

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 647 154**

51 Int. Cl.:

C12Q 1/68 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **04.11.2013 PCT/EP2013/072965**

87 Fecha y número de publicación internacional: **08.05.2014 WO14068124**

96 Fecha de presentación y número de la solicitud europea: **04.11.2013 E 13786660 (4)**

97 Fecha y número de publicación de la concesión europea: **13.09.2017 EP 2914739**

54 Título: **Combinaciones de biomarcadores para tumores colorrectales**

30 Prioridad:

05.11.2012 EP 12191236

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

19.12.2017

73 Titular/es:

**NOVIGENIX SA (100.0%)
Biopôle IV Route de la Corniche 3
1066 Epalinges, CH**

72 Inventor/es:

**THERIANOS, STAVROS;
RUEGG, CURZIO;
MONNIER-BENOIT, SYLVAIN;
CIARLONI, LAURA y
HOSSEINIAN, SAHAR**

74 Agente/Representante:

ELZABURU, S.L.P

ES 2 647 154 T3

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

DESCRIPCIÓN

Combinaciones de biomarcadores para tumores colorrectales

Solicitudes relacionadas

5 La presente invención reivindica prioridad y beneficio para la solicitud EP nº 12191234.4 presentada el 5 de noviembre de 2012.

Campo de la invención

La presente invención se refiere de un modo general a los biomarcadores de sangre periférica relacionados con tumores colorrectales, y a métodos de uso de los mismos.

Antecedentes de la invención

10 De un modo universal, el cáncer colorrectal (CCR) es el tercer cáncer más común, después del cáncer de pulmón y el de mama, y es causa destacada de las 650.000 muertes anuales relacionadas con el cáncer (*Jernal, A., Siegel, R., Ward, E., Hao, Y, Xu J, Thun, M Cancer Statistics 2009. CA Cancer J Clin 2009; 59; 225 - 249*). En Europa, es la segunda forma mayor de cáncer y la segunda causa de muerte, después del cáncer de pulmón. Sin embargo, las tasas de cribado de CCR se mantienen subóptimas (~20%) y están muy por debajo de las del cáncer de mama, de
15 cuello uterino y de próstata.

Los documentos US20100196889 y US20090215053 describen marcadores de la expresión génica para el CRC.

Así pues, hay una necesidad urgente de un método nuevo y más asequible para el cribado del CCR.

Compendio de la invención

20 La presente invención se define en las reivindicaciones anexas. La presente descripción se refiere a biomarcadores y métodos relativos para el cribado, la detección, el diagnóstico y el seguimiento de tumores colorrectales.

La presente descripción proporciona un método que incluye las etapas de (a) medir en una muestra obtenida de un sujeto la cantidad de cada biomarcador de un panel de biomarcadores incluyendo al menos un biomarcador de alta prioridad y al menos dos biomarcadores nucleares seleccionados entre el grupo que consiste en IL1B, PTGS2, S100A8, LTF, CXCL10 y CACNB4; (b) calcular una puntuación de probabilidad basada en la medición de la etapa
25 (a); y (c) descartar un tumor colorrectal para el sujeto si la puntuación en la etapa (b) es más baja que una puntuación pre-determinada; o admitir la probabilidad de tumor colorrectal para el sujeto si la puntuación en la etapa (b) es más alta que una puntuación predeterminada. La probabilidad de tumor colorrectal adicional puede determinarse además por la sensibilidad, especificidad, valor predictivo negativo (NPV) o valor predictivo positivo (PPV) asociados con la puntuación.

30 La presente invención proporciona también un método que incluye las etapas de (a) medir en una muestra obtenida de un sujeto la cantidad de cada biomarcador de un panel de biomarcadores que incluye al menos un biomarcador de alta prioridad y al menos dos biomarcadores nucleares seleccionados entre el grupo que consiste en IL-1B, PTGS2, S100A8, LTF, CXCLIO y CACNB4; (b) comparar la cantidad medida en la etapa (a) con un valor de referencia; y (c) clasificar el sujeto como más probable que tenga un tumor colorrectal cuando se detecta un
35 aumento o una disminución en la cantidad de cada biomarcador del panel de biomarcadores en relación con el valor de referencia en la etapa (b). El método puede incluir además una etapa de administrar al sujeto clasificado por la etapa (c) una cantidad terapéuticamente eficaz de al menos un agente de modulación colorrectal.

