rotativas durante aproximadamente dos semanas hasta que las células son menos del 10% viables. El sobrenadante de cultivo se extrae de estas botellas rotativas para purificación.

Breve descripción de los inmunoensayos.

Todos los anticuerpos son ensayos de tipo ELISA heterogéneos formateados para el sistema Bayer Immuno 1 o placas de 96 pozos. El sistema emplea anticuerpos de captura marcados con fluoresceína (denotado R1) y anticuerpos marcados con latasa de fosfatasa alcalina (denotado R2). Los conjugados de anticuerpos se disuelven en un regulador fisiológico a una concentración entre 2 y 50 mg/L. Los reactivos inmunorreactivos se incuban con una cantidad fija de muestra de paciente que contiene el antígeno por analizar. La muestra del paciente siempre se pipetea primero en una cubeta de reacción seguida de R1 treinta segundos después. Normalmente se agrega R2 30 segundos a 20 minutos después de la adición de R1. La mezcla se incuba durante un máximo de 20 minutos, aunque otras realizaciones de los inmunoensayos pueden requerir más tiempos de incubación más cortos. Posteriormente, se añaden partículas inmunomagnéticas a la mezcla. Las partículas consisten en óxido de hierro que contiene perlas de poliacrilamida con anticuerpos anti-fluoresceína conjugados con la superficie de la partícula. Las partículas están disponibles comercialmente en Bayer HealthCare Diagnostics.

Tras la incubación de las partículas inmunomagnéticas con el inmunocomplejo sándwich formado a partir del antígeno y los conjugados R1 y R2, el inmunocomplejo sándwich se captura a través del marcador de fluoresceína del anticuerpo R1 por los anticuerpos anti-fluoresceína sobre las partículas inmunomagnéticas. El supercomplejo formado es precipitado por un campo magnético externo. Todo el material no unido, especialmente el conjugado fosfato alcalino R2 se elimina mediante lavado. El complejo lavado se resuspende en solución de p-nitrofenolfosfato. La tasa de formación de color es proporcional a la cantidad de fosfatasa que queda en la cubeta, que es proporcional a la cantidad de antígeno. La cuantificación se logra al registrar una curva de calibración de seis puntos y una curva de calibración, construida por una regresión cúbica o un ajuste de Rodbard.

(a) Rendimiento del ensayo.

El rendimiento de cada uno de los ensayos se determina de forma aislada. La sensibilidad y la especificidad, la variación inter e intraensayo, las interferencias, la linealidad y el paralelismo se determinan para cada inmunoensayo. Se determinan los rangos de resultados obtenidos para sujetos sanos de ambos sexos y un rango de edades de 18 a 75 años, para establecer valores "normales". Los ensayos se aplican a sujetos con una variedad de trastornos patológicos.

La invención se ilustra adicionalmente en los siguientes ejemplos no limitativos.

Ejemplo 1

Tratamiento de pacientes con cáncer colorrectal

Resumen

Se logró una discriminación estadísticamente significativa de la supervivencia general del paciente (p menor que aproximadamente el nivel de 0.05 cuando se calculó con gráficos de Kaplan-Meier) (incluso en el análisis de un solo parámetro) utilizando los métodos de la invención. Se compararon niveles elevados o disminuidos de marcadores séricos con niveles de control normales o niveles medios ajustados de cohortes enfermas. La importancia de los marcadores individuales se determinó calculando los gráficos de Kaplan-Meier de los pacientes (utilizando el cuartil superior o inferior de los niveles de marcadores individuales). Una disminución o un aumento en los niveles de los marcadores en el paciente con cáncer en comparación con los niveles en los controles normales indicó un aumento

5 en el estadio, el grado, la gravedad, el avance o la progresión del cáncer del paciente y/o una falta de eficacia o beneficio del tratamiento o terapia del cáncer. En particular, los altos niveles de Gastrina, CA 19-9, TIMP-1 y el bajo nivel de EGFr, MMP-2 se correlacionaron con un mal pronóstico. Además, el análisis combinado de niveles altos de colágeno VI, tenascina, uPA y niveles bajos de PIINP, VEGF se correlacionó con un buen pronóstico. Algunos parámetros séricos singulares produjeron valores medios estadísticamente significativos y diferenciaron las cohortes según las diferencias en los puntos finales del estudio.

Metodología clínica

10 Se estudiaron cuarenta y cuatro pacientes que padecían carcinoma colorrectal metastásico al hígado. El carcinoma primario se confirmó histológicamente. También se obtuvo confirmación histológica para metástasis hepáticas sincrónicas. Cuando se identificó la metástasis hepática metacrónica, la confirmación histológica solo se realizó cuando las técnicas de imagen (tomografía computarizada en espiral (TC) del abdomen o MRT del hígado) no mostraron resultados claros.

15 Los pacientes recibieron quimioterapia de primera línea, que consistió en una infusión semanal de ácido folínico de 1 a 2 horas (500 mg m^{-2}) seguida de una infusión de 5-fluorouracilo (2600 mg m^{-2}) durante 24 horas. Un ciclo comprendió seis infusiones semanales seguidas de 2 semanas de descanso. Un total de 23 pacientes recibieron oxaliplatino cada dos semanas (85 mg m^{-2}) y tres pacientes también recibieron irinotecán una vez por semana (80 mg m^{-2}). La respuesta al tratamiento se controló cada 8 semanas mediante TC espiral y la actividad antitumoral se evaluó de acuerdo con los criterios de la OMS. La duración media del tratamiento fue de 7 meses. La tabla 3 a continuación enumera los tamaños de tumores ajustados por computotomografía en cada ciclo de terapia para evaluar la respuesta del tumor al tratamiento.

Tabla 3

Evaluación clínica								
ID del paciente	Pretratamiento	después del 1 ^{er} ciclo	después del 2 ^o ciclo	después del 3 ^{er} ciclo	después del 4 ^o ciclo	después del 5 ^o ciclo	después del 6 ^o ciclo	Tamaño del tumor % del tamaño inicial
G 42	4.8	3.6	3.6	3.6	3.6	7.5		75
G 52	91.6	97.1	72.3	72.3	79.4			74.4
G 53	132.8	54	16	54				3.4
G 56	10.6	10.6	10.6	2.4				22.7
G 60	36	26	21.1	18.9				52.5
G 226	9	12.3						136
G 73	15.2	11.2	8.4	4	4	6.3	18	26.3
G 79	34	18.5	5.6	4.5	9.5			13.2
G 85	180	104	69	61.8	65.2			34.3
G 86	216.3	31.8	19.8	30.6				9.1
G 87	9.3	1.8	0					0
G 88	182.2	73.2	43.3	26				14.2
G 92								

ES 2 717 627 T3

Evaluación clínica								
ID del paciente	Pretratamiento	después del 1 ^{er} ciclo	después del 2 ^o ciclo	después del 3 ^{er} ciclo	después del 4 ^o ciclo	después del 5 ^o ciclo	después del 6 ^o ciclo	Tamaño del tumor % del tamaño inicial
G 96	116.2	62.8	42	28.3				24,2
G 98	14.3	3	1.4	1.4				9.8
G100	13.3	9	6.3	1	0.3	0.25	0.3	1.9
G101	9	3.3	3.2	1.4	1.2			13
G103	3.3	1.8	2.3	2.3	3.8			54.5
G111	15.2	11.8	8.4	8				52.6
G116	49	9		2.4				4.9
G119	5.3	4	1.8					34
G131	4	4	0.8	0.5				12.5
G218	12.3	0						0
G136	21	13.5	6.3	4	6			19
G138	102	37.5	13.3					13
G148	25	25	16	20.3				64,0
G151	60	33	22.4	20	20	30.3		33
G152	32	15.4	8.2	6.3				19.7
G154	110.3	36	18	10.6				9.6
G166	30	21.4	12	12				40
G169	45.3	27	27	35.3				59.6
G170	25	14.7	8.4	8.4	8.4	9	7.5	30
G173	22	16	6	6.2				27.2
G177	3.2	5.3						160
G178	225	143	125.4					55.7
G179	16	13.7	9	7.5	22.1			46.8

5 Se obtuvo suero de cada paciente inmediatamente antes del tratamiento y se tomaron muestras longitudinales de suero en cada ciclo. Se determinaron los siguientes parámetros de suero y plasma: MMP2, TIMP1, MMP9, Colágeno IV, ColágenoVI, PIIINP, Tenascina, Laminina, CEA, CA15-3, CA19-9, sHer-2/neu, EGFR, uPA y PAI-1. Los pacientes se clasificaron según su supervivencia global y la supervivencia libre de enfermedad.

Ejemplo 2

Determinación de valores de predicción y derivación de algoritmos relacionados

Resumen

10 Se analizaron muestras de suero obtenidas de cada paciente como se describe en el Ejemplo 1 y se usaron valores de nivel de marcador de enfermedad neoplásica para generar valores predictores algorítmicos que se correlacionaban con la supervivencia del paciente.

Transformaciones de datos

15 Se reportaron valores para los siguientes diecisiete marcadores antes del inicio de la quimioterapia y durante cada uno de los ciclos de quimioterapia descritos a continuación: MMP2, TIMP1, MMP9, Colágeno IV, Colágeno VI, PIIINP, Tenascina, Laminina, CEA, CA19-9, sHer-2/neu, EGFR, uPA, PAI-1, Gastrina, IL2R e IL6.

Las tablas 4A y 4B muestran datos experimentales determinados por mediciones duplicadas o por triplicado para cada uno de los 17 marcadores indicados en la muestra de suero pretratamiento.

Tabla 4A

Datos experimentales y determinación de umbral.												
ID del paciente	Supervivencia	Supervivencia	CO1037.6	CO 674	CO 316	CO 33.7	Corte 1083	Corte 9.17	Corte 7.2	Corte 15ng		
ID	Estatus	Mes	TIMP-1	MMP-2	COLIV	Laminina	Tenascina	PIIINP	Col VI	Her2/neu		
			[ng/ml]	[ng/ml]	[ng/ml]	[ng/ml]	[ng/ml]	[ng/ml]	[ng/ml]	[ng/ml]		
G 111	vivo	41	468.7	540.9	105.3	18.8	358.3	5.9	8.4	9.66		
G 60	vivo	35	665.8	553.7	176.9	23.4	533.0	7.8	7.1	12.1		
G 208	vivo	23	471.6	1128.7	151.6	12.5	287.1	6.6	5.5	12.55		
G 18	vivo	42	653.8	819.3						9.3		
G 20	vivo	33	648.3	416.4	160.7	22.7	323.4	4.9	4.3	7.3		
G 226	vivo	21	1242.2	432.6	229.7	35.5	774.5	25.1	6.0	8.9		
G 14	vivo	33	1897.2	483.9	590.9	60.8	1122.8	30.1	5.7	16.9		
G 88	vivo	30	1085.7	520.9	366.0	27.5	470.5	22.5	5.7	12.4		
G 116	vivo	42	917.5	554.9	184.3	48.7	973.0	12.7	5.0	6.86		
G 13	vivo	47	1022.9	541.5	216.7	45.6	2199.3	21.6	7.7	9.8		
G 87	vivo	61	848.7	1620.1	671.6	108.0	2364.7	71.8	22.0	20.41		
G 100	vivo	45	1528.4	1079.7	510.4	73.7	1021.9	29.8	17.0	11.22		
G 119	vivo	42	640.6	920.0	232.3	32.0				9.76		

Datos experimentales y determinación de umbral.												
ID del paciente	Supervivencia	Supervivencia	CO1037.6	CO 674	CO 316	CO 33.7	Corte 1083	Corte 9.17	Corte 7.2	Corte 15ng		
ID	Estatus	Mes	TIMP-1	MMP-2	COLIV	Laminina	Tenascina	PIIINP	Col VI	Her2/neu		
			[ng/ml]	[ng/ml]	[ng/ml]	[ng/ml]	[ng/ml]	[ng/ml]	[ng/ml]	[ng/ml]		
G 57	vivo	45	639.7	817.4	210.5	25.1	175.4	8.4	4.4	7.7		
G 98	vivo	40	821.9	838.1	170.7	31.9	821.4	20.2	9.9	10.24		
G 148	muerto	22	1420.2	464.4	329.1	31.3	979.4	19.8	4.9	10.18		
G 103	muerto	26	502.7	757.2	117.5	19.2	315.5	5.2	8.4	9.62		
G 169	muerto	9	1220.0	428.8	165.1	19.4	276.6	12.7	4.7	8.57		
G 182	muerto	9	1580.0	465.5	247.6	24.8	851.1	16.0	6.6	13.99		
G 19	muerto	6	671.1	542.6	149.9	24.3	728.2	6.0	5.9	6.6		
G 196	muerto	15	1387.6	438.7	312.6	31.6	750.0	21.2	5.1	25.02		
G 42	muerto	25	728.6	495.0	166.7	24.2	842.7	14.2	8.2	5.9		
G 92	muerto	30	765.2	619.0	155.3	24.5	1150.8	8.4	5.3	10.38		
G 52	muerto	11	1381.1	564.0	273.4	44.2	739.0	15.7	6.2	7.9		
G 33	muerto	15	658.4	611.2	186.7	44.8	410.3	11.0	6.0	8.2		
G 49	muerto	11	2020.1	643.0	559.6	53.3	2201.7	26.5	5.7	55.8		
G 178	muerto	11	1523.5	451.3	341.8	29.4	829.3	25.4	4.6	11.2		

Datos experimentales y determinación de umbral.												
ID del paciente	Supervivencia	Supervivencia	CO1037.6	CO 674	CO 316	CO 33.7	Corte 1083	Corte 9.17	Corte 7.2	Corte 15ng		
ID	Estatus	Mes	TIMP-1	MMP-2	COLIV	Laminina	Tenascina	PIIINP	Col VI	Her2/neu		
			[ng/ml]	[ng/ml]	[ng/ml]	[ng/ml]	[ng/ml]	[ng/ml]	[ng/ml]	[ng/ml]		
G 15	muerto	16	1193.5	1173.4						7.7		
G 79	muerto	16	835.1	571.7	194.5	44.4	492.8	10.1	6.7	5.7		
G 85	muerto	18	1297.0	522.7	294.0	32.1	1128.5	18.2	4.1	6.95		
G 96	muerto	23	741.8	326.8	161.5	38.7	577.6	11.7	2.3	7.36		
G 218	muerto	18	805.6	767.5	196.4	27.6	577.0	25.4	4.0	14.6		
G 86	muerto	13	1801.1	753.2	581.7	76.4	849.7	27.3	7.2	8.48		
G 192	muerto	15	553.7	577.9	128.4	10.5	331.5	4.7	3.9	13.28		
G 152	muerto	37	461.5	313.1	112.5	12.7	315.0	4.1	2.4	2.33		
G 73	muerto	22	623.9	591.6	136.2	16.9	678.0	4.9	2.4	4.6		
G 101	muerto	35	720.3	662.7	171.6	28.1	341.9	3.8	10.7	9.76		
G 136	muerto	15	852.3	588.0	148.0	36.5	327.5	6.9	5.0	14.60		
G 53	muerto	7	1882.2	518.3	243.5	69.7	817.4	13.0	10.0	6.8		
G 179	muerto	14	1068.0	460.0	204.5	29.0	563.9	11.4	5.8	10.70		
G 131	muerto	58	587.4	790.2	199.0	24.6	448.7			9.3		

Datos experimentales y determinación de umbral.												
ID del paciente	Supervivencia	Supervivencia	CO1037.6	CO 674	CO 316	CO 33.7	Corte 1083	Corte 9.17	Corte 7.2	Corte 15ng		
ID	Estatus	Mes	TIMP-1	MMP-2	COLIV	Laminina	Tenascina	PIIINP	Col VI	Her2/neu		
			[ng/ml]	[ng/ml]	[ng/ml]	[ng/ml]	[ng/ml]	[ng/ml]	[ng/ml]	[ng/ml]		
G 138	muerto	14	1159.4	599.3	371.5	42.1	931.0	14.7	5.1	13.55		
G 170	muerto	24	506.5	611.2	142.5	18.1	288.9	5.8	5.0	10.54		
G 151	muerto	28	919.0	599.2	238.2	34.0	853.3	21.5	5.9	7.8		
G 154	muerto	14	515.0	765.6	226.9	20.2	489.5	9.6	6.6	10.62		
G 184	muerto	28	566.5	527.1	138.3	15.5	315.9	3.7	5.8	11.33		
G 173	muerto	32	570.4	395.0	127.0	15.4	454.6	9.1	3.3	5.9		
G 166	muerto	7	511.7	486.7	114.6	17.5	234.6	8.4	2.7	7.44		

Tabla 4B

Datos experimentales y determinación de umbral											
	Supervivencia	Supervivencia	EGFR	TIMP-1	uPA	VEGF	CEA	CA 19-9	IL2R	IL6	Gastrina
ID	estatus	mes	[ng/ml]	[ng/ml]	[pg/ml]	[pg/ml]					
G 111	vivo	41	49.12	142.9	803.6	162.5	1.9	10.0	382.0	5.0	58.0
G 60	vivo	35	58.42	218.8	1225.9	171.2	6.8	6.0	372.0	5.0	
G 208	vivo	23	61.53	53.3	749.5	166.0					
G 18	vivo	42	50.68	169.9	1490.5	162.1	0.6	26.0	1336.0	31.0	15.0
G 20	vivo	33	46.94	188.0	1307.5	169.5	894.0	2.0			
G 226	vivo	21	38.22	439.6	2206.6	169.1	1098.0	1449.0	1215.0	5.0	
G 14	vivo	33	49.89	573.3	2936.7	346.9	1614.0	4596.0	414.0	5.0	
G 88	vivo	30	50.21	377.3	1912.4	165.8	52.8	17.0	532.0	5.0	
G 116	vivo	42	54.28	337.9	946.7	292.9	6.0	3.0	847.0	5.0	15.0
G 13	vivo	47	53.05	320.8	1689.5	170.6	191.4	4136.0	640.0	14.6	19.0
G 87	vivo	61	106.21	321.8		172.5	9.1	37.0			11.0
G 100	vivo	45	44.87	337.7	1623.1	165.8	45.1	166.0	519.0	5.0	23.0
G 119	vivo	42					3.3	2.0			22.0
G 57	vivo	45	20.57	244.0	1105.6	163.3	8.0	30.0	1251.0	5.0	16.0
G 98	vivo	40	62.7	316.0	1736.5	196.2	0.5	2.0	690.0	5.0	23.0

ES 2 717 627 T3

Datos experimentales y determinación de umbral											
	Supervivencia	Supervivencia	EGFR	TIMP-1	uPA	VEGF	CEA	CA 19-9	IL2R	IL6	Gastrina
ID	estatus	mes	[ng/ml]	[ng/ml]	[pg/ml]	[pg/ml]					
G 148	muerto	22	65.86	346.9	1685.6	436.9	5.6	72.0	769.0	13.1	30.0
G 103	muerto	26	43.12	86.3	888.6	170.6	12.1	12.0	350.0	5.0	
G 169	muerto	9	42.43	357.4	950.6	194.4	57.3	1833.0	971.0	5.0	36.0
G 182	muerto	9	42.2	417.1	1673.9	166.4	1700.0	30.0	834.0	5.0	
G 19	muerto	6	35.28	199.9	1000.9	162.5	6.8	2.0	606.0	5.0	18.0
G 196	muerto	15	100.1	523.3	2265.7	342.5	279.9	1964.0	1301.0	5.0	-99
G 42	muerto	25	15.7	244.2	927.3	237.5	171.0	667.0	510.0	5.0	20.0
G 92	muerto	30	65.6	203.5	1412.6	674.8	33.4	17.0	367.0	5.0	13.0
G 52	muerto	11	44.1	382.8	1979.0	297.7	39.2	1618.0	765.0	5.0	132.0
G 33	muerto	15	58.30	173.3	772.6	196.0	4.4	53.0	328.0	5.0	
G 49	muerto	11	43.8	574.3	2018.2	294.6	2050.0	5866.0	756.0	5.0	
G 178	muerto	11	43.4	318.3	2537.4	167.4	2952.0	1102.0	1072.0	5.0	36.0
G 15	muerto	16	44.18	337.4	1787.2	230.9	210.0	120.0			75.0
G 79	muerto	16	36.18	253.5	1241.5	162.1	25.5	37.0	492.0	5.0	22.0
G 85	muerto	18	40.48	382.8	1709.0	304.0	1620.0	2.0	1132.0	5.0	33.0
G 96	muerto	23	39.7	209.4	-99	162.2	7050.0	90.0	354.0	17.5	13.0

ES 2 717 627 T3

Datos experimentales y determinación de umbral											
	Supervivencia	Supervivencia	EGFR	TIMP-1	uPA	VEGF	CEA	CA 19-9	IL2R	IL6	Gastrina
ID	estatus	mes	[ng/ml]	[ng/ml]	[pg/ml]	[pg/ml]					
G 218	muerto	18	45.80	240.7	1024.2	165.4	4.4	91.0	463.0	5.0	
G 86	muerto	13	63.19	484.0	1709.0	316.1	690.0	175.0	1235.0	5.0	
G 192	muerto	15	64.0	170.9	927.3	166.9	7.7	4.0	417.0	5.0	19.0
G 152	muerto	37	29.83	110.1	502.6	162.4	335.8	690.0			19.0
G 73	muerto	22	28.23	182.5	1284.2	163.0	11.1	217.0			12.0
G 101	muerto	35	61.34	258.8	556.5	165.2	12.1	17.0			16.0
G 136	muerto	15	83.26	227.8	1292.0	170.3	3112.0	1582.0	549.0	5.0	
G 53	muerto	7	56.78	494.9	1517.8	167.3	59.9	96.0	1299.0	9.2	16.0
G 179	muerto	14	55.2	336.8	1245.3	366.4	25.6	439.0	2285.0	5.0	33.0
G 131	muerto	58	56.0	130.8			5	12			
G 138	muerto	14	49.17	375.6	1237.6	278.7	46.4	9.0	744.0	5.0	46.0
G 170	muerto	24	41.07	148.1	425.6	169.4	16.6	32.0	381.0	5.0	23.0
G 151	muerto	28	47.85	260.1	1218.2	274.4	67.4	2.0	616.0	5.0	
G 154	muerto	14	60.69	150.7	931.2	162.1	441.8	2.0	524.0	5.0	31.0
G 184	muerto	28	37.7	417.1	718.6	192.6	7.6	57.0	433.0	5.0	
G 173	muerto	32	25.3	189.5	707.0	169.1	1.5	9.0	1058.0	5.0	26.0

ES 2 717 627 T3

Datos experimentales y determinación de umbral											
	Supervivencia	Supervivencia	EGFR	TIMP-1	uPA	VEGF	CEA	CA 19-9	IL2R	IL6	Gastrina
ID	estatus	mes	[ng/ml]	[ng/ml]	[pg/ml]	[pg/ml]					
G 166	muerto	7	42.60	198.1	1066.8	205.3	277.2	782.0	393.0	5.0	40.0

Para asegurar la posibilidad de comparación entre los marcadores, se obtuvo el logaritmo natural (base e) de cada valor de marcador.

Imputación de datos

- 5 Hasta el 18% de los valores para cualquier variable predictiva dada faltaban en el conjunto de datos. El análisis del valor faltante (SPSS versión 11) se realizó en las transformadas de registro de las variables de ensayo. Sobre la base de un modelo de regresión múltiple general, los valores faltantes se imputaron para casos incompletos.

Modelo de regresión de Cox

- 10 Se desarrolló un modelo de regresión de Cox utilizando el conjunto completo de datos con valores imputados. La eliminación por pasos en retroceso produjo un modelo con cinco covariables.

La tabla 5 presenta resultados de ejemplo de un análisis de regresión de cox utilizando todas las variables, incluidos los datos imputados.

Tabla 5

Resultados de la regresión de Cox

- 15 Modelos de regresión seleccionados por criterio de puntuación

Número de puntos		
Variables Chi-cuadrado Variables incluidas en el modelo		
1	3.4467	MMP_2ng_ml_Promedio
1	3.3318	Tumor_final_con_respecto_a_tumor_original
1	2.4987	Colágeno_VI_ng_ml_Promedio
1	2.2177	Gastrina
1	1.9357	PIIINP_ng_ml_Promedio
1	1.3859	Tenascina_ng_ml_Promedio
1	0.7419	Laminina_ng_ml_Promedio
1	0.6592	VEGF_pg_ml_Promedio
1	0.5635	TIMP_1_ng_ml_Promedio
1	0.4845	IL2R

ES 2 717 627 T3

1	0.4121	CEA
1	0.2969	Género_1
1	0.2802	IL6
1	0.2521	TIMP_1_ng_ml_Promedio_1
1	0.1229	COLÁGENo_IV_ng_ml_Promedio
1	0.1203	CA_19_9
1	0.1062	Edad_en_diagnóstico_inicial
1	0.0940	Her2_neu_ng_ml_Promedio
1	0.0691	uPA_pg_ml_Promedio
1	0.0297	EGFR_ng_ml_Promedio
2	13.3130	PIIINP_ng_ml_Promedio TIMP_1_ng_ml_Promedio
2	6.9550	PIIINP_ng_ml_Promedio TIMP_1_ng_ml_Promedio_1
2	6.1894	Laminina_ng_ml_Promedio TIMP_1_ng_ml_Promedio
2	5.3721	Colágeno_VI_ng_ml_Promedio TIMP_1_ng_ml_Promedio
2	5.2526	MMP_2_ng_ml_Promedio TIMP_1_ng_ml_Promedio
2	4.9809	MMP_2ng_ml_Promedio Tenascina_ng_ml_Promedio
2	4.9108	Gastrina MMP_2_ng_ml_Promedio
2	4.7721	Colágeno_VI_ng_ml_Promedio Gastrina
2	4.6743	Colágeno_VI_ng_ml_Promedio Tumor_final_con_respecto_a_tumor_original
2	4.4602	IL6 MMP_2_ng_ml_Promedio
2	4.4306	COLAGENo_IV_ng_ml_Promedio MMP_2_ng_ml_Promedio
2	4.4006	Tumor_final_con_respecto_a_tumor_original Tenascina ng_ml_Promedio
2	4.3305	Tumor_final_con_respecto_a_tumor_original TIMP_1_ng_ml_Promedio
2	4.3235	Tumor_final_con_respecto_a_tumor_original MMP_2ng_ml_Promedio
2	4.3036	Edad_en_diagnóstico_inicial MMP_2_ng_ml_Promedio
2	4.2922	Tumor_final_con_respecto_a_tumor_original PIIINP_ng_ml_Promedio

2	4.2883	Edad_en_diagnóstico_inicial Final Tumo _con_respecto_a_tumor_original
2	4.2579	PIIINP_ng_ml_Promedio uPA_pg_ml_Promedio
2	4.1569	MMP_2ng_ml_Promedio TIMP_1_ng_ml_Promedio 1
2	4.1331	Tumor_final_con_respecto_a_tumor_original TIMP_1_ng_ml_Promedio_1
3	17.1662	Edad en diagnóstico inicial PIIINP_ng_ml_Promedio TIMP_1_ng_ml_Promedio
3	14.5799	Tumor_final_con_respecto_a_tumor_original PIIINP_ng_ml_Promedio TIMP_1_ng_ml_Promedio
3	14.4887	Laminina_ng_ml_Promedio PIIINP_ng_ml_Promedio TIMP_1_ng_ml_Promedio
3	14.0809	PIIINP_ng_ml_Promedio TIMP_1ng_ml_Promedio VEGF_pg_ml_Promedio
3	13.9596	Género_1 PIIINP_ng_ml_Promedio TIMP_1_ng_ml_Promedio
3	13.8212	CA_19_9 PIIINP_ng_ml_Promedio TIMP_1ml_Promedio
3	13.8127	PIIINP_ng_ml_Promedio TIMP_1_ng_ml_Promedio uPA_pg_ml_Promedio
3	13.7812	IL2R PIIINP_ng_ml_Promedio TIMP_1_ng_ml_Promedio
3	13.7473	COLÁGENO_IV_ng_ml_Promedio PIIINP_ng_ml_Promedio TIMP_1_ng_ml_Promedio
3	13.6423	Gastrina PIIINP_ng_ml_Promedio TIMP_1_ng_ml_Promedio
3	13.5334	MMP_2_ng_ml_Promedio PIIINP_ng_ml_Promedio TIMP_1_ng_ml_Promedio
3	13.4361	CEA PIIINP_ng_ml_Promedio TIMP_1_ng_ml_Promedio
3	13.3831	IL6 PIIINP_ng_ml_Promedio TIMP_1_ng_ml_Promedio

En el análisis, los aumentos en TIMP-1 y GASTRINA se asocian con aumentos en el riesgo de fracaso. Los aumentos en los valores de Tenascina, Colágeno VI y UPA se asocian con disminuciones en el riesgo de fracaso. La estadística de Wald se utilizó para determinar la importancia de cada parámetro de estimación. La estadística se calcula como

$$Wald = \left(\frac{B}{s.e.B} \right)^2$$

La estadística se distribuye como una distribución de Chi-cuadrado con un grado de libertad.

Determinación de los valores de predicción por regresión de Cox

Las estimaciones de parámetros enumeradas en la Tabla 6A se usaron para calcular un valor de predictor Z para cada paciente. El algoritmo de valor del predictor es:

$$Z = 4.48 \ln(\text{TIMP-1}) + 0.92(\text{GASTRINA}) - 2.08 \ln(\text{TENASCINA}) - 1.1 \ln(\text{ColágenoVI}) - 1.56 \ln(\text{UPA})$$

Estos valores se utilizaron en un análisis ROC. La tabla 6B muestra las coordenadas de la curva ROC. El área bajo la curva (AUC) para estos datos fue de 0.8 (IC del 95%: 0.67 a 0.94), lo que indica una asociación significativa con el fracaso.

ES 2 717 627 T3

Las Tablas 6A y 6B, que enumeran las estimaciones del Parámetro de Regresión de Cox y las coordenadas ROC que se determinaron de acuerdo con el experimento del Ejemplo 2 del presente documento.

Parámetros de regresión de Cox estimados y coordenadas ROC

Bifurcación y análisis de Kaplan Meier.

- 5 Los valores Z del predictor se bifurcaron a un valor de 8.62. El examen de las Tablas 6A y 6B indica que a este valor, la fracción positiva verdadera (TPF) es 0.81 y la fracción negativa verdadera (TNF) es 0.6. La Tabla 6B ilustra las curvas de supervivencia de Kaplan-Meier para esta división de cohorte en una Z de 8.62. Una prueba de log-rank indica que estas curvas son significativamente diferentes. (LR=11.08, p=0.0009). La supervivencia media para los pacientes cuyo valor predictivo Z fue inferior a 8.62 (BC) fue de 58 meses. Para los pacientes con valores superiores a este punto de corte (UC), la supervivencia media fue de 18 meses.

Tabla 6A

Estimaciones de los parámetros de regresión de Cox			
Variable	Parámetro (B)	Estadística de Wald	p
Ln(TIMP-1)	4.48	13.9	0.000
Ln(GASTRINA)	0.94	5.46	0.019
Ln(TENASCINA)	-2.08	5.22	0.022
Ln(Colágeno VI)	-1.10	5.21	0.022
Ln(UPA)	-1.56	4.36	0.037

TABLA 6B

Coordenadas ROC

Z	TPF	TNF	TP	TN	FP	FN
4.08	100.0%	6.7%	33	1	14	0
6.69	100.0%	13.3%	33	2	13	0
7.18	97.0%	13.3%	32	2	13	1
7.70	97.0%	20.0%	32	3	12	1
7.71	97.0%	26.7%	32	4	11	1
Z	TPF	TNF	TP	TN	FP	FN
7.85	97.0%	33.3%	32	5	10	1
8.09	97.0%	40.0%	32	6	9	1
8.12	93.9%	40.0%	31	6	9	2
8.13	90.9%	40.0%	30	6	9	3
8.17	90.9%	46.7%	30	7	8	3
8.18	90.9%	53.3%	30	8	7	3
8.19	87.9%	53.3%	29	8	7	4
8.51	84.8%	53.3%	28	8	7	5
8.53	84.8%	60.0%	28	9	6	5

ES 2 717 627 T3

8.62	81.8%	60.0%	27	9	6	6
8.72	78.8%	60.0%	26	9	6	7
8.73	75.8%	60.0%	25	9	6	8
8.77	75.8%	66.7%	25	10	5	8
8.79	75.8%	73.3%	25	11	4	8
8.82	72.7%	73.3%	24	11	4	9
8.84	69.7%	73.3%	23	11	4	10
8.87	66.7%	73.3%	22	11	4	11
8.91	66.7%	80.0%	22	12	3	11
8.91	63.6%	80.0%	21	12	3	12
9.20	60.6%	80.0%	20	12	3	13
9.27	60.6%	86.7%	20	13	2	13
9.32	57.6%	86.7%	19	13	2	14
9.56	54.5%	86.7%	18	13	2	15
9.60	51.5%	86.7%	17	13	2	16
9.65	48.5%	86.7%	16	13	2	17
9.72	45.5%	86.7%	15	13	2	18
9.77	42.4%	86.7%	14	13	2	19
9.79	42.4%	93.3%	14	14	1	19
9.86	39.4%	93.3%	13	14	1	20
9.87	36.4%	93.3%	12	14	1	21
10.01	33.3%	93.3%	11	14	1	22
10.16	30.3%	93.3%	10	14	1	23
10.29	30.3%	100.0%	10	15	0	23
10.30	27.3%	100.0%	9	15	0	24
10.32	24.2%	100.0%	8	15	0	25
10.36	21.2%	100.0%	7	15	0	26
10.40	18.2%	100.0%	6	15	0	27
10.41	15.2%	100.0%	5	15	0	28
10.47	12.1%	100.0%	4	15	0	29
11.00	9.1%	100.0%	3	15	0	30
11.47	6.1%	100.0%	2	15	0	31
12.00	3.0%	100.0%	1	15	0	32
12.19	0.0%	100.0%	0	15	0	33

Análisis Kaplan Meier de conjuntos de marcadores singulares y combinados.

Para cada corte de marcador singular, los valores se determinaron como se establece en las Tablas 4A y 4B. Posteriormente, se realizó el análisis de Kaplan Meier para cada uno de los marcadores singulares. Como se

representa en las figuras 2-7, esta partición en "Por debajo del punto de corte" ("BC") (que se estableció como el valor numérico "0") y en "Por encima del punto de corte" ("UC") (que fue establecido como el valor numérico "1"), permitió la bifurcación y la discriminación estadísticamente significativa de pacientes con resultados clínicos buenos y malos (es decir, tiempo de supervivencia general).

5 La tabla 5 presenta los resultados del análisis de Kaplan Meier de un solo parámetro utilizando los valores de corte para cada uno de los marcadores seleccionados de la Tabla 1.

10 Como se muestra en las Figuras 2-8, Gastrina, CA19-9, TIMP-1 (Immuno-1), MMP-2 y EGFR produjeron resultados estadísticamente significativos a un nivel de $p=0.05$ para los valores umbral indicados. VEGF y CEA mostraron una tendencia hacia la significación estadística a un nivel de 0.08 para los valores de umbral indicados. Como se muestra en la Tabla 1, los valores de corte indicados de cada uno de los marcadores individuales se transformaron en los valores numéricos 1 o 0, dependiendo de si las mediciones individuales estaban por encima o por debajo del valor de corte, respectivamente.

15 Estos valores se usaron para desarrollar algoritmos simples basados en parámetros dicotómicos. Como se indica en la Tabla 7, se derivó un algoritmo ejemplar "MCT-V" (fila I) mediante la adición de los valores dicotómicos de MMP-2 (fila L), Colágeno VI (fila N) y Tenascina (fila P) y sustracción del valor dicotómico de VEGF (fila R). Luego se utilizaron los valores de la suma para la división en dos grupos ("UC">1 y "BC"<1"), y posteriormente se empleó el análisis de Kaplan Meier.

La Tabla 7 representa la evaluación de los valores del algoritmo MCT-V.

Tabla 7

Análisis combinatorio de parámetros dicotómicos											
		"MCT-V"	Supervivencia	Corte 674		Corte 7,2		Corte 1083		>Media 221,1	
		Algorithm	Mes	MMP-2		Colágeno VI		Tenascina		VEGF	
ID	Supervivencia			[ng/ml]		[ng/ml]		[ng/ml]		[pg/ml]	
G 111	0	1	41	540,9	0	8,4	1	358,3	0	162,5	0
G 60	0	0	35	553,7	0	7,1	0	533,0	0	171,2	0
G 208	0	1	23	1128,7	1	5,5	0	287,1	0	166,0	0
G 18	0	1	42	819,3	1		0		0	162,1	0
G 20	0	0	33	416,4	0	4,3	0	323,4	0	169,5	0
G 226	0	0	21	432,6	0	6,0	0	774,5	0	169,1	0
G 14	0	0	33	483,9	0	5,7	0	1122,8	1	346,9	1
G 88	0	0	30	520,9	0	5,7	0	470,5	0	165,8	0
G 116	0	-1	42	554,9	0	5,0	0	973,0	0	292,9	1
G 13	0	2	47	541,5	0	7,7	1	2199,3	1	170,6	0
G 87	0	3	61	1620,1	1	22,0	1	2364,7	1	172,5	0
G 100	0	2	45	1079,7	1	17,0	1	1021,9	0	165,8	0
G 119	0	1	42	920,0	1		0		0		0
G 57	0	1	45	817,4	1	4,4	0	175,4	0	163,3	0
G 98	0	2	40	838,1	1	9,9	1	821,4	0	196,2	0
G 148	1	-1	22	464,4	0	4,9	0	979,4	0	436,9	1
G 103	1	2	26	757,2	1	8,4	1	315,5	0	170,6	0
G 169	1	0	9	428,8	0	4,7	0	276,6	0	194,4	0
G 182	1	0	9	465,5	0	6,6	0	851,1	0	166,4	0