En ciertas realizaciones, el al menos un biomarcador de alta prioridad se selecciona entre el grupo que consiste en S100A8, LTF, CXCL 10 y CACNB4.

40 En ciertas realizaciones, el al menos un biomarcador de alta prioridad se selecciona entre el grupo que consiste en S100A8, LTF, CXCL10, CACNB4, MMP9, CXCL11, EGR1, JUN, TNFSF 13B, GATA2, MMP II, NME I, PTGES, CCR 1, CXCR3, FXD5, IL8, ITGA2, ITGB5, MAPK6, RHOC, BCL3, CD63, CES 1, MAP2K3, MSL 1 y PPARG.

En ciertas realizaciones, los al menos dos biomarcadores nucleares son IL 1B y PTGS2.

45 En ciertas realizaciones, los al menos dos biomarcadores nucleares son (a) IL 1B y PTGS2; (b) IL 1B, PTGS2 y S100A8; (c) IL 1B, PTGS2, S100A8 y LTF; (d) IL1B, PTGS2, S100A8, LTF, y CXCL10; o (e) IL1B, PTGS2, S100A8, LTF, CXCL 10 y CACNB4.

En ciertas realizaciones, cuando se descarta el tumor colorrectal, el sujeto no recibe un protocolo de tratamiento.

En ciertas realizaciones, cuando se descarta el tumor colorrectal en el sujeto, éste recibe un protocolo de tratamiento. Por ejemplo, el protocolo de tratamiento es una colonoscopia, una biopsia, una cirugía, una quimioterapia, una radioterapia, o cualquier combinación de estas técnicas.

5 En ciertas realizaciones, la puntuación de probabilidad puede ser calculada a partir de un modelo de predicción de regresión logística aplicado a la medida.

La muestra puede ser células mononucleares de sangre periférica, células sanguíneas, sangre entera, suero, plasma, células endoteliales, células tumorales circulantes, biopsias de tejido, fluido linfático, líquido ascítico, fluido intersticial, médula ósea, líquido cefalorraquídeo (CSF), saliva, mucosa, esputo, sudor u orina.

En ciertas realizaciones, el tumor colorrectal es adenoma o carcinoma.

10 En ciertas realizaciones, el sujeto está en riesgo de desarrollo de un tumor colorrectal.

La descripción también proporciona un kit para ser utilizado de acuerdo con el método mencionado anteriormente para detectar la presencia de tumores colorrectales. El kit comprende uno o más de un par de cebadores para la medición de uno o más biomarcadores, en particular el panel de biomarcadores que se describe en el presente texto (Tabla 1).

15 El kit puede comprender además una o más sondas, muestras de referencia para la realización de controles de calidad de medición, envases de material plástico y reactivos para llevar a cabo reacciones de prueba e instrucciones para el uso de los reactivos en el método de una cualquiera de las reivindicaciones precedentes.

A menos que se establezca otra cosa, todos los términos técnicos y científicos usados en el presente texto tienen el mismo significado que entiende normalmente un experto en la técnica a la que pertenece esta invención. Aunque en la práctica de la presente invención se pueden usar los métodos y materiales similares o equivalentes a los descritos en el presente texto, se describen a continuación métodos y materiales adecuados.

20 En caso de discrepancia, controlará la presente memoria descriptiva, incluyendo las definiciones.

Además, los materiales, métodos, y ejemplos descritos en el presente texto son solamente ilustrativos y no pretenden ser limitantes.

25 Otras características y ventajas de la invención serán evidentes a partir de la siguiente descripción detallada y de las reivindicaciones.

Breve descripción de los dibujos

30 La siguiente descripción detallada, que se da a título de ejemplo, pero que no pretende limitar la invención a realizaciones específicas descritas, puede entenderse junto con la figura adjunta, que se incorpora en el presente texto como referencia, en la que:

Figura 1. Los diagramas de caja representan IL1B, PTGS2, S100A8, MMP9, LTF, CXCL10 y CCR1, la expresión de genes durante la evolución y el progreso del cáncer colorrectal: controles, adenomas (POL) entre 1 - 2 cm, adenomas mayores que 2 cm y carcinomas en estadios I a IV. La unidad de medida corresponde a los valores deltaCp. Se ha de observar que se utilizaron diferentes escalas de unidad para diferentes gráficos de genes.