ES 2 717 627 T3

G 19	1	0	6	542,6	0	5,9	0	728,2	0	162,5	0
G 196	1	-1	15	438,7	0	5,1	0	750,0	0	342,5	1
G 42	1	0	25	495,0	0	8,2	1	842,7	0	237,5	1
G 92	1	0	30	619,0	0	5,3	0	1150,8	1	674,8	1
G 52	1	-1	11	564,0	0	6,2	0	739,0	0	297,7	1
G 33	1	0	15	611,2	0	6,0	0	410,3	0	196,0	0
G 49	1	0	11	643,0	0	5,7	0	2201,7	1	294,6	1
G 178	1	0	11	451,3	0	4,6	0	829,3	0	167,4	0
G 15	1	0	16	1173,4	1		0		0	230,9	1
G 79	1	0	16	571,7	0	6,7	0	492,8	0	162,1	0
G 85	1	0	18	522,7	0	4,1	0	1128,5	1	304,0	1
G 96	1	0	23	326,8	0	2,3	0	577,6	0	162,2	0
G 218	1	1	18	767,5	1	4,0	0	577,0	0	165,4	0
G 86	1	0	13	753,2	1	7,2	0	849,7	0	316,1	1
G 192	1	0	15	577,9	0	3,9	0	331,5	0	166,9	0
G 152	1	0	37	313,1	0	2,4	0	315,0	0	162,4	0
G 73	1	0	22	591,6	0	2,4	0	678,0	0	163,0	0
G 101	1	1	35	662,7	0	10,7	1	341,9	0	165,2	0
G 136	1	0	15	588,0	0	5,0	0	327,5	0	170,3	0
G 53	1	1	7	518,3	0	10,0	1	817,4	0	167,3	0
G 179	1	-1	14	460,0	0	5,8	0	563,9	0	366,4	1
G 131	1	1	58	790,2	1		0	448,7	0		0
G 138	1	-1	14	599,3	0	5,1	0	931,0	0	278,7	1
G 170	1	0	24	611,2	0	5,0	0	288,9	0	169,4	0
G 151	1	-1	28	599,2	0	5,9	0	853,3	0	274,4	1
G 154	1	1	14	765,6	1	6,6	0	489,5	0	162,1	0
G 184	1	0	28	527,1	0	5,8	0	315,9	0	192,6	0
G 173	1	0	32	395,0	0	3,3	0	454,6	0	169,1	0
G 166	1	0	7	486,7	0	2,7	0	234,6	0	205,3	0

TABLA 8

Comparación de curvas de supervivencia (supervivencia mes/porcentaje de supervivencia)											
	Gastrina CO 25.4	CA 19-9 CO 37	TIMP-1 Immuno CO 1037.6	MMP-2 CO 675	EGFr CO 45	VEG F CO 221,1	CEA CO 100	MMP 2 Tenascina - VEG F	Col6	TIMP-1 CO 949 y EGFr CO 45	
Prueba de Logrank											
Chi cuadrado	7.237	7.485	6.757	5.208	3.896	3.279	3.052	10.75		4.557	
df	1	1	1	1	1	1	1	1		1	
Valor de P	0.0071	0.0062	0.0093	0.0225	0.0484	0.070 2	0.0806	0.001 0		0.0328	
Resumen de valores de p	**	**	**	*	*	ns	ns	**		*	
¿Son las curvas de supervivencia significativamente diferentes?	Si	Si	Si	Si	Si	No	No	Si		Si	
Supervivencia media											
	Gastrina alta	CA 19-9 alta	TIMP-1 Immuno alto	MMP2>6 75	EGFr<4 5	VEG F alta	CEA>100	MCT-V alta		TIMP-1 alto y EGFr bajo	
Dato 1:	14.00	16.00	14.00	58.00	22.00	17.00	16.00	58.00		11.00	
	Gastrina baja	CA 19-9 bajo	TIMP-1 Inmuno Bajo	MMP2<6 75	EGFr>4 5	VEG F bajo	CEA<100	MCT-V bajo		TIMP-1 bajo y/o EGFr normal	
Dato 1:	30.00	35.00	30.00	22.00	30.00	28.00	30.00	18.00		28.00	
Relación	0.4667	0.4571	0.4667	2.636	0.7333	0.607 1	0.5333	3.222		0.3929	
95% CI de relación	0.01269 0.9206	-0.04426 0.9585	- 0.03078 0.9641	2.223 a 3.049	0.2294 1.237	0.115 1 1.099	0.03946 1.027	2.809 a 3.635		- 0.04119 a 0.8269	
Relación Hazard											

Comparación de curvas de supervivencia (supervivencia mes/porcentaje de supervivencia)											
	Gastrina CO 25.4	CA 19-9 CO 37	TIMP-1 Immuno CO 1037.6	MMP-2 CO 675	EGFr CO 45	VEG F CO 221,1	CEA CO 100	MMP 2 Col6 Tenascina - VEG F	TIMP-1 CO 949 y EGFr CO 45		
Prueba de Logrank											
Chi cuadrado	7.237	7.485	6.757	5.208	3.896	3.279	3.052	10.75	4.557		
df	1	1	1	1	1	1	1	1	1		
Valor de P	0.0071	0.0062	0.0093	0.0225	0.0484	0.070 2	0.0806	0.001 0	0.0328		
Resumen de valores de p	**	**	**	*	*	ns	ns	**	*		
¿Son las curvas de supervivencia significativamente diferentes?	Si	Si	Si	Si	Si	No	No	Si	Si		
Supervivencia media											
	Gastrina alta	CA 19-9 alta	TIMP-1 Immuno alto	MMP2>6 75	EGFr<4 5	VEG F alta	CEA>100	MCT-V alta	TIMP-1 alto y EGFr bajo		
Relación	2.716	2.443	2.362	0.3913	1.928	1.865	1.821	0.275 3	2.346		
95% CI de relación	1.440 a 10.18	1.342 a 5.936	1.327 a 7.509	0.1967 a 0.8839	1.005 a 4.282	0.938 6 a 4.964	0.9164 a 4.571	0.140 2 a 0.609 9	1.109 a 11.32		
	Gastrina alta/ Gastrina baja	CA 19-9 alta/ CA 19-9 baja	TIMP-1 Immuno alto I TIMP-1 Immuno bajo	MMP2>6 75/ MMP2<6 75	EGFr<4 5/ EGFr>4 5	VEG F alto/ VEG F bajo	CEA>100/ CEA<100	MCT-V alto/ MCT-V bajo	TIMP-1 alto y EGFr bajo/ TIMP-1 bajo y/o EGFr normal		
Número o filas	48	48	48	48	48	48	48	48	48		

Las figuras 10 y 10A representan el análisis de Kaplan Meier de los respectivos valores del algoritmo "MCT-V".

Pruebas estadísticas múltiples

5 Los datos séricos también se transformaron para el análisis en el software Genedata Expressionist™. La población de pacientes se dividió en "respondedores" y "no respondedores" como se muestra en la fila D o "supervivientes (supervivencia de más de 40 meses)" y "no supervivientes (muertos dentro de los 18 meses)" como se muestra en la fila G. Posteriormente, se han realizado múltiples pruebas estadísticas utilizando las técnicas estadísticas T-Test, Welch, Kolmogorov-Smirnov y Wilcoxon. Se muestran los valores de p resultantes para las pruebas estadísticas respectivas.

10 La Tabla 9 muestra los resultados de múltiples pruebas estadísticas para discriminar a los pacientes con CCR metastásico que sobrevivieron durante más de 40 meses o menos de 18 meses desde el tratamiento primario mediante la evaluación de los parámetros séricos.

Tabla 9

Pruebas estadísticas múltiples - Análisis de supervivencia global					
CEA	0.005	0.003	0.044	0.007	1
cm2 pre-terapia	0.025	0.030	0.013	0.022	2
Colágeno VI	0.008	0.078	0.011	0.043	3
MMP-2	0.011	0.049	0.030	0.039	4
Gastrina	0.054	0.046	0.053	0.059	5
TIMP-1	0.154	0.127	0.116	0.147	6
CA 19-9	0.167	0.181	0.266	0.176	7
Laminina	0.236	0.281	0.430	0.275	8
VEGF	0.221	0.160	0.609	0.318	9
Tenascina	0.326	0.457	0.160	0.231	10
PIIINP	0.275	0.401	0.617	0.417	11
TIMP-1	0.581	0.537	0.193	0.422	12
IL6	0.190	0.448	0.996	0.380	13
uPA	0.609	0.614	0.877	0.605	14
EGFr	0.959	0.970	0.433	0.524	15
IL2R	0.635	0.600	0.882	0.687	16
Her-2/neu	0.764	0.709	0.816	1.000	17
Colágeno_IV_A1	0.87949997	0.89459997	0.8664	0.93790001	18

Los valores de p respectivos se indican para cada uno de los marcadores medidos. Se ha realizado la prueba de suma de rangos para elegir marcadores óptimos para el análisis posterior, como el análisis de componentes principales. Como se indica, CEA, el tamaño inicial del tumor, el Colágeno VI, MMP-2 y la gastrina fueron estadísticamente significativos para discriminar entre "supervivientes" y "no supervivientes" mediante el uso de diversas pruebas estadísticas para analizar las variables continuas.

La figura 10 y la figura 10A muestra la partición inicial en dos grupos cuando se usan los 17 parámetros de la Tabla 9. "Supervivientes" se muestran como bolas verdes y "no supervivientes" se muestran como bolas rojas.

La figura 11 y la figura 11A muestra la partición mejorada en dos grupos por análisis de componentes principales (PCA) cuando se utilizan los 5 parámetros más importantes (es decir, CEA, tamaño inicial del tumor, Colágeno VI, MMP-2 y gastrina) mostrados en la Tabla 9. Los "supervivientes" se muestran como las bolas verdes y los "no supervivientes" se muestran como bolas rojas.

Ejemplo 3

Análisis de expresión de tejido tumoral primario y metastásico mediante análisis de tejido tumoral incrustado en parafina

Resumen

Se analizaron tejidos incrustados en parafina, fijados con formalina, de resectados quirúrgicos del paciente como se describe en el Ejemplo 1 y se determinaron los valores de los niveles de marcadores de enfermedades neoplásicas mediante técnicas de qRT-PCR y se correlacionaron con la supervivencia del paciente.

Perfiles de expresión utilizando RT-PCR cinética cuantitativa.

El ARN se aisló a partir de tejidos fijados con formalina y embebidos en parafina (= tejidos FFPE). Los expertos en la técnica pueden realizar procedimientos de extracción de ARN. Por ejemplo, el ARN total de un rizo de 5 a 10 μm de tejido tumoral FFPE se puede extraer utilizando el kit de parafina de ARN de alta pureza (Roche, Basilea, Suiza), cuantificado por el ensayo de cuantificación de ARN Ribogreen (Molecular Probes, Eugene, OR) y calificado por RT-PCR de fluorescencia en tiempo real de un fragmento de RPL37A. En general, se ensayaron de 0.5 a 2 ng de ARN de cada extracción de ARN cualificada mediante qRT-PCR como se describe a continuación. Para un análisis detallado de la expresión génica mediante métodos de PCR cuantitativos, se utilizarán cebadores que flanquean la región genómica de interés y una sonda marcada con fluorescencia que se hibrida en el medio. Utilizando el sistema de detección de secuencias PRISM 7700 o 7900 de PE Applied Biosystems (Perkin Elmer, Foster City, CA, Estados Unidos) con la técnica de una sonda fluorogénica, que consiste en un oligonucleótido etiquetado con un tinte reportero fluorescente y un tinte extintor, uno puede realizar tal medida de expresión. La amplificación del producto específico de la sonda provoca la escisión de la sonda, generando un aumento en la fluorescencia del indicador. Los cebadores y las sondas se seleccionaron utilizando el software Primer Express y se localizaron principalmente en los bordes de exón/intrón y las grandes secuencias no transcritas intermedias (>800 pb) para garantizar la especificidad del ARN o dentro de la región 3' de la secuencia de codificación o en el 3' región no traducida. El diseño de cebadores y la selección de una región objetivo apropiada es bien conocido por los expertos en la materia. El cebador y las sondas predefinidas para los genes enumerados en la Tabla 2 también se pueden obtener de proveedores, por ejemplo, Applied Biosystems de PE. Se comprobó la especificidad de todos los pares de cebadores mediante reacciones de PCR convencionales y electroforesis en gel. Para estandarizar la cantidad de muestra de ARN, GAPDH, RPL37A, RPL9 y CD63 se seleccionaron como referencias, ya que no estaban reguladas diferencialmente en las muestras analizadas. Para realizar tal análisis de expresión de genes dentro de muestras biológicas, los cebadores/sondas respectivos se preparan mezclando 25 μl de la solución madre 100 μM de "Cebador Superior", 25 μl de la solución madre 100 μM de "Cebador Inferior" con 12.5 μl de la solución madre 100 μM de solución TaqMan (FAM/Tamra) y se ajustó a 500 μl con agua destilada (mezcla cebador/sonda). Para cada reacción, se mezclaron 1.25 μl de cADN de las muestras de pacientes con 8.75 μl de agua libre de nucleasas y se agregaron a un pozo de una placa de reacción de 96 pozos ópticos (Applied Biosystems Part No. 4306737). A continuación, se agregan 1.5 μl de la mezcla Cebador/Sonda descrita anteriormente, 12.5 μl de Taq Man Universal-PCR-mix (2x) (Applied Biosystems Part No. 4318157) y 1 μl de agua. Las placas de 96 pozos se cierran con 8 tapas/tiras (Pat No. Applied Biosystems 4323032) y se centrifugan durante 3 minutos. Las mediciones de la reacción de PCR se realizan de acuerdo con las instrucciones del fabricante con un TaqMan 7700 de Applied Biosystems (No. 20114) en condiciones apropiadas (2 min. 50°C, 10 min. 95°C, 0.15 min. 95°C, 1 min. 60°C; 40 ciclos). Antes de la medición de las muestras biológicas hasta ahora no clasificadas, los experimentos de control, por ejemplo, se podrían usar líneas celulares, muestras de control sanas, muestras de respuesta de terapia definida para la estandarización de las condiciones experimentales.

Se realizaron experimentos de validación con TaqMan que demostraron que las eficiencias del objetivo y las amplificaciones de control son aproximadamente iguales, lo que es un requisito previo para la cuantificación relativa de la expresión génica mediante el método comparativo $\Delta\Delta\text{CT}$, conocido por los expertos en la técnica. Por lo tanto, el software SDS 2.0 de Applied Biosystems se puede utilizar de acuerdo con las instrucciones respectivas. Los valores de CT se analizan posteriormente con el software apropiado (Microsoft Excel™) de los paquetes de software estadístico (SAS).

Además de la tecnología descrita anteriormente, proporcionada por Perkin Elmer, se pueden usar implementaciones de otras técnicas como Lightcycler™ de Roche Inc. o iCycler de Stratagene Inc., capaces de detectar en tiempo real una reacción de RT-PCR.

5 La figura 12 y la figura 12A muestra la expresión relativa de los miembros de la familia tirosina quinasa del receptor ERB en tejidos FFPE de resectados de tumores primarios de pacientes como se describe en el Ejemplo 1 y como se determina mediante el perfil de qRT-PCR. Los genes se muestran en líneas. La supervivencia de los pacientes se representa arriba de cada fila, con 1 o 0 que significa "muerto" o "vivo" y los números entre paréntesis significan el mes de supervivencia desde el diagnóstico primario.

10 Como se representa, la expresión de los miembros de la familia EGFR se correlaciona con la respuesta clínica de la metástasis hepática de pacientes con CCR que se tratan con un régimen basado en 5'FU según lo determinado por las determinaciones de TC de las lesiones metastásicas. La respuesta clínica se denota como "Respuesta parcial" (= PR o barra de color verde en la parte superior), "Enfermedad estable" (= SD o barra de color naranja en la parte superior) y "Enfermedad progresiva" (= PD o barra de color rojo oscuro en la parte superior). La supervivencia se representa para cada paciente sobre cada columna (supervivencia = 0 o muerte = 1 seguida de un mes de supervivencia entre paréntesis [x mes]). Es evidente que la sobreexpresión de al menos un miembro de la familia ERB es evidente en el grupo de mal pronóstico, es decir, la cohorte de pacientes SD y PD que no responde. La expresión particularmente alta de EGFR en el tumor primario se correlaciona con una respuesta no favorable al tratamiento antitumoral. Esto se demostró aún más realizando múltiples pruebas estadísticas como se muestra en la Tabla 10 (independientemente del método de normalización).

20 La Tabla 10 muestra los resultados de múltiples pruebas estadísticas para discriminar a los pacientes con CCR metastásico cuyas lesiones metastásicas responden a un régimen basado en 5'FU (Respuesta parcial) o no responden (enfermedad estable y enfermedad progresiva) mediante la determinación de ARN de miembro de la familia EGFR en muestras de tejido FFPE.

Tabla 10

Pruebas estadísticas múltiples - Respuesta clínica - Análisis FFPE de miembros de la familia ERB						
Nombre del gen	Descripción del gen	Prueba T	Welch	Kolmogorov - Smirnov	Wilcoxon	Suma de rango
Her2/neu	normalizado a la media de RPL37A	0.01977	0.02106	0.05303	0.03788	1
EGFR	normalizado a la media de RPL37A	0.02762	0.02805	0.05303	0.02622	2
EGFR II	normalizado a la media de RPL37A GAPDH, RPL9, CD63	0.0397	0.03977	0.2121	0.05303	3
EGFR I	normalizado a la media de GAPDH	0.05634	0.05636	0.2121	0.09732	4
Her2/neu I	normalizado a la media de GAPDH	0.15549999	0.1556	0.05303	0.07284	5
ERBB3 II	normalizado a la media de RPL37A, GAPDH, RPL9, CD63	0.0906	0.09065	0.2121	0.12819999	6
ERBB3	normalizado a la media de RPL37A	0.06432	0.06656	0.57520002	0.2243	7
Her2/neu II	normalizado a la media de RPL37A, GAPDH, RPL9, CD63	0.083	0.08317	0.57520002	0.1649	8
VEGF-C I	normalizado a la media de GAPDH	0.22149999	0.2237	0.2121	0.21969999	9

Pruebas estadísticas múltiples - Respuesta clínica - Análisis FFPE de miembros de la familia ERB						
Nombre del gen	Descripción del gen	Prueba T	Welch	Kolmogorov - Smirnov	Wilcoxon	Suma de rango
VEGF-C II	normalizado a la media de RPL37A, GAPDH, RPL9, CD63	0.2326	0.235	0.2121	0.1373	10
VEGF-C	normalizado a la media de RPL37A	0.23989999	0.243	0.2121	0.1543	11

La alta expresión de ARNm de EGFR en tumores primarios de pacientes con mal pronóstico contrasta con el bajo nivel sérico de EGFr en suero de pacientes con mal pronóstico. Sin embargo, como los niveles séricos de EGFr y TIMP-1 eran simultáneamente altos en pacientes con mal pronóstico, los niveles comparativamente bajos de EGFr sérico aparentemente reflejan la degradación reducida de EGFr por las proteinasas en lugar de la expresión reducida dentro del tejido tumoral, que están sorprendentemente elevadas. Esto es de importancia crítica para las estrategias terapéuticas dirigidas a los miembros de la familia de receptores de EGF (como, por ejemplo, Iressa®, Erbitux® o Herceptin®), que son inesperadamente útiles en particular en pacientes con niveles bajos de EGFr en suero. Además, de acuerdo con los datos representados en la figura 12 y la figura 12A, la organización de la red de miembros de la familia ERB es de vital importancia para el resultado clínico. Los tumores colorrectales que expresan altos niveles de EGFR y simultáneamente bajos niveles de Her-2/neu tienen una supervivencia global significativamente más corta, que los pacientes con altos niveles de EGFR y Her-2/neu. Esto parece reflejar impactos biológicos muy diferentes de los receptores ERB hetero o homodimerizados en la tumorigénesis y el resultado clínico de las terapias contra el cáncer. Supuestamente, la composición de la red ERB influye entre otras cosas en la tasa de proliferación, por lo que es de gran importancia para los agentes quimioterapéuticos antiproliferativos, como los regímenes basados en 5FU. Esto explicaría en parte el sorprendente hallazgo de que los tumores CCR positivos a Her-2/neu tienen un mejor pronóstico que los tumores negativos a Her-2/neu.

En línea con esto, el análisis combinado de TIMP-1 y EGFr en muestras de suero pretratamiento identificó una población de alto riesgo de pacientes con niveles altos de TIMP-1 y EGFr, que mostraron un resultado peor (supervivencia general de 11 meses) en comparación con la evaluación de un solo parámetro.

La Tabla 11 muestra datos experimentales determinados por mediciones duplicadas o por triplicado para TIMP-1 y EGFr en la muestra de suero pretratamiento y su análisis combinado.

Tabla 11

Datos séricos de TIMP-1 y EGFR											
ID	Edad de diagnóstico	Respuesta	Respuesta	Supervivencia Estatus	Supervivencia mes	Supervivencia	TIMP-1 ALT y EGFR bajo	TIMP-1 [ng/ml]	EGFR [ng/ml]		
G 111	39	SD	0	vivo	41	0	0	468,7	0	49,12	0
G 60	60	SD	0	vivo	35	0	0	665,8	0	58,42	0
G 208	62	-99	0	vivo	23	0	0	471,6	0	61,53	0
G 18	63	SD	0	vivo	42	0	0	653,8	0	50,68	0
G 20	63	SD	0	vivo	33	0	0	648,3	0	46,94	0
G 226	72	PD	0	vivo	21	0	1	1242,2	1	38,22	1
G 14	43	PR	1	vivo	33	0	0	1897,2	1	49,89	0
G 88	50	PR	1	vivo	30	0	0	1085,7	1	50,21	0
G 116	52	PR	1	vivo	42	0	0	917,5	0	54,28	0
G 13	60	PR	1	vivo	47	0	0	1022,9	0	53,05	0
G 87	60	CR	1	vivo	61	0	0	848,7	0	106,21	0
G 100	61	PR	1	vivo	45	0	1	1528,4	1	44,87	1
G 119	67	PR	1	vivo	42	0	0	640,6	0		0
G 57	71	PR	1	vivo	45	0	0	639,7	0	20,57	1
G 98	71	PR	1	vivo	40	0	0	821,9	0	62,7	0

Datos séricos de TIMP-1 y EGFr												
ID	Edad de diagnóstico	Respuesta	Respuesta	Supervivencia Estatus	Supervivencia mes	Supervivencia	TIMP-1 ALT y EGFR bajo	TIMP-1 [ng/ml]	EGFR [ng/ml]			
G 148	34	SD	0	muerto	22	1	0	1420,2	1	65,86	0	
G 103	52	SD	0	muerto	26	1	0	502,7	0	43,12	1	
G 169	55	SD	0	muerto	9	1	1	1220,0	1	42,43	1	
G 182	59	SD	0	muerto	9	1	1	1580,0	1	42,2	1	
G 19	61	SD	0	muerto	6	1	0	671,1	0	35,28	1	
G 196	61	SD	0	muerto	15	1	0	1387,6	1	100,1	0	
G 42	62	SD	0	muerto	25	1	0	728,6	0	15,7	1	
G 92	63	-99	0	muerto	30	1	0	765,2	0	65,6	0	
G 52	66	SD	0	muerto	11	1	1	1381,1	1	44,1	1	
G 33	70	SD	0	muerto	15	1	0	658,4	0	58,30	0	
G 49	70	-99	0	muerto	11	1	1	2020,1	1	43,8	1	
G 178	70	SD	0	muerto	11	1	1	1523,5	1	43,4	1	
G 15	74	SD	0	muerto	16	1	1	1193,5	1	44,18	1	
G 79	43	PR	1	muerto	16	1	0	835,1	0	36,18	1	
G 85	46	PR	1	muerto	18	1	1	1297,0	1	40,48	1	
G 96	46	PR	1	muerto	23	1	0	741,8	0	39,7	1	

Datos séricos de TIMP-1 y EGFr												
ID	Edad de diagnóstico	Respuesta	Respuesta	Supervivencia Estatus	Supervivencia mes	Supervivencia	TIMP-1 ALT y EGFR bajo	TIMP-1 [ng/ml]	EGFR [ng/ml]			
G 218	51		CR	1	18	1	0	805,6	0	45,80	0	0
G 86	57		PR	1	13	1	0	1801,1	1	63,19	0	0
G 192	57		PR	1	15	1	0	553,7	0	64,0	0	0
G 152	58		PR	1	37	1	0	461,5	0	29,83	1	1
G 73	59		PR	1	22	1	0	623,9	0	28,23	1	1
G 101	59		PR	1	35	1	0	720,3	0	61,34	0	0
G 136	59		PR	1	15	1	0	852,3	0	83,26	0	0
G 53	61		PR	1	7	1	0	1882,2	1	56,78	0	0
G 179	62		PR	1	14	1	0	1068,0	1	55,2	0	0
G 131	64		PR	1	58	1	0	587,4	0	56,0	0	0
G 138	66		PR	1	14	1	0	1159,4	1	49,17	0	0
G 170	66		PR	1	24	1	0	506,5	0	41,07	1	1
G 151	67		PR	1	28	1	0	919,0	0	47,85	0	0
G 154	70		PR	1	14	1	0	515,0	0	60,69	0	0
G 184	72		PR	1	28	1	0	566,5	0	37,7	1	1
G 173	73		PR	1	32	1	0	570,4	0	25,3	1	1

Datos séricos de TIMP-1 y EGFr										
ID	Edad de diagnóstico	Respuesta	Respuesta	Supervivencia Estatus	Supervivencia mes	Supervivencia	TIMP-1 ALT y EGFR bajo	TIMP-1 [ng/ml]	EGFR [ng/ml]	
G 166	75	PR	1	muerto	7	1	0	511,7	0	42,60
										1

La Figura 13 ilustra las curvas de supervivencia de Kaplan-Meier del análisis combinado de los niveles séricos de TIMP-1 y EGFr

Ejemplo 4

Análisis de expresión de tejido tumoral primario y metastásico mediante análisis de biopsias de tejido tumoral fresco

5 Resumen

Se analizaron las biopsias de pacientes como se describe en el Ejemplo 1 y se realizaron análisis de expresión amplia del genoma mediante tecnologías de matriz y se correlacionaron con la supervivencia del paciente.

Las sondas específicas de las secuencias de polinucleótidos de la Tabla 2 y la Tabla 11 se obtienen como sigue.

10 Las sondas polinucleotídicas se inmovilizan en un chip de ADN en una matriz organizada. Los oligonucleótidos se pueden unir a un soporte sólido mediante una variedad de procesos, incluida la litografía. Por ejemplo, un chip puede contener hasta 410 000 oligonucleótidos (GeneChip, Affymetrix).

15 Se obtiene una muestra biológica (por ejemplo, una muestra de biopsia que se fracciona opcionalmente por medio de un criostato para enriquecer las células enfermas en aproximadamente el 80% de la población celular total, o una muestra de líquidos corporales como suero u orina, suero o célula que contiene líquidos, por ejemplo, derivado de aspirados con aguja fina). El ADN o ARN luego se extrae, amplifica y analiza con un chip de ADN para determinar la presencia o ausencia de secuencias polinucleotídicas marcadoras. Las sondas de polinucleótidos se siembran sobre un sustrato en una matriz o matriz bidimensional. Las muestras de polinucleótidos se marcan y luego se hibridan con las sondas. Los polinucleótidos de doble cadena, que comprenden los polinucleótidos de muestra marcados unidos a polinucleótidos de sonda, pueden detectarse una vez que la porción no unida de la muestra se elimina por lavado.

20 Los polinucleótidos de la sonda pueden sembrarse sobre sustratos que incluyen vidrio, nitrocelulosa, etc. Las sondas se pueden unir al sustrato mediante enlaces covalentes o mediante interacciones no específicas, como las interacciones hidrófobas. Los polinucleótidos de muestra pueden marcarse utilizando marcadores radiactivos, fluoróforos, cromóforos, etc. Las técnicas para construir matrices y los métodos para usar estas matrices se describen en el documento EP0 799 897; WO 97/29212; WO 97/27317; EP 0 785 280; WO 97/02357; Patente de los Estados Unidos No. 5,593,839; Patente de los Estados Unidos No. 5,578,832; EP 0 728 520; Patente de los Estados Unidos No. 5,599,695; EP 0 721 016; Patente de los Estados Unidos No. 5,556,752; WO 95/22058; y la Patente de los Estados Unidos No. 5,631,734. Además, las matrices se pueden usar para examinar la expresión diferencial de los genes y se pueden usar para determinar la función de los genes. Por ejemplo, se pueden usar matrices de las secuencias de polinucleótidos instantáneas para determinar si alguna de las secuencias de polinucleótidos se expresa diferencialmente entre células normales y células enfermas, por ejemplo. La alta expresión de un mensaje particular en una muestra enferma, que no se observa en una muestra normal correspondiente, puede indicar una proteína específica del cáncer.

Análisis de datos de experimentos de perfiles de expresión.

35 De acuerdo con la técnica de medición de Affymetrix (Affymetrix GeneChip Expression Analysis Manual, Santa Clara, CA), una medición de expresión de un solo gen en un chip produce el valor de diferencia promedio y la llamada absoluta. Cada chip contiene 16-20 pares de sondas de oligonucleótidos por gen o clon de cADN. Estos pares de sondas incluyen conjuntos perfectamente emparejados y conjuntos no coincidentes, ambos de los cuales son necesarios para el cálculo de la diferencia promedio, o valor de expresión, una medida de la diferencia de intensidad para cada par de sondas, calculada restando la intensidad de la falta de coincidencia de la intensidad de la pareja perfecta. Esto tiene en cuenta la variabilidad en la hibridación entre pares de sondas y otros artefactos de hibridación que podrían afectar las intensidades de fluorescencia. La diferencia promedio es un valor numérico que se supone representa el valor de expresión de ese gen. La llamada absoluta puede tomar los valores 'A' (ausente), 'M' (marginal) o 'P' (presente) y denota la calidad de una sola hibridación. Utilizamos la información cuantitativa dada por la diferencia promedio y la información cualitativa dada por la llamada absoluta para identificar los genes que se expresan diferencialmente en muestras biológicas de individuos con cáncer versus muestras biológicas de la población normal. Con otros algoritmos distintos al de Affymetrix, se ha obtenido diferentes valores numéricos que representan los mismos valores de expresión y diferencias de expresión en comparación.

45 La expresión diferencial E en uno de los grupos de cáncer en comparación con la población normal se calcula de la siguiente manera. Dados n valores de diferencia promedio d1, d2, ..., dn en la población de cáncer y m valores de diferencia promedio c1, c2, ..., cm en la población de individuos normales, se calcula mediante la ecuación:

$$E \equiv \exp\left(\frac{1}{m} \sum_{i=1}^m \ln(c_i) - \frac{1}{n} \sum_{i=1}^n \ln(d_i)\right) \quad (\text{ecuación 1})$$

Si $d_j < 50$ o $c_i < 50$ para uno o más valores de i y j , estos valores particulares c_i y/o d_j se configuran en un valor de expresión "artificial" de 50. Este cálculo particular de E permite una comparación correcta con los resultados de TaqMan.

5 Un gen se llama sobrerregulado en el cáncer de resultado bueno o malo, si $E \geq$ factor de cambio promedio 2 y si el número de llamadas absolutas igual a 'P' en la población de cáncer es mayor que $n/2$.

10 Las figuras 14 y 14A muestran la expresión relativa de la fase aguda y los marcadores inmunes en muestras de tumores frescos de pacientes como se describe en el Ejemplo 1 y según lo determinado por el análisis Affymetrix GeneChip. La respuesta de las lesiones metastásicas según lo determinado por computotomografía se representa como "PR" = Respuesta parcial, "SD" = Enfermedad estable y "PD" = Enfermedad progresiva. Se presentan los niveles de expresión de los tejidos normales adyacentes (Muc = Mucosa; Liv = hígado). Los niveles absolutos de expresión normalizados por el escalamiento global de cada gen indicado se representan en líneas. Los pacientes se representan en filas, comenzando con el número de paciente seguido del tipo de tumor (tumor primario "PR" o lesión metastásica "LM"). El código de color se representa en la parte superior izquierda para visualizar la respuesta del tumor.

15 Como se representa en las figuras 14 y 14A, la expresión de la fase aguda y los marcadores inmunitarios se correlacionan con la respuesta clínica de la metástasis hepática de los pacientes con CCR que se tratan con un régimen basado en 5'FU según lo determinado por las determinaciones de TC de las lesiones metastásicas. El tipo de muestra se indica de la siguiente manera: Muc1-3 = tejido de mucosa normal 1-3, LIV1 = tejido de hígado normal, LM = metástasis hepática, PR = tumor primario. La respuesta clínica se denota como sigue: "Respuesta parcial" (= PR o barra de color verde en la parte superior), "Enfermedad estable" (= SD o barra de color naranja en la parte superior) y "Enfermedad progresiva" (= PD o barra de color roja en la parte superior). La expresión de la fase aguda y los marcadores inmunes se observan únicamente en la lesión metastásica y no en el tejido tumoral primario. La expresión es específicamente elevada en las lesiones metastásicas que no responden al régimen anticanceroso.

25 Las figuras 15 y 15A muestran la expresión relativa de los genes candidatos que son en sí mismos fase aguda y marcadores inmunes o se corregulan en muestras de tumores frescos de pacientes como se describe en el Ejemplo 1 y según lo determinado por el análisis Affymetrix GeneChip. La respuesta de las lesiones metastásicas según lo determinado por computotomografía se representa como "PR" = Respuesta parcial, "SD" = Enfermedad estable y "PD" = Enfermedad progresiva. Se presentan los niveles de expresión de los tejidos normales adyacentes (Muc = Mucosa; Liv = hígado). Los niveles absolutos de expresión normalizados por el escalamiento global de cada gen indicado se representan en líneas. Los pacientes se representan en filas, comenzando con el número de paciente seguido del tipo de tumor (tumor primario "PR" o lesión metastásica "LM"). El código de color se representa en la parte superior izquierda para visualizar la respuesta del tumor.

30 Como se representa en las figuras 15 y 15A, la expresión de marcadores de fase aguda y genes corregulados se correlacionan con la respuesta clínica de la metástasis hepática de pacientes con CCR que se tratan con un régimen basado en 5'FU según lo determinado por las determinaciones de TC de las lesiones metastásicas.

35 La Tabla 12 enumera secuencias de nucleótidos representativas de la fase aguda y marcadores inmunes que pueden expresarse para producir marcadores que son útiles en los métodos de la invención.

Tabla 12

Fase aguda y conjunto de marcadores inmunes y genes corregulados de ejemplo				
Símbolo genético	Ref. Descripción de secuencias	Ref. secuencias	Unigene_ID	OMIM
APOB	precursor de apolipoproteína B	NM_000384	Hs.585	107730
APOC1	precursor de apolipoproteína C-I	NM_001645	Hs.268571	107710
APOE	apolipoproteína E complemento componente 1, q	NM_000041	Hs.169401	107741
C1QA	subcomponente, precursor de polipéptido alfa	NM_015991	Hs.9641	120550
C1QB	componente del complemento 1, subcomponente q, precursor de polipéptido beta	NM_000491	Hs.8986	120570
C3	componente complementario 3 precursor	NM_000064	Hs.284394	120700

ES 2 717 627 T3

Fase aguda y conjunto de marcadores inmunes y genes corregulados de ejemplo				
Símbolo genético	Ref. Descripción de secuencias	Ref. secuencias	Unigene_ID	OMIM
C4A	Complemento componente 4A preproteína	NM_007293	Hs.278625	120810
CRP	Proteína C reactiva, relacionada con pentraxina	NM_000567	Hs.76452	123260
F2	factor de coagulación II precursor	NM_000506	Hs.76530	176930
F5	factor de coagulación V precursor	NM_000130	Hs.30054	227400
FGA	fibrinógeno, cadena alfa isoforma alfa-E preproteína	NM_000508	Hs.90765	134820
FGB	fibrinógeno, preproteína de la cadena beta	NM_005141	Hs.7645	134830
FGG	fibrinógeno, precursor cadena gamma isoforma gamma-A	NM_000509	Hs.75431	134850
ITIH3	Inhibidor de la pre-alfa (globulina), polipéptido H3	NM_002217	Hs.76716	146650
ITIH4	Inhibidor inter-alfa (globulina) H4 (glucoproteína sensible a la calicreína plasmática)	NM_002218	Hs.76415	600564
ORM1	precursor orosomucoide 1	NM_000607	Hs.572	138600
ORM2	orosomucoide 2	NM_000608	Hs.278388	138610
SAA2	suero amiloide A1	NM_000331	Hs.18162	104750
TF	transferrina	NM_001063	Hs.284176	190000
APCS	precursor del componente amiloide P del suero	NM_001639	Hs.1957	104770
ARL7	factor de ribosilación ADP tipo-7	NM_005737	Hs.111554	604787
BBOX1	gamma-butirotetaina hidroxilasa	NM_003986	Hs.9667	603312
C4B	Complemento componente 4B preproteína	NM_000592	Hs.278625	120820
C4BPA	Complemento componente 4 proteína de unión, alfa	NM_000715	Hs.1012	120830
C8B	componente del complemento 8, polipéptido beta	NM_000066	Hs.38069	120960
CAST	isoforma de calpastatina a	NM_001750	Hs.279607	114090
CPB2	plasma carboxipeptidasa B2 isoforma a, preproteína	NM_001872	Hs.274495	603101
FBP17	proteína de unión a formina 1	NM_015033	Hs.301763	606191
FGL1	precursor 1 similar a fibrinógeno	NM_004467	Hs.107	605776

ES 2 717 627 T3

Fase aguda y conjunto de marcadores inmunes y genes corregulados de ejemplo				
Símbolo genético	Ref. Descripción de secuencias	Ref. secuencias	Unigene_ID	OMIM
FLJ11560	proteína hipotética FLJ11560	NM_025182	Hs.301696	-
FSTL3	glicoproteína 3 similar a la folistatina	NM_005860	Hs.25348	605343
GC	componente específico del grupo (proteína de unión a la vitamina D)	NM_000583	Hs.198246	139200
HXB	tenascina C(hexabrachion)	NM_002160	Hs.289114	187380
IGFBP1	proteína de unión al factor de crecimiento similar a la insulina 1	NM_000596	Hs.102122	146730
ITIH2	Inhibidor inter-alfa (globulina), polipéptido H2	NM_002216	Hs.75285	146640
KMO	quinurenina 3-monooxigenasa (quinurenina 3-hidroxilasa)	NM_003679	Hs.107318	603538
MAGP2	glicoproteína asociada a microfibrillas 2	NM_003480	Hs.512842	601103
MGC4638	inhibina beta E	NM_031479	Hs.279497	-
NNMT	nicotinamida N-metiltransferasa	NM_006169	Hs.76669	600008
PBX3	Factor de transcripción de leucemia de células pre-B 3	NM_006195	Hs.294101	176312
PCDH17	protocaderina 17	NM_014459	Hs.106511	-
PLOD	procolágeno-lisina 5-dioxigenasa	NM_000302	Hs.75093	153454
PPP3R1	proteína fosfatasa 3, subunidad reguladora B, isoforma alfa 1	NM_000945	Hs.278540	601302
PRKCDBP	proteína quinasa C, proteína de unión delta	NM_145040	Hs.85181	-
SERPINA1	inhibidor de la serina (o cisteína) proteinasa, clado A (alfa-1 antiproteinasa, antitripsina), miembro 1	NM_000295	Hs.297681	107400
SERPINA1	inhibidor del activador del plasminógeno-1	NM_000602	Hs.82085	173360
SERPINA1	Componente 1 complemento inhibidor precursor	NM_000062	Hs.151242	606860
TEGT	Transcripto de gen de testículo mejorado (inhibidor de BAX 1)	NM_003217	Hs.74637	600748
TUBB	tubulina, beta polipéptido UDP	NM_001069	Hs.179661	191130
UGT2B4	Familia de la glicosiltransferasa 2, polipéptido B4	NM_021139	Hs.89691	600067

ES 2 717 627 T3

La Tabla 13 muestra los niveles de expresión de la fase aguda y los marcadores inmunitarios que discriminan entre los tumores que responden y los que no responden, según lo determinado por el perfil de expresión génica utilizando Affymetrix GeneChip HG U133A.