35 Figura 2. Gráficos representativos de curvas de características de funcionamiento del receptor (ROC: Receiver Operating Characteristics) para actuaciones del modelo de predicción del carcinoma colorrectal (panel izquierdo) o adenoma (panel derecho). El AUC (área bajo la curva) con 95% CI son 0,82 a 0,91 y 0,67 a 0,79 para el modelo de predicción de carcinoma o de adenoma, respectivamente. Se sacaron 1000 conjuntos de datos aleatorios con reemplazo de conjunto de entrenamiento (*bootstrap*); cada *bootstrap* tenía el mismo tamaño que el conjunto de formación. En cada iteración, los modelos son equipados y las muestras fuera de bolsas (no seleccionadas en cada *bootstrap*) se utilizaron para validar estos modelos. Los valores promedio sobre más de 1000 *bootstraps* para tasa de positivo verdadero y de positivo falsos están representados por las curvas.

45 Figura 3. Diagramas de dispersión de especificidad y sensibilidad para todos los modelos de predicción de carcinoma o de adenoma calculado en el conjunto de entrenamiento (azul), por *bootstrap* (verde), o en el conjunto de validación independiente (*rojo*).

Descripción detallada

La presente invención se basa parcialmente en el descubrimiento de que un pequeño panel de biomarcadores de la sangre es capaz de identificar específicamente y distinguir sujetos con lesiones colorrectales malignas y benignas procedentes de un sujeto sin dichas lesiones.

En consecuencia, la invención proporciona al paciente ventajas singulares, asociadas a la detección temprana de un tumor colorrectal en un paciente, incluyendo una mayor esperanza de vida, la disminución de la morbilidad y la mortalidad, la disminución de la exposición a la radiación durante las exploraciones y la repetición de las exploraciones, y un modelo de diagnóstico mínimamente invasivo. Es importante destacar que los métodos de la invención permiten que un paciente evite procedimientos invasivos, aumentando así la observancia por el paciente.

En la actualidad, las tasas de detección del cáncer colorrectal (CRC) siguen siendo subóptimas (~20%) y están muy por detrás de cáncer de mama, de cuello uterino y de próstata. Los países de la Unión Europea han introducido programas de cribado del CRC mediante colonoscopia, sigmoidoscopia, guayacol o prueba inmunoquímica de sangre fecal oculta (FOBT y FIT) para las personas mayores de 50 años. La falta de observancia de las recomendaciones de cribado se atribuye en gran parte a los aspectos desagradables y engorrosos de estos métodos. Por tanto, es imperativo encontrar una prueba de cribado de uso fácil, no invasiva y fiable, que satisfaga las necesidades de una numerosa población no explorada y de edad avanzada. Un análisis de sangre tendría las mayores posibilidades de aceptación por los pacientes y por la comunidad médica.

El diseño y las características de la invención descrita en el presente texto, en particular el uso de la sangre y las células mononucleares de sangre periférica (PBMCs) como espécimen de prueba, establece un método de cribado nuevo y más compatible para las pruebas de CRC previas a la colonoscopia.

Específicamente, la presente descripción proporciona biomarcadores relacionados con tumores colorrectales que, cuando se usan juntos en combinaciones de al menos dos biomarcadores nucleares con al menos un biomarcador de alta prioridad, que se selecciona individualmente de un panel de biomarcadores candidatos, se pueden usar tales combinaciones de biomarcadores para detectar tumores colorrectales. Por consiguiente, la presente invención proporciona métodos para el cribado, detección, diagnóstico y seguimiento de tumores colorrectales midiendo la cantidad de cada biomarcador de al menos tres biomarcadores de la Tabla I en una muestra (tal como PBMCs o células sanguíneas).

Particularmente, la presente descripción proporciona un método que incluye las etapas de (a) medir en una muestra obtenida de un sujeto la cantidad de cada biomarcador de un panel de biomarcadores que incluye al menos tres biomarcadores de la Tabla 1; (b) calcular una puntuación de probabilidad (o un valor de probabilidad), basado en la medición de la etapa (a); y (c-1) descartar el tumor colorrectal para el sujeto si la puntuación en la etapa (b) es menor que una puntuación predeterminada (o un preumbral determinado) o (c-2) admitir la probabilidad de tumor colorrectal para el sujeto si la puntuación en la etapa (b) es mayor que una puntuación predeterminada (o un umbral predeterminado).