5 Datos de expresión de genes candidatos que comparan a pacientes que respondieron (Resp) versus no respondedores (No-Resp) que fueron tratados con quimioterapia paliativa basada en 5-FU

10 Los factores de cambio de plegamiento promedio se representan para aquellos pacientes que responden a un tumor (grupo de muestra 1, metástasis hepática de respuesta), o que no responden a un régimen basado en 5-FU (grupo de muestra 2, metástasis hepática que no responde). La intensidad de la señal promedio dentro de cada subgrupo, la relación de cambio de plegamiento ("Fc") entre los dos subgrupos, la significación estadística según la prueba t de Student y la dirección de cambio se indican para cada gen especificado por nombre y abreviatura.

Tabla 13

	Nr Affy	Promedio respondedor	Promedio no respondedor	Fc_ Resp vs No Resp	Prueba T	Dirección Resp vs No Resp	Nombre del gen	Gen
1	202953_at	305.67	1092.85	-3.58	0.033	Abajo	componente del complemento 1, subcomponente q, polipéptido beta	C1QB
2	203382_s_at	132.48	513.97	-3.88	0.001	Abajo	apolipoproteína E	APOE
3	204416_x_at	856.1	3347.48	-3.91	0.002	Abajo	apolipoproteína C-I	APOC1
4	204714_s_at	231.23	1197.85	-5.18	0.005	Abajo	Factor de coagulación V (proacelerina, factor lábil)	F5
5	204988_at	2708	13973.7 2	-5.16	0.031	Abajo	fibrinógeno, polipéptido B beta	FGB
6	205041_s_at	196.2	2513.82	- 12.81	0.021	Abajo	orosomucoide 1	ORM1
7	205108_s_at	209.98	845.3	-4.03	0.025	Abajo	apolipoproteína B (incluido el antígeno Ag (x))	APOB
8	205650_s_at	493.55	2904.28	-5.88	0.041	Abajo	fibrinógeno, polipéptido alfa un	FGA
9	205754_at	209.75	662.38	-3.16	0.026	Abajo	factor de coagulación II (trombina)	F2
10	214063_s_at	237.9	1677.93	-7.05	0.046	Abajo	transferrina	TF
11	214428_x_at	779.75	2975.48	-3.82	0.005	Abajo	complemento componente 4A	C4A
12	214456_x_at	264.88	4909.38	- 18.53	0.038	Abajo	suero amiloide A2	SAA2
13	214465_at	66.25	762.43	- 11.51	0.026	Abajo	orosomucoide 2	ORM2

	Nr Affy	Promedio respondedor	Promedio no respondedor	Fc_Resp vs No Resp	Prueba T	Dirección Resp vs No Resp	Nombre del gen	Gen
14	217767_at	953.25	5588.37	- 5.86	0.038	Abajo	complemento componente 3	C3
15	218232_at	93.93	390.22	- 4.15	0.005	Abajo	componente del complemento 1, subcomponente q, polipéptido alfa	C1QA
16	219612_s_at	1133.45	7741.42	- 6.83	0.038	Abajo	fibrinógeno, polipéptido gamma	FGG
17	37020_at	462.3	3024.4	- 6.54	0.029	Abajo	Proteína C reactiva, relacionada con pentraxina	CRP

Los cambios en el plegamiento mayores que 1 se refieren a una diferencia en la expresión génica entre la primera y la segunda muestra de cohorte. Estos factores de regulación son valores medios y pueden diferir individualmente, aquí los perfiles combinados de 17 genes enumerados en la Tabla 12 en un análisis de agrupamiento o un análisis de componentes principales (PCA) indicarán el grupo de clasificación para dicha muestra.

Filtración de datos:

Los datos sin procesar del análisis de la matriz de genes se adquirieron utilizando el software Microsuite 5.0 de Affymetrix y se normalizaron siguiendo una práctica estándar de escalar el promedio de todas las intensidades de la señal del gen a un valor arbitrario común. Se eliminaron del análisis 59 genes correspondientes a los controles de Affymetrix (genes de mantenimiento, etc.). La única excepción se ha hecho para los genes para GAPDH y Beta-actina, cuyos niveles de expresión se usaron para los propósitos de normalización. Cien genes, cuyos niveles de expresión se usan de forma rutinaria para normalizarse entre HG-U133A y HG-U133B GeneChips, también se eliminaron del análisis. Los genes con niveles de ruido potencialmente altos (81 conjuntos de sondas), que se observan para los genes con valores de expresión absolutos bajos (genes, cuyos niveles de expresión no alcanzaron 30 RLU (TGT=100) en todos los experimentos), se eliminaron del conjunto de datos. Los genes restantes fueron preprocesados para eliminar los genes (3196 conjuntos de sondas) cuyas intensidades de señal no fueron significativamente diferentes de sus niveles de fondo y, por lo tanto, fueron etiquetados como "Ausentes" por Affymetrix MicroSuite 5.0 en todos los experimentos. Eliminamos los genes que no estaban presentes en al menos el 10% de las muestras (3841 series de sondas). Los datos para los 15,006 conjuntos de sondas restantes se analizaron posteriormente mediante métodos estadísticos.

Análisis estadístico:

Para optimizar la predicción del resultado, se puede utilizar esta clase de la cohorte de entrenamiento y realizar múltiples pruebas estadísticas, adecuadas para la comparación de grupos, incluida la prueba no paramétrica de la suma de rangos de Wilcoxon, la prueba t de Student independiente de dos muestras, la prueba de Welch, la prueba de Kolmogorov-Smirnov (para varianza), y la prueba SUM-Rank. Como se muestra, podemos identificar dichos genes con una expresión diferencial en el grupo que responde frente a no respondedor y un nivel de significación (valor p) inferior a 0.05. Por este medio verificamos la significación estadística de los genes candidatos seleccionados que se muestran en la Tabla 12.

Adicionalmente, se puede aplicar la corrección para múltiples errores de prueba, como Benjamini-Hochberg, y puede aplicar pruebas para la detección de falso descubrimiento, como las permutaciones con los algoritmos Bootstrap o Jack-knife.

Ejemplo 5

Análisis en suero de la PCR en muestras de suero en serie de pacientes con tumor que padecen cáncer colorrectal metastásico antes y durante la quimioterapia basada en 5'FU.

Resumen

Las muestras de suero en serie obtenidas de cada paciente como se describe en el Ejemplo 1 se analizaron para determinar los niveles de proteína en fase aguda (es decir, CRP) mediante el uso de la prueba de amplio rango disponible comercialmente para la CRP (#74038) de Bayer Diagnostics en la plataforma ADVIA 2400, según instrucciones del fabricante y en comparación con el tamaño determinado clínicamente de la lesión tumoral metastásica.

5

Como se puede ver en las figuras 16A y 16B, las mediciones en serie de muestras de suero de varios pacientes revelaron un aumento en los niveles séricos de CRP (columnas rojas [mg/l]) en pacientes que sufrieron progresión de la enfermedad metastásica posteriormente como lo muestran los cambios en el tamaño del tumor (columnas grises [cm²]). Las muestras de pretratamiento se representan como "A". Posteriormente se obtuvieron muestras de suero antes de cada ciclo de quimioterapia. Como se puede ver en el paciente G73, el aumento de la PCR de 14.7 mg/l en el punto de tiempo "E" a 47.5 mg/l en el punto de tiempo "F" precede a la progresión masiva de la lesión hepática metastásica un mes más tarde en el punto temporal "G" de 6.3 a 18 cm². De manera similar para el paciente 179, la elevación de la PCR de 0.4 mg/ml en el punto de tiempo "C" a 4.3 mg/ml en el punto de tiempo "D" precede al crecimiento del tumor en el punto de tiempo "E" de 7.5 cm² a 22.1 cm². Por lo tanto, se ha descubierto que el aumento de los procesos inflamatorios es una reacción muy temprana a la recurrencia/progresión del tumor antes de que pueda determinarse mediante las posibilidades de evaluación del estándar clínico de oro (es decir, la exploración por TC). Sin embargo, la identificación temprana de la progresión tumoral se puede usar para modificar los programas de tratamiento aplicados y, por lo tanto, se puede usar para monitorear la efectividad de la terapia y optimizar el régimen antitumoral para vencer los mecanismos de resistencia y finalmente ahorrar tiempo y potencialmente generar un beneficio de supervivencia.

10

15

20

REIVINDICACIONES

1. Un método implementado por ordenador para predecir un resultado clínico relacionado con un paciente que padece cáncer colorrectal, que comprende: (a) determinar un valor predictivo mediante el uso de valores de muestra de pacientes para (1) al menos un marcador seleccionado del grupo que consiste en marcadores tumorales, marcadores inmunes, marcadores de fase aguda, y (2) al menos un marcador que (i) se selecciona del grupo que consiste en un marcador de matriz extracelular (ECM), un marcador que es indicativo de la síntesis de matriz extracelular (fibrogénesis), y marcador que es indicativo de la degradación de la matriz extracelular (fibrólisis), y (b) predice el resultado clínico del cáncer colorrectal mediante la evaluación del valor predictivo en donde el marcador de (1) y el marcador de (2) no son los mismos, en donde el valor del predictor se determina (1) mediante un algoritmo derivado del análisis de regresión de Cox y (2) se utiliza para evaluar la probabilidad de que el paciente responda favorablemente a un régimen de tratamiento antineoplásico.
2. Un método de la reivindicación 1, en donde se usan valores para los siguientes marcadores: (1) valores para al menos un marcador tumoral, marcador inmune, o marcadores de fase aguda seleccionados del grupo que consiste en CEA, CA15-3, CA19-9, EGFR, ERBB2, ERBB3, ERBB4, c-Kit, KDR, FLT4, FLT3, c-Met, un miembro de la superfamilia FGFR, miembro de la familia de ligandos FGFR y variantes de empalme relacionadas, un miembro de la familia del factor de crecimiento, miembros de la superfamilia VEGFR (KDR, FLT3, FLT4), miembros de la familia de ligandos VEGFR (VEGFA, VEGFB, VEGFC, VEGFD), dominios compartidos de miembros de factores de crecimiento, interleucinas, receptores de interleucina, factores de complemento, proteínas de fase aguda y hormonas; y (2) al menos un marcador que es: (i) un marcador de matriz extracelular (ECM) seleccionado del grupo que consiste en colágenos, proteínas de adhesión basal, entactina, proteoglicanos y glicosaminoglicanos P, un miembro de la superfamilia de colágeno; o (ii) un marcador que sea indicativo de la síntesis de matriz extracelular (fibrogénesis) seleccionada del grupo que consiste en preformas de colágenos, proteínas de adhesión basal, entactina, proteoglicanos y glicosaminoglicanos o prepropéptidos de los mismos; o (iii) un marcador que sea indicativo de la degradación de la matriz extracelular (fibrólisis) seleccionado del grupo que consiste en la superfamilia MMP, o inhibidores asociados de la misma, incluida la superfamilia TIMP.
3. Un método de la reivindicación 1, en donde el paciente padece cáncer colorrectal y en donde se usan valores para uno o más de los siguientes marcadores: PI1INP, Colágeno IV, Colágeno VI, Tenascina, Laminina, HA, CRP, MMP-2, TIMP-I, complejo MMP-9/TEVIP-I, CEA, CA15-3, CA19-9, IL2R, IL6, Gastrina, Her-2/neu, EGFR, uPA y VEGF165.
4. Un método de la reivindicación 1, que comprende además comparar un valor de predictor con un conjunto de datos comparativos que comprende uno o más valores numéricos, o rango de valores numéricos, que están asociados con un cáncer colorrectal.
5. Un método de la reivindicación 1, en donde el valor del predictor se determina utilizando un algoritmo de función lineal o no lineal que se deriva de: (a) compilar un conjunto de datos que comprende datos de marcadores relacionados con la enfermedad neoplásica para un primer grupo de sujetos, en donde los datos de marcadores relacionados con enfermedades neoplásicas se relacionan con (1) al menos un marcador tumoral o al menos un marcador inmunitario, y (2) al menos un marcador que es (i) un marcador de matriz extracelular (ECM) (ii) un marcador que es indicativo de la síntesis de la matriz extracelular (fibrogénesis), o (iii) un marcador que es indicativo de la degradación de la matriz extracelular (fibrólisis); (b) derivar un algoritmo de función lineal o no lineal del conjunto de datos compilados a través de la aplicación de al menos una metodología analítica seleccionada del grupo que consiste en análisis de función discriminante, análisis de regresión no paramétrica, árboles de clasificación y redes neuronales; (c) calcular los valores del predictor de validación para un segundo grupo de sujetos ingresando datos que comprenden datos de marcadores relacionados con enfermedades neoplásicas para el segundo grupo de sujetos en el algoritmo derivado en el paso (b); (d) comparar los valores del predictor de validación calculados en el paso (c) con las puntuaciones relacionadas con la enfermedad neoplásica para el segundo grupo de sujetos; y (e) si los valores del predictor de validación determinados en el paso (c) no se correlacionan dentro de un nivel de tolerancia clínicamente aceptable con los valores del predictor de validación para el segundo grupo de sujetos, realizando las siguientes operaciones (i)-(iii) hasta que tal tolerancia se cumpla: (i) modificar el algoritmo sobre una base o bases que comprende (1) revisar el conjunto de datos para el primer grupo de sujetos, y (2) revisar o cambiar la metodología analítica (ii) calcular valores de predictor de validación para el segundo grupo de sujetos ingresando datos que comprenden datos de marcadores relacionados con neoplasma para el segundo grupo de sujetos en el algoritmo modificado (iii) que evalúa si los valores de la puntuación de la biopsia de validación calculados utilizando el algoritmo modificado se correlacionan con los valores del predictor para el segundo grupo de sujetos dentro de lo clínicamente aceptable nivel de tolerancia.
6. Un medio legible por ordenador que tiene almacenado en él una estructura de datos que comprende un campo de datos que contiene datos que representan un algoritmo que genera valores predictivos que pueden usarse para predecir la respuesta de un paciente a un régimen de tratamiento antineoplásico, en donde el algoritmo usa valores para (1) al menos un marcador tumoral o al menos un marcador inmunitario, y (2) al menos un marcador que es (i) un marcador de matriz extracelular (ECM) (ii) un marcador que es indicativo de la síntesis de matriz extracelular (fibrogénesis), o (iii) un marcador que es indicativo de la degradación de la matriz extracelular (fibrólisis).

7. Una estructura de datos almacenada en un medio legible por ordenador que puede ser leído por un microprocesador y que comprende al menos un código que identifica de manera única los valores de predicción determinados de acuerdo con la reivindicación 1.
- 5 8. Una estructura de datos que comprende un campo de datos que contiene datos que representan un algoritmo que genera valores predictivos que pueden usarse para predecir la respuesta de un paciente a un régimen de tratamiento antineoplásico, almacenándose dicha estructura de datos en un medio legible por ordenador que puede leerse por un microprocesador y que comprende al menos un código que identifica de manera única los datos que representan valores para (1) al menos un marcador tumoral o al menos un marcador inmunitario, y (2) al menos un marcador que es (i) un marcador de matriz extracelular (ECM) (ii) un marcador que es indicativo de la síntesis de la matriz extracelular (fibrogénesis), o (iii) un marcador que es indicativo de la degradación de la matriz extracelular (fibrólisis).
- 10 9. Un kit que comprende: (a) una estructura de datos que comprende un campo de datos que contiene datos que representan un algoritmo que genera valores predictivos que se pueden usar para predecir la respuesta de un paciente a un régimen de tratamiento antineoplásico, almacenándose dicha estructura de datos en un medio legible por ordenador que puede ser leído por un microprocesador y que comprende al menos un código que identifica de manera única los valores de predicción determinados de acuerdo con la reivindicación 1; y (b) uno o más inmunoensayos que comprenden reactivos inmunorreactivos, en donde los inmunoensayos detectan y determinan valores para (1) al menos un marcador tumoral o al menos un marcador inmune, y (2) al menos un marcador que es (i) un marcador extracelular de matriz (ECM) (ii) un marcador que es indicativo de la síntesis de matriz extracelular (fibrogénesis), o (iii) un marcador que es indicativo de la degradación de la matriz extracelular (fibrólisis).
- 15 10. Un kit que comprende uno o más inmunoensayos que detectan y determinan valores para (1) al menos un marcador tumoral o al menos un marcador inmunitario, y (2) al menos un marcador que es (i) un marcador de matriz extracelular (ECM) (ii) un marcador que es indicativo de la síntesis de la matriz extracelular (fibrogénesis), o (iii) un marcador que es indicativo de la degradación de la matriz extracelular (fibrólisis).
- 20 11. Un sistema que comprende: (a) una estructura de datos almacenada en un medio legible por ordenador que puede ser leída por un microprocesador y que comprende al menos un código que identifica de manera única los valores del predictor determinados de acuerdo con la reivindicación 1; y (b) uno o más inmunoensayos que detectan y determinan valores para (1) al menos un marcador tumoral o al menos un marcador inmunitario, y (2) al menos un marcador que es (i) un marcador de matriz extracelular (ECM) (ii) un marcador que sea indicativo de la síntesis de la matriz extracelular (fibrogénesis), o (iii) un marcador que sea indicativo de la degradación de la matriz extracelular (fibrólisis).
- 25 12. Un sistema que comprende: (a) una estructura de datos que comprende un campo de datos que contiene datos que representan un algoritmo que genera valores predictivos que se pueden usar para predecir la respuesta de un paciente a un régimen de tratamiento antineoplásico, almacenándose dicha estructura de datos en un medio legible por ordenador que puede ser leído por un microprocesador y que comprende al menos un código que identifica de manera única los valores de (1) al menos un marcador tumoral o al menos un marcador inmunitario, y (2) al menos un marcador que es (i) un marcador de matriz extracelular (ECM) (ii) un marcador que es indicativo de la síntesis de matriz extracelular (fibrogénesis), o (iii) un marcador que es indicativo de la degradación de la matriz extracelular (fibrólisis); y (b) uno o más inmunoensayos que detectan y determinan valores para (1) al menos un marcador tumoral o al menos un marcador inmunitario, y (2) al menos un marcador que es (i) un marcador de matriz extracelular (ECM) (ii) un marcador que sea indicativo de la síntesis de la matriz extracelular (fibrogénesis), o (iii) un marcador que sea indicativo de la degradación de la matriz extracelular (fibrólisis).
- 30 35 40 13. Un método implementado por ordenador que comprende predecir el estado o progreso de un cáncer colorrectal en un paciente mediante la evaluación de dos o más valores predictivos determinados en uno o más puntos de tiempo por un método de la reivindicación 1.
- 45 14. Un método implementado por ordenador que comprende predecir el estado o el progreso de un cáncer colorrectal en un paciente mediante la evaluación de dos o más valores predictivos para (1) al menos un marcador tumoral o al menos un marcador inmunitario, y (2) al menos uno marcador que es (i) un marcador de matriz extracelular (ECM) (ii) un marcador que es indicativo de síntesis de matriz extracelular (fibrogénesis), o (iii) un marcador que es indicativo de degradación de matriz extracelular (fibrólisis).
- 50 55 15. Un método implementado por ordenador para evaluar el pronóstico de un paciente que padece o está en riesgo de desarrollar un cáncer colorrectal que comprende evaluar los valores de los predictores determinados en dos o más puntos temporales, en donde: (a) los valores de los predictores se determinan algorítmicamente utilizando valores de muestras de pacientes para (1) al menos un marcador seleccionado del grupo que consiste en marcadores tumorales, marcadores inmunes y marcadores de fase aguda, y (2) al menos un marcador que es (i) un marcador de matriz extracelular (ECM) (ii) un marcador que es indicativo de la síntesis de la matriz extracelular (fibrogénesis), o (iii) un marcador que es indicativo de la degradación de la matriz extracelular (fibrólisis); y (b) el pronóstico del paciente se evalúa mediante la evaluación de los valores predictivos determinados utilizando un algoritmo derivado del Análisis de Regresión de Cox.

FIG. 1

ANÁLISIS DE KAPLAN MEIER DE ALGORÍTMO

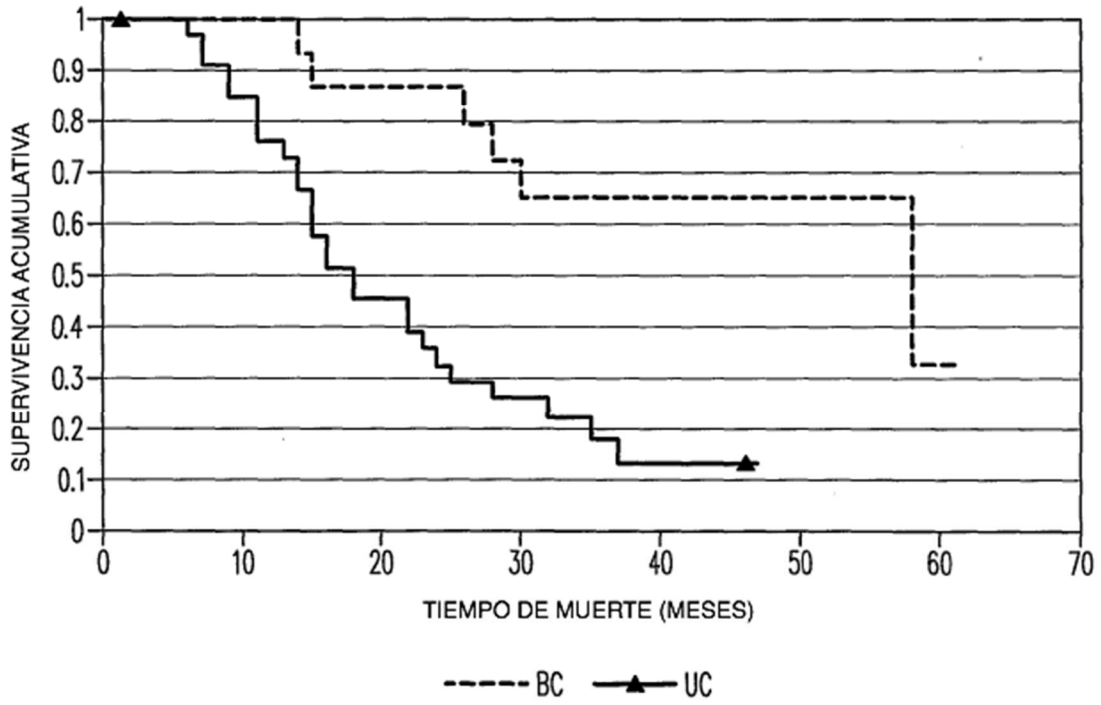


FIG. 2

ANÁLISIS DE KAPLAN MEIER DE GASTRINA

--- ALTA GASTRINA

—▲— BAJA GASTRINA

GASTRINA CO 25,4

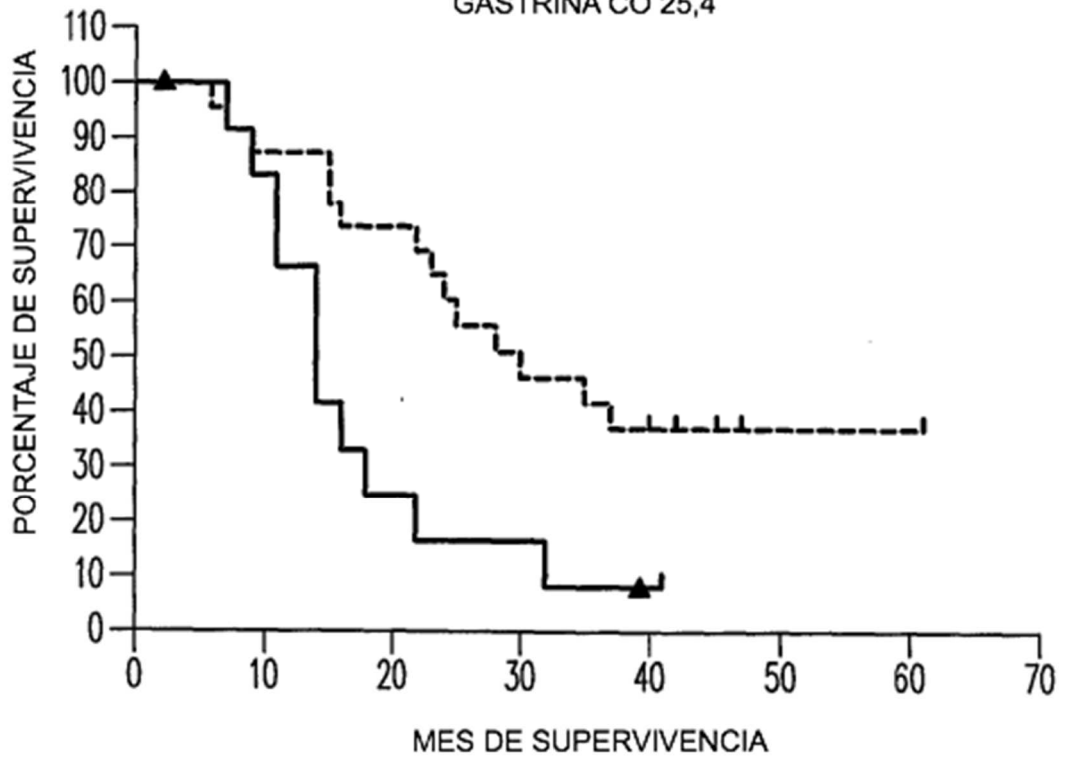


FIG. 3

ANÁLISIS DE KAPLAN MEIER DE CA 19-9

----- ALTO CA 19-9
—▲— BAJO CA 19-9

CA 19-9 CO 37

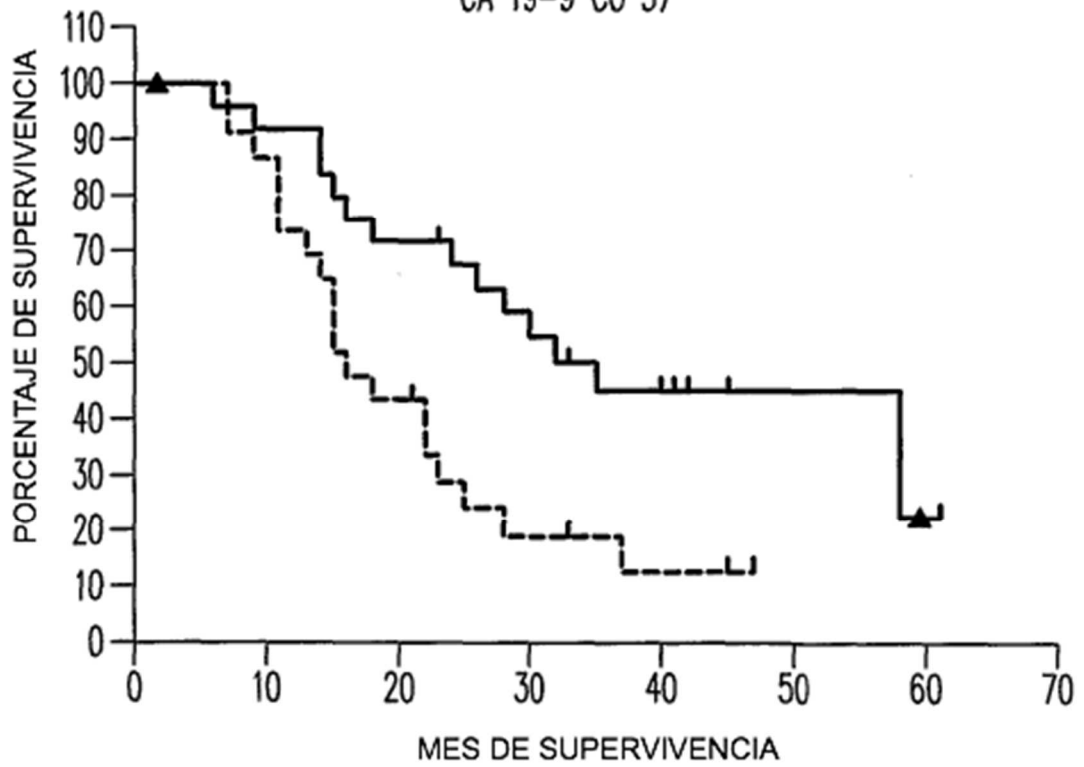


FIG. 4

ANÁLISIS DE KAPLAN MEIER DE TIMP-1

- ALTA INMUNO TIMP-1
- ▲ BAJA INMUNO TIMP-1

INMUNO TIMP-1 CO 1037,6

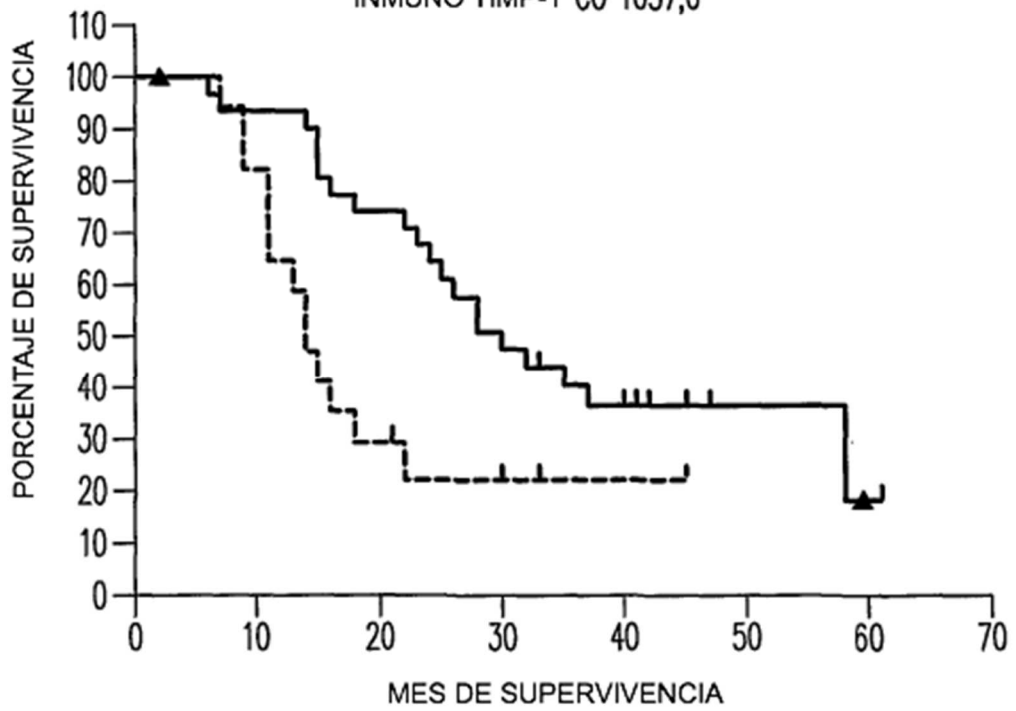


FIG. 5

ANÁLISIS DE KAPLAN MEIER DE MMP-2

--- MMP-2 > 675

—▲— MMP-2 < 675

MMP-2 CO 675

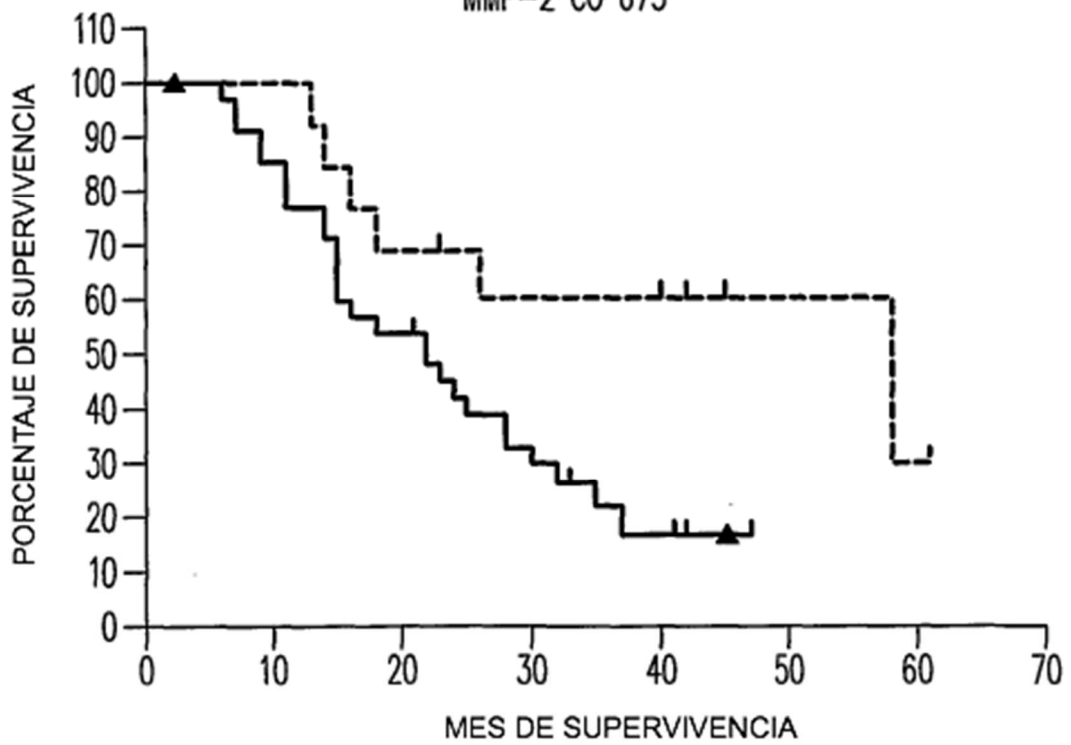


FIG. 6

ANÁLISIS DE KAPLAN MEIER DE EGFr

--- EGFr < 45

—▲— EGFr > 45

EGFr CO 45

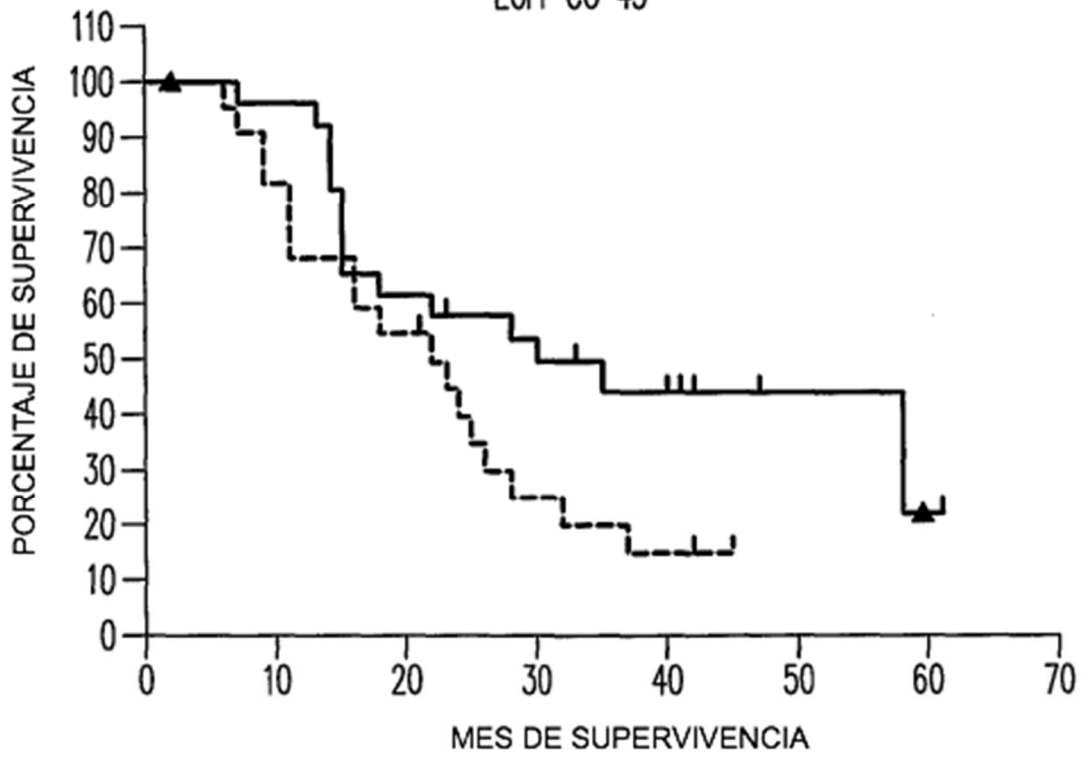