En algunas realizaciones, el método incluye etapas de (a) recoger una muestra de ácido nucleico de una muestra biológica (por ejemplo, células o células sanguíneas mononucleares de sangre periférica) obtenida de un sujeto; (b) medir en la muestra de ácido nucleico la cantidad de cada biomarcador de un panel de biomarcadores que incluye al menos tres biomarcadores de la Tabla 1; (c) calcular una puntuación de probabilidad (o un valor de probabilidad), basado en la medición de la etapa (b); y (d-1) descartar el tumor colorrectal para el sujeto si la puntuación en la etapa (c) es menor que una puntuación predeterminada (o un umbral predeterminado) o (d-2) aceptar la probabilidad de tumor colorrectal para el sujeto si la puntuación en la etapa (c) es mayor que una puntuación predeterminada (o un umbral predeterminado).

Por ejemplo, los al menos tres biomarcadores de la Tabla I incluyen al menos un biomarcador de alta prioridad y por lo menos dos biomarcadores nucleares seleccionados del grupo que consiste en IL1B, PTGS2, S100A8, LTF, CXCL10 y CACNB4. Por ejemplo, los al menos tres biomarcadores de la Tabla 1 incluyen IL1B, PTGS2 y S100A8.

Cuando se descarta el tumor colorrectal el sujeto no recibe un protocolo de tratamiento. Sin embargo, cuando se acepta el tumor colorrectal en el sujeto, éste recibe un protocolo de tratamiento. El protocolo de tratamiento puede incluir, pero no se limita a estas técnicas, una colonoscopia, una biopsia, una cirugía, una quimioterapia, una radioterapia, o cualquier combinación de dichas técnicas.

La puntuación de probabilidad puede ser calculada de acuerdo con cualquier método conocido en la técnica. Por ejemplo, la puntuación de probabilidad se calcula a partir de un modelo de predicción de regresión logística aplicado a la medición. Por ejemplo, la puntuación de probabilidad se calcula por

$$\log\left(\frac{\Pr(y_i=1)}{1-\Pr(y_i=1)}\right) = \beta_0 + \beta_1 x_{1,i} + \dots + \beta_m x_{m,i}$$

y donde $x_{m,i}$ es un valor medido para el biomarcador m y sujeto i , y $(\beta_0, \beta_1, \dots, \beta_m)$ es un vector de coeficientes. En otras palabras, β_0 es una constante específica del panel y β_m es el correspondiente coeficiente de regresión logística del biomarcador m .

En algunas realizaciones, la probabilidad de tumor colorrectal también está determinada por la sensibilidad, la especificidad, el valor predictivo negativo (NPV) o el valor predictivo positivo (PPV) asociados con la puntuación.

La presente descripción proporciona también un método que incluye las etapas de (a) medir en una muestra obtenida de un sujeto la cantidad de cada biomarcador de un panel de biomarcadores que incluye al menos tres biomarcadores de la Tabla 1; (b) comparar la cantidad medida en la etapa (a) con un valor de referencia; y (c) clasificar al sujeto como sujeto con más probabilidades de tener un tumor colorrectal cuando se detecta en la etapa (b) un aumento o una disminución de la cantidad de cada biomarcador del panel de biomarcadores en relación con el valor de referencia.

En algunas realizaciones, el método incluye las etapas de (a) obtener una muestra de ácido nucleico de una muestra biológica (por ejemplo, células mononucleares de sangre periférica o células sanguíneas) obtenidas de un sujeto; (b) medir en la muestra de ácido nucleico la cantidad de cada biomarcador de un panel de biomarcadores que incluye al menos tres biomarcadores de la Tabla 1; (c) comparar la cantidad medida en la etapa (b) con un valor de referencia; y (d) clasificar al sujeto como sujeto con más probabilidades de tener un tumor colorrectal cuando se detecta en la etapa (c) un aumento o una disminución de la cantidad de cada biomarcador del panel de biomarcadores en relación con el valor de referencia.