FIG. 7

ANÁLISIS DE KAPLAN MEIER DE VEGF

- ALTO VEGF
- ▲— BAJO VEGF

VEGF CO 221,1

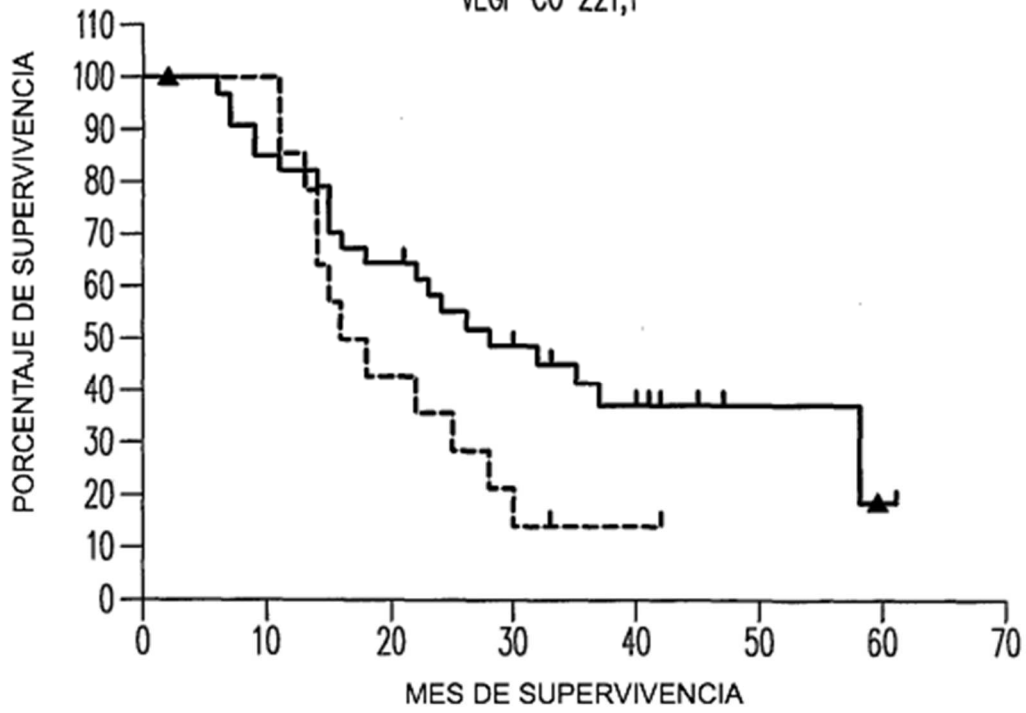


FIG. 8

ANÁLISIS DE KAPLAN MEIER DE CEA

- CEA > 100
- ▲ CEA < 100

CEA CO 100

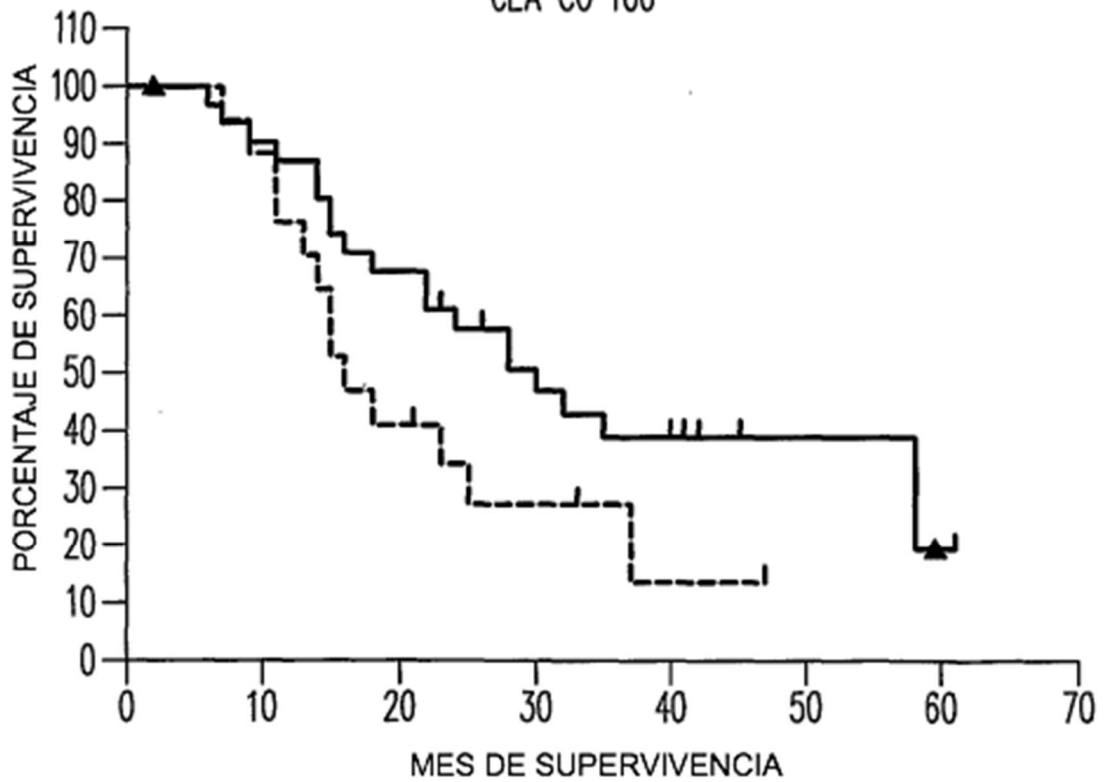


FIG. 9

ANÁLISIS DE KAPLAN MEIER DE VALORES DEL ALGORITMO MCT-V

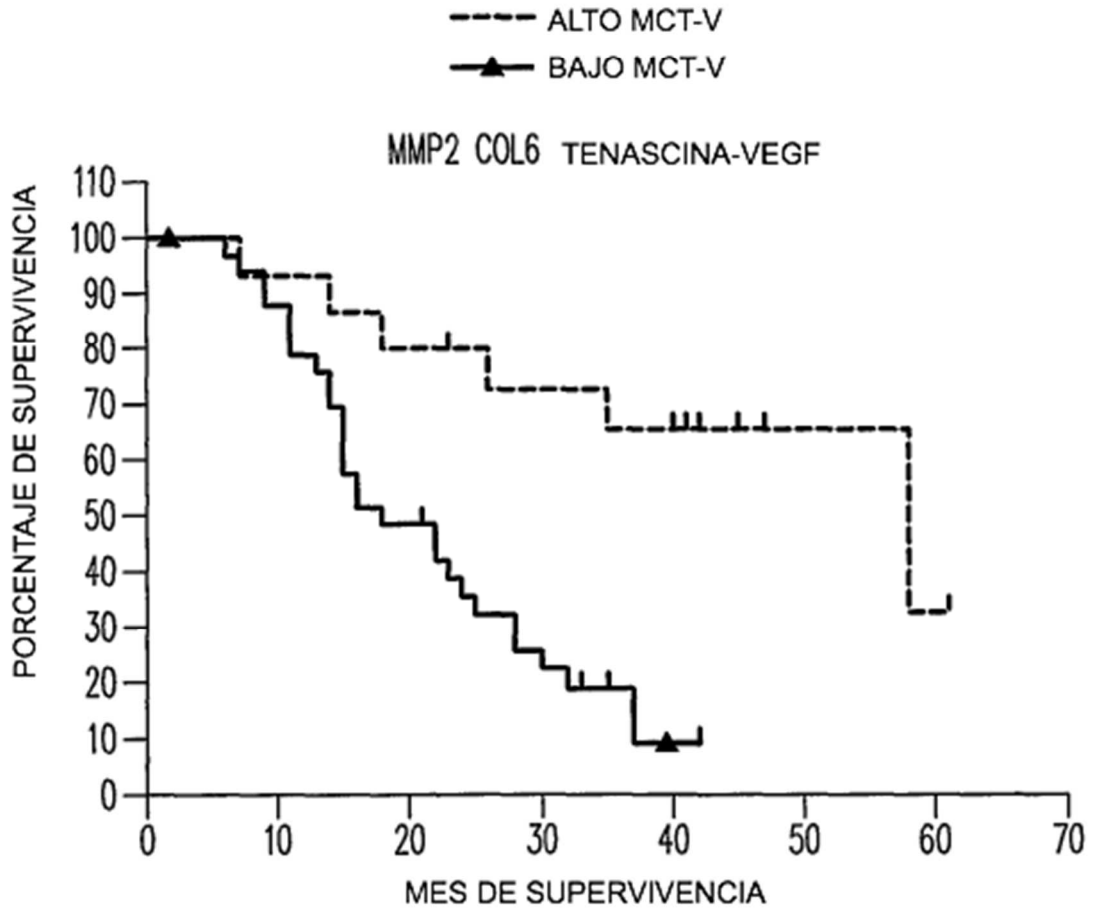
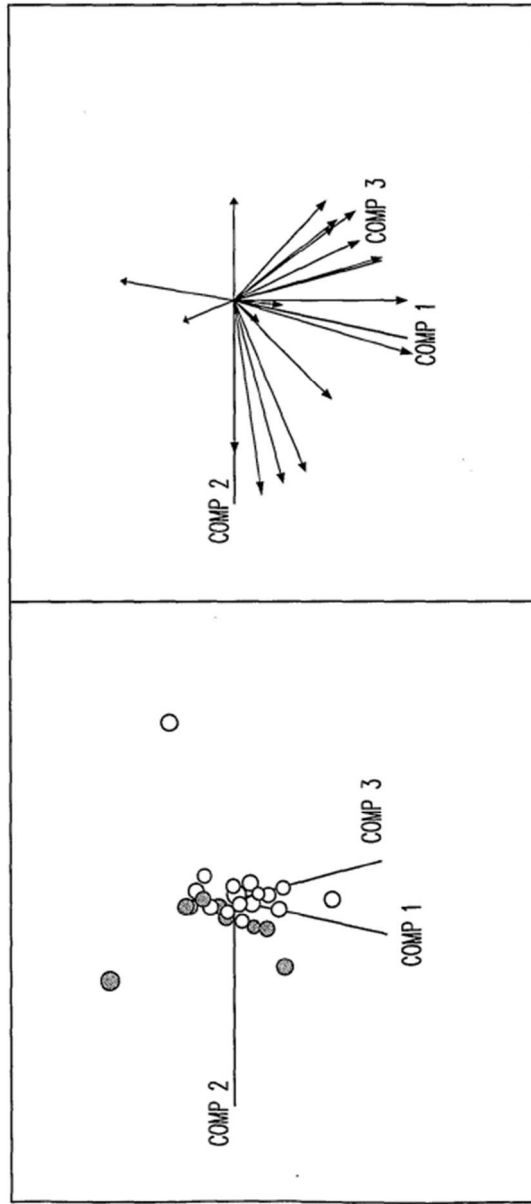


FIG. 10

ANÁLISIS DEL COMPONENTE PRINCIPAL MEDIANTE EL USO DE TODOS LOS PARÁMETROS



ANALISTA EXPRESIONISTA DE DATOS DE GEN

- GRUPOS DE EXPERIMENTO
- MUERTO < 19 meses
 - VIVO > 40 meses sin G18
- GRUPOS DE GEN
- TODOS LOS GENES

FIG. 10A

ANÁLISIS DEL COMPONENTE PRINCIPAL MEDIANTE EL USO DE TODOS LOS PARÁMETROS

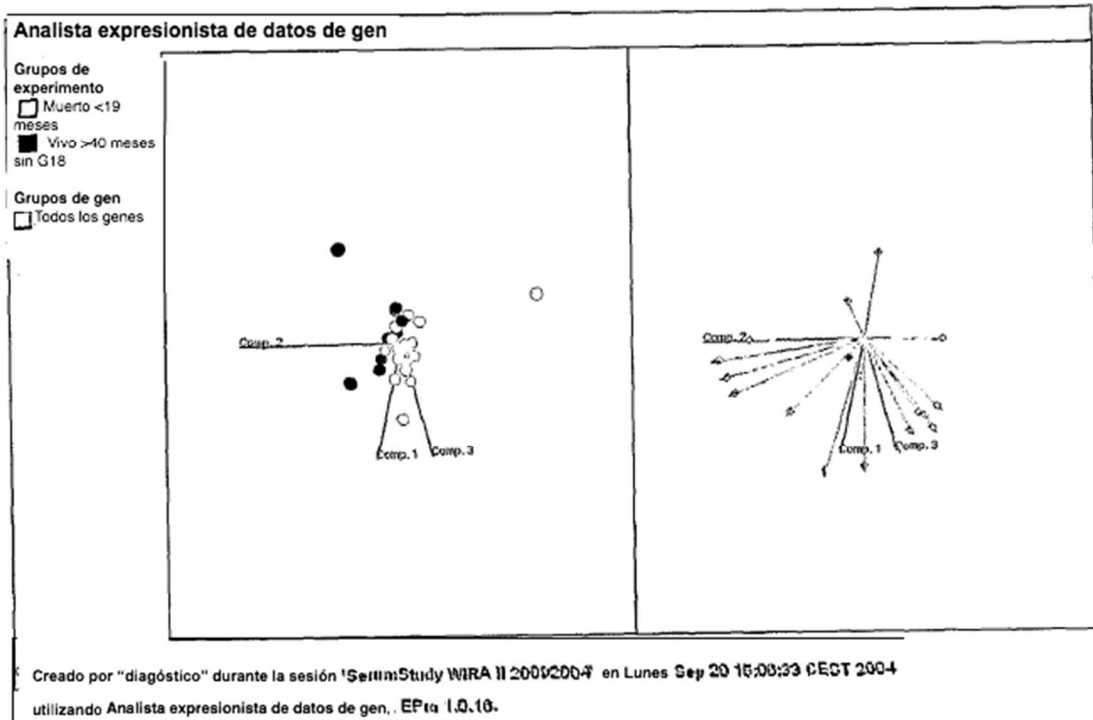
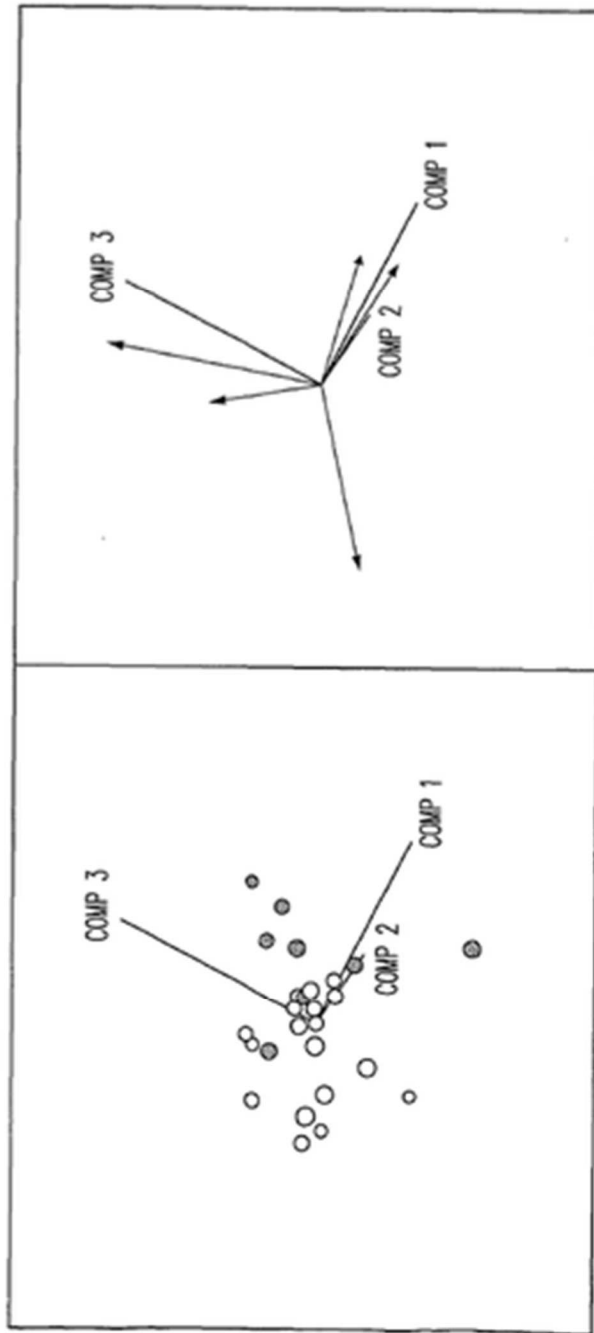