Por ejemplo, los al menos tres biomarcadores de la Tabla 1 incluyen al menos un biomarcador de alta prioridad y al menos dos biomarcadores nucleares seleccionados del grupo que consiste en IL1B, PTGS2, S100A8, LTF, CXCL 10 y CACNB4. Por ejemplo, los al menos tres biomarcadores de la Tabla 1 incluyen IL1B, PTGS2 y S100A8.

En algunas realizaciones, el método incluye además una etapa de (i) seleccionar un régimen de tratamiento (o protocolo) para el sujeto clasificado como con más probabilidades de tener tumor colorrectal o (ii) administrar al sujeto clasificado como más propenso a tener un tumor colorrectal una cantidad terapéuticamente eficaz de al menos un agente modulador del cáncer colorrectal.

El régimen de tratamiento para el cáncer colorrectal es el estándar para los cuidados para el tratamiento de un tumor colorrectal (por ejemplo, pólipos colorrectales tales como adenomas y carcinomas colorrectales) como se describe en las más actuales directrices de National Comprehensive Cancer Network (NCCN). El régimen de tratamiento puede incluir administrar una cantidad terapéuticamente eficaz de al menos un agente modulador del cáncer colorrectal.

El uno o más agentes moduladores del cáncer colorrectal, pueden comprender un agente de alquilación, un agente antibiótico, un agente antimetabólico, un agente hormonal, un agente derivado de plantas, un agente retinoide, un inhibidor de la tirosina quinasa, un agente biológico, un agente de terapia génica, un inhibidor de histona desacetilasa, otro agente anticanceroso, o combinaciones de los mismos. Entre los ejemplos de agentes moduladores del cáncer colorrectal se incluyen, pero sin limitarse a ellos, Aducril (fluorouracilo), Avastin (bevacizumab), Bevacizumab, Camptosar (clorhidrato de Irinotecan), Capecitabina, Cetuximab, Efundex (Fluorouracilo), Eloxatin (oxaliplatino), Erbitux (Cetuximab), Fluoroplex (Fluorouracilo), Fluorouracilo, Irinotecan clorhidrato, Leucovorina cálcica, Oxaliplatino, Panitumumab, Regorafenib, Stivarga (Regorafenib), Vectibix (Panitumumab), Wellcovorin (Leucovorin cálcico), Xeloda (Capecitabina), Zaltrap (Ziv-Aflibercept), y Ziv -Aflibercept.

"Tratar" o "tratamiento" como se usa en el presente texto con respecto a una condición, puede referirse a la prevención de la condición, el retardo de la aparición o la velocidad de desarrollo de la condición, la reducción del riesgo de desarrollo de la condición, prevención o retardo del desarrollo de los síntomas asociados con la condición, reducción o finalización de los síntomas asociados con la condición, generación de una regresión completa o parcial de la condición, o alguna combinación de los mismos. Por ejemplo, las mejoras en los factores de riesgo del cáncer colorrectal como resultado de un tratamiento con uno o más agentes moduladores del cáncer colorrectal pueden comprender una reducción en la formación de pólipos, una reducción en el tamaño de los pólipos, una reducción en el número de pólipos, una reducción en los síntomas de la colitis ulcerosa, la enfermedad inflamatoria intestinal, y / o la enfermedad de Crohn, o combinaciones de los mismos.

La presente descripción proporciona también al menos tres biomarcadores de la Tabla 1 para su uso en un método de determinación de la probabilidad de un tumor colorrectal, la detección del tumor colorrectal, el diagnóstico del tumor colorrectal y/o el seguimiento del tumor colorrectal. El método puede incluir las etapas de:

(I) (a) medir en una muestra de ácido nucleico procedente de una muestra biológica (p. ej. células mononucleares de sangre periférica o células sanguíneas) la cantidad de cada biomarcador de los al menos tres biomarcadores de la Tabla 1; (b) calcular una puntuación de probabilidad basándose en la medición de la etapa (a); y (c-1) descartar el tumor colorrectal para el sujeto si la puntuación en la etapa (c) es menor que una puntuación predeterminada (o un umbral predeterminado) o (c-2) aceptar la probabilidad de tumor colorrectal para el sujeto si la puntuación en la etapa (b) es mayor que una puntuación predeterminada (o un umbral predeterminado); o

(II) (a) medir en una muestra de ácido nucleico procedente de una muestra biológica (p. ej. células mononucleares de sangre periférica o células de la sangre) la cantidad de cada biomarcador de los al menos tres biomarcadores de la Tabla 1; (b) comparar la cantidad medida en la etapa (a) con un valor de referencia; y (c) clasificar al sujeto como con más probabilidades de tener un tumor colorrectal cuando se detecta un aumento o una disminución de la cantidad de cada biomarcador del panel de biomarcadores en relación con el valor de referencia en la etapa (b).

Por ejemplo, los al menos tres biomarcadores de la Tabla 1 incluyen al menos un biomarcador de alta prioridad y al menos dos biomarcadores nucleares seleccionados del grupo que consiste en IL1B, PTGS2, S100A8, LTF, CXCL 10 y CACNB4. Por ejemplo los al menos tres biomarcadores de la Tabla 1 incluyen IL1B, PTGS2 y S100A8.

5 La etapa de medición de cualquier método o uso descrito en el presente texto puede incluir una etapa de puesta en contacto de la muestra de ácido nucleico obtenida del sujeto con uno o más cebadores descritos en el presente texto que se hibridan específicamente con el biomarcador de interés. La etapa de medición puede incluir además una etapa de amplificación del biomarcador de interés con dichos uno o más cebadores.

10 La medición real de los niveles de los biomarcadores se puede determinar a nivel de ácido nucleico o de proteína usando cualquier método conocido en la técnica. Por ejemplo, a nivel de ácido nucleico los biomarcadores se pueden medir extrayendo ácidos ribonucleicos de la muestra y realizando cualquier tipo de PCR cuantitativa sobre los ácidos nucleicos de transcripción inversa. Otra forma de detectar los biomarcadores puede ser también mediante un análisis del transcriptoma completo basado en metodologías de secuenciación de alto rendimiento, p. ej. RNA-seq, o en la tecnología de micromatrices, p. ej. matrices de Affymetrix.

15 A título de ejemplo, otros métodos que pueden ser utilizados para medir el biomarcador pueden implicar cualquier otro método de cuantificación conocido en la técnica de los ácidos nucleicos, tales como, pero sin limitarse a ellas, la amplificación de secuencias específicas, las sondas de oligonucleótidos, la hibridación de los genes diana con sondas complementarias, la fragmentación por endonucleasas de restricción y estudio de los fragmentos resultantes (polimorfismos), técnicas de geles de campo pulsado, amplificación isotérmica de desplazamiento múltiple, amplificación o replicación por círculo rodante, inmuno-PCR, entre otros conocidos por los expertos en la técnica.

20 Usando la información proporcionada por las entradas de base de datos para las secuencias de biomarcadores, los niveles de expresión de biomarcadores pueden detectarse y medirse usando técnicas bien conocidas por los expertos en la técnica. Por ejemplo las secuencias de biomarcadores dentro de las entradas de base de datos de secuencia, o dentro de las secuencias descritas en el presente documento, se pueden utilizar para construir sondas y cebadores para la detección de secuencias de ARNm de los biomarcadores en métodos que específicamente y de forma preferible, cuantitativamente amplifican secuencias específicas de ácidos nucleicos tales como la reacción en cadena de la polimerasa en tiempo real basada en la transcripción inversa (RT-qPCR).

25 Los niveles de los biomarcadores también se pueden determinar a nivel de proteínas, p. ej. midiendo los niveles de péptidos codificados por los productos génicos descritos en el presente texto, o las actividades de los mismos. Tales métodos son bien conocidos en la técnica e incluyen, p. ej., inmunoensayos basados en anticuerpos contra proteínas codificadas por los genes, aptámeros o huellas moleculares. Alternativamente, puede seleccionarse un método adecuado para determinar la actividad de las proteínas codificadas por los genes de biomarcadores de acuerdo con la actividad de cada proteína analizada.

30 Las proteínas de biomarcadores, polipéptidos, mutaciones, y polimorfismos de los mismos pueden ser detectados de cualquier manera adecuada, pero se detectan típicamente poniendo en contacto una muestra del sujeto con un anticuerpo que se une a la proteína biomarcadora, polipéptido, mutación, o polimorfismo y después detectando la presencia o ausencia de un producto de reacción. El anticuerpo puede ser monoclonal, policlonal, quimérico, o un fragmento de los precedentes, como se discutió detalladamente con anterioridad, y la etapa de detectar el producto de reacción puede llevarse a cabo con cualquier inmunoensayo adecuado. La muestra del sujeto es típicamente una muestra biológica como se describió anteriormente, y puede ser la misma muestra utilizada para llevar a cabo el método antes descrito.