FIG. 11

ANÁLISIS DEL COMPONENTE PRINCIPAL MEDIANTE EL USO DE LOS PRIMEROS 5 PARÁMETROS



ANÁLISIS EXPRESIONISTA DE DATOS DE GEN

GRUPOS DE EXPERIMENTO

- MUERTO < 19 meses
- VIVO > 40 meses sin G18

GRUPOS DE GEN

- DE MUERTO < 19 vs VIVO > 40: PRUEBAS MÚLTIPLES: WELCH, KOLMOGOROV-SMIMOV, WILCOXER

FIG. 11A

ANÁLISIS DEL COMPONENTE PRINCIPAL MEDIANTE EL USO DE LOS PRIMEROS 5 PARÁMETROS

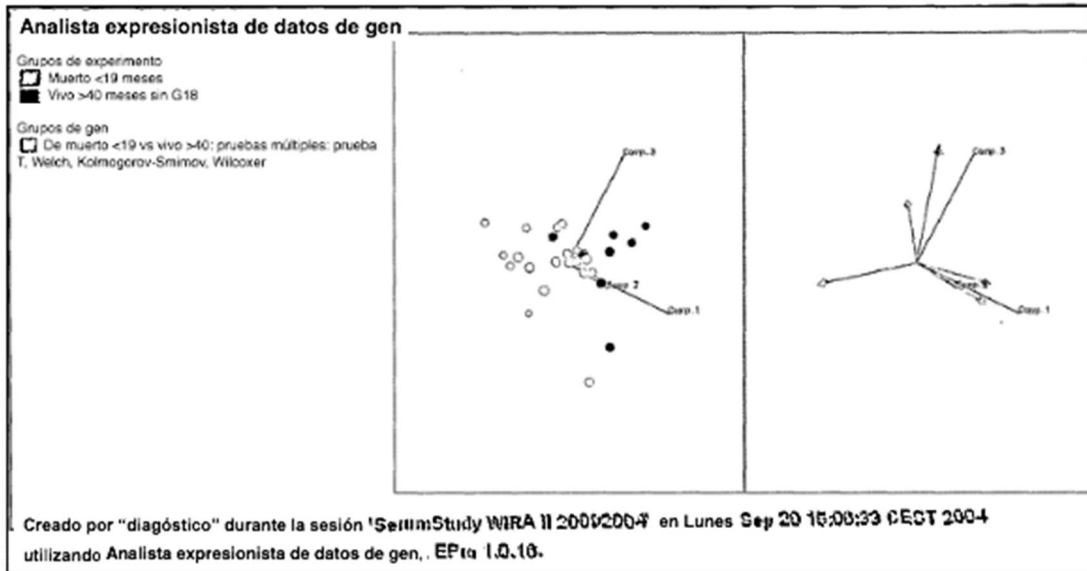


FIG. 12

ANÁLISIS DE EXPRESIÓN DE qRT DE LOS MIEMBROS DE LA FAMILIA ERB EN TEJIDO FFPE DE TUMORES PRIMARIOS

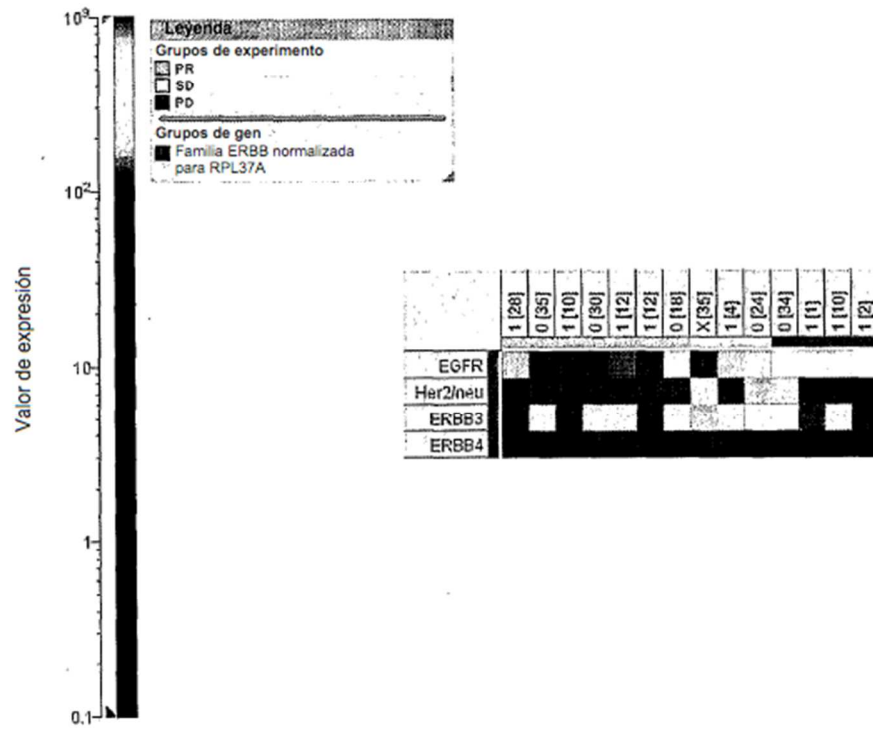


FIG. 12A

ANÁLISIS DE EXPRESIÓN DE qRT DE LOS MIEMBROS DE LA FAMILIA ERB EN TEJIDO FFPE DE TUMORES PRIMARIOS

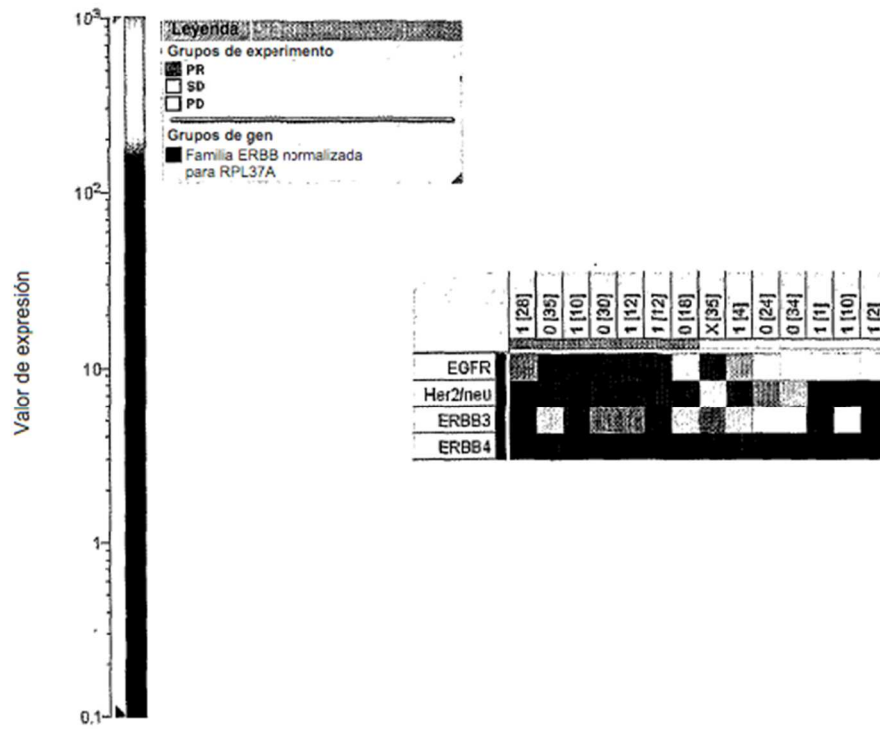


FIG. 13

ANÁLISIS DE KAPLAN MEIER DE TIMP-1 Y EGFr EN SUERO

- ALTO TIMP-1 Y BAJO EGFr
- ▲— BAJO TIMP-1 Y/O NORMAL EGFr

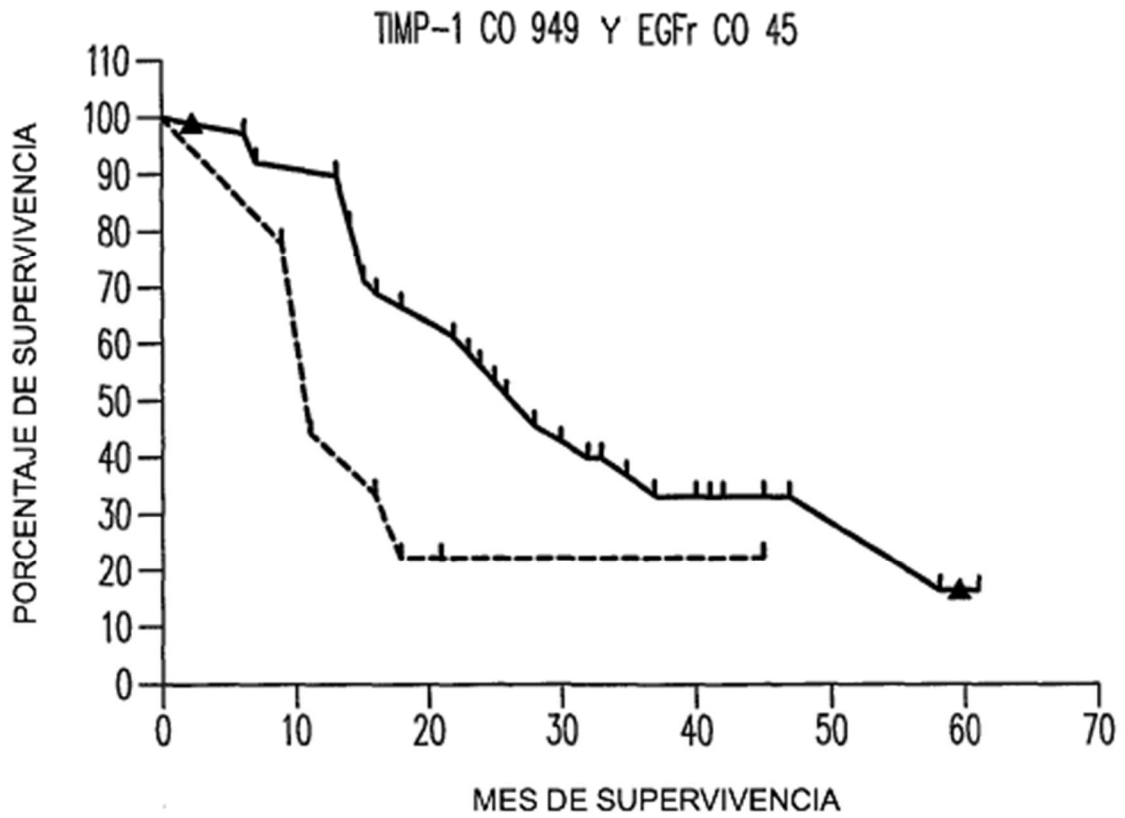


FIG. 14

PERFILACIÓN DE EXPRESIÓN DE MARCADORES PREDICTIVOS DEL SUERO
 PARECIDOS A LA REACCIÓN DEL HÍGADO A LA LESIÓN METÁSTICA

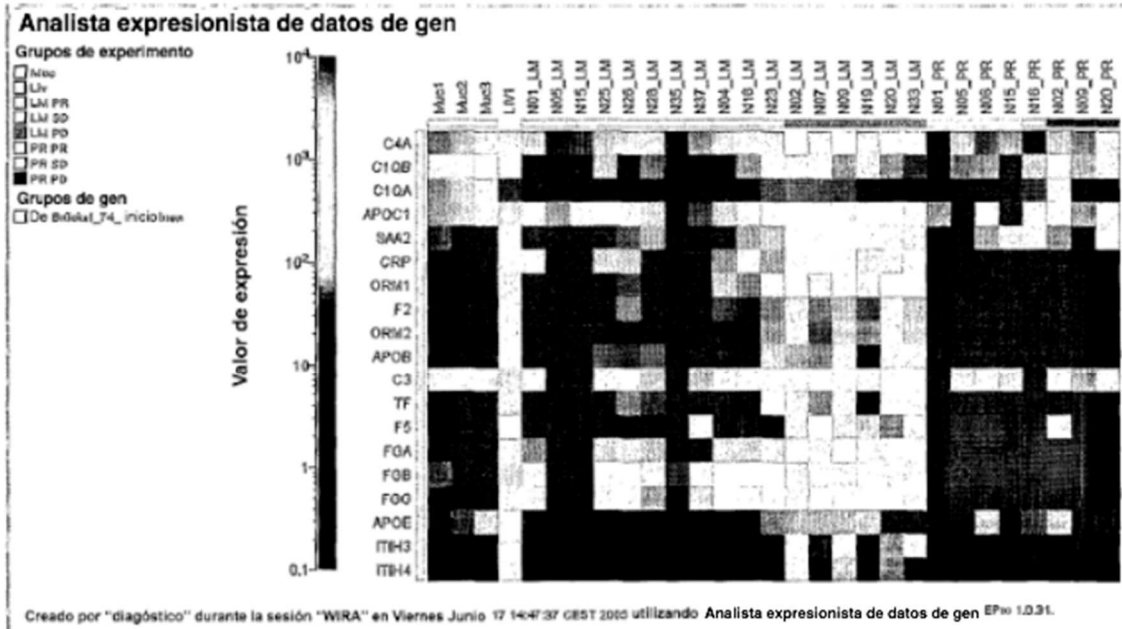


FIG. 14A

PERFILACIÓN DE EXPRESIÓN DE MARCADORES PREDICTIVOS DEL SUERO
PARECIDOS A LA REACCIÓN DEL HÍGADO A LA LESIÓN METÁSTICA

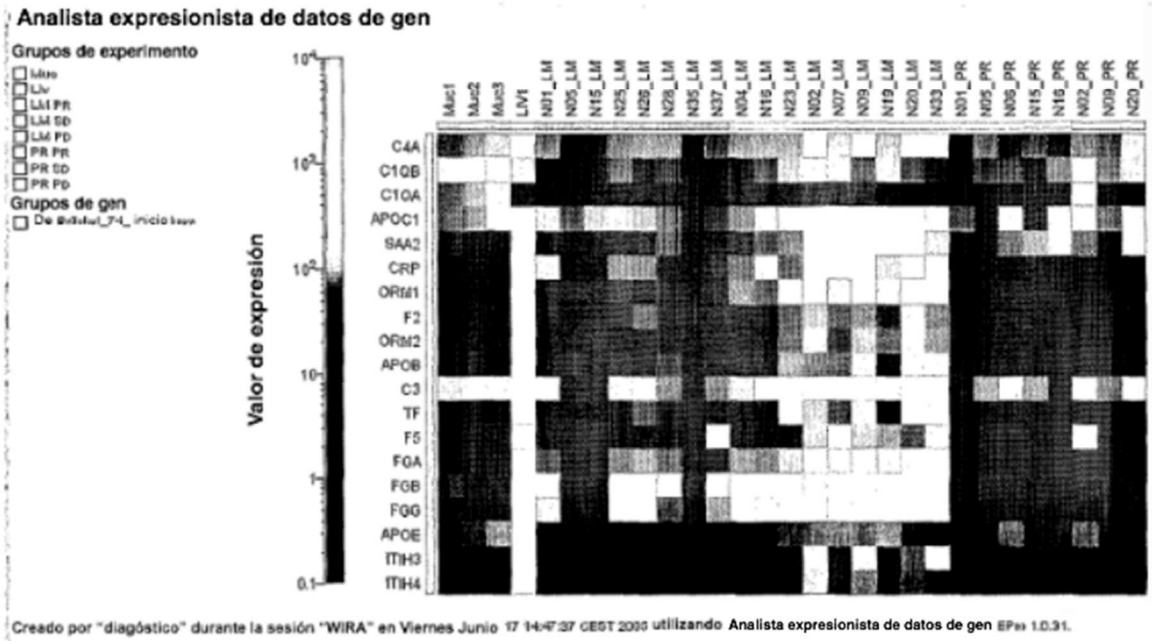


FIG. 15

PERFILACIÓN DE EXPRESIÓN DE MARCADORES PREDICTIVOS DEL SUERO PARECIDOS A LA REACCIÓN DEL HÍGADO A LA LESIÓN METÁSTICA

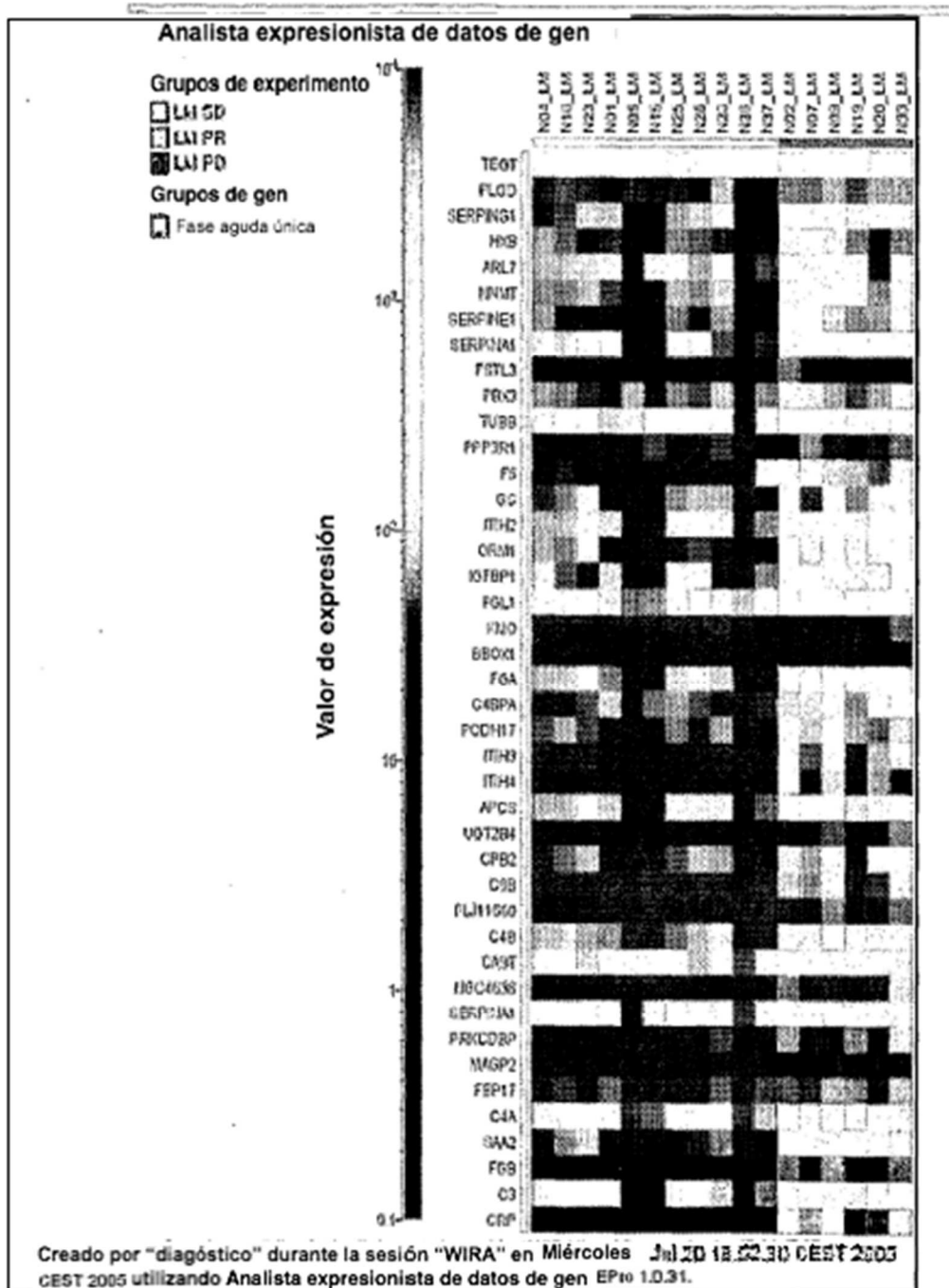


FIG. 15A

PERFILACIÓN DE EXPRESIÓN DE MARCADORES PREDICTIVOS DEL SUERO PARECIDOS A LA REACCIÓN DEL HÍGADO A LA LESIÓN METÁSTICA

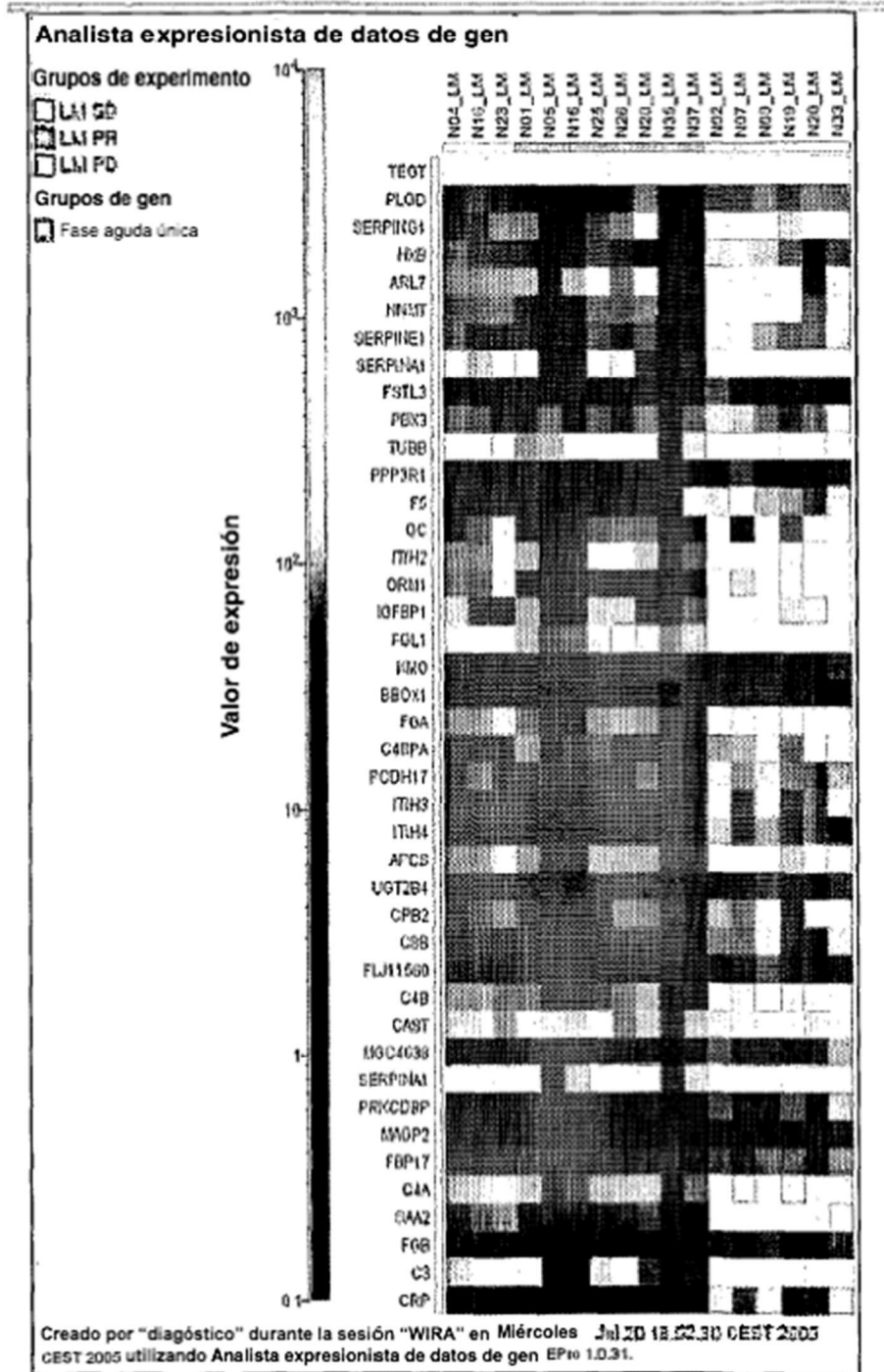


FIG. 16A

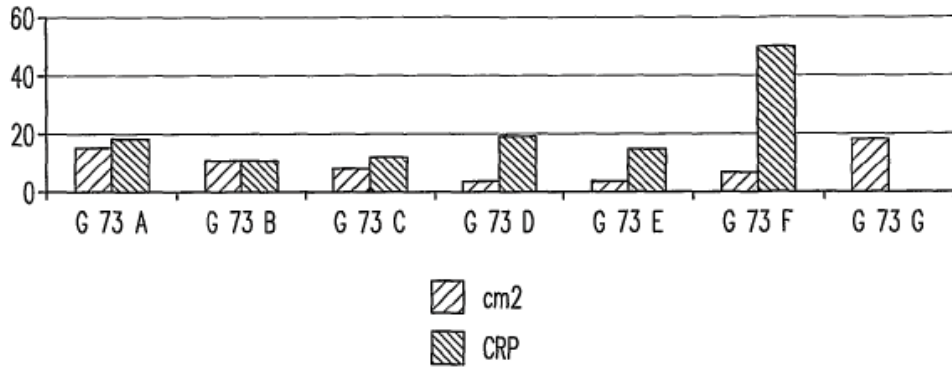


FIG. 16B