Los expertos en la técnica estarán familiarizados con numerosos inmunoensayos específicos y formatos de ensayo de amplificación de ácidos nucleicos y variaciones de los mismos que pueden ser útiles para llevar a cabo las realizaciones de la invención descrita en el presente texto.

45 Preferiblemente, los niveles de expresión de los biomarcadores de la presente invención son detectados por RT-qPCR, y en particular por PCR en tiempo real, como se describe más adelante en el presente texto.

50 En general, el ARN total puede ser aislado de la muestra objetivo, tal como sangre periférica o PBMC, usando cualquier procedimiento de aislamiento. Este ARN se puede usar después para generar primera copia de la cadena de ADN (ADNc) usando cualquier procedimiento, p. ej. usando cebadores aleatorios, cebadores oligo-dT o cebadores aleatorios-oligo-dT que son cebadores oligo-dT acoplados en el extremo 3' para tramos cortos de secuencia específica que cubren todas las combinaciones posibles. El ADNc se puede entonces utilizar como plantilla en PCR cuantitativa.

55 En PCR en tiempo real la cuantificación de productos de la PCR se basa p. ej. en incrementos de fluorescencia, liberada en cada ciclo de amplificación de la reacción, p. ej. por una sonda que se hibrida con una porción del producto de amplificación. Los planteamientos de fluorescencia utilizados en PCR cuantitativa en tiempo real se basan típicamente en un colorante fluorescente reportero tal como FAM, fluoresceína, HEX, TET, etc. y un apagador tal como TAMRA, DABSYL, Black Hole, etc. Cuando el apagador se separa de la sonda durante la fase de extensión

de la PCR, puede medirse la fluorescencia del reportero. Sistemas como Universal ProbeLibrary, Molecular Beacons (balizas moleculares), sondas TaqMan, cebadores Scorpion o cebadores Sunrise y otros, utilizan este método para realizar PCR cuantitativa en tiempo real. Alternativamente, se puede medir la fluorescencia desde fluorocromos de ADN-intercalante tal como SYBR Green.

- 5 La abundancia de moléculas de ARN diana se puede realizar mediante PCR en tiempo real de una manera relativa o absoluta. Los métodos relativos se pueden basar en la determinación del ciclo umbral (Ct) o, en el caso de los instrumentos de PCR de Roche, el punto de cruce (Cp). La abundancia relativa de moléculas ARN se calcula entonces por el método delta Ct (delta Cp) restando el valor Ct (Cp) de uno o más genes de limpieza o *housekeeping*. Un ejemplo de genes de limpieza que pueden ser utilizados se expone en la Tabla 2.
- 10 Alternativamente, la medición absoluta se puede realizar determinando el número de copias de la molécula de ARN diana por medio de curvas de calibración.

La Tabla 1 enumera un ejemplo de cebadores directos e inversos, así como el número de identificación de la sonda Universal Probelibrary (Roche) que podría ser utilizada para la medición del biomarcador correspondiente por PCR en tiempo real.

- 15 Los biomarcadores y métodos de la presente invención permiten que un experto en la técnica detecte, identifique, diagnostique o evalúe de otra manera a aquellos sujetos que no presentan ningún síntoma de tumores colorrectales, pero que no obstante pueden estar en riesgo de desarrollar tumores colorrectales, o de experimentar los síntomas característicos de una enfermedad cancerosa.

- 20 La Tabla 1 proporciona información que incluye una lista no exhaustiva de biomarcadores de sangre periférica relacionados con tumores colorrectales según la invención. Un experto en la técnica reconocerá que los biomarcadores presentados en el presente texto incluyen proteínas, ácidos nucleicos (ADNcs, ARNms, ARNs, ADNs) y metabolitos, junto con sus polimorfismos, mutantes, variantes de isoformas, metabolitos relacionados, derivados, precursores incluyendo ácidos nucleicos y pro-proteínas, productos de escisión, complejos de proteína-ligando, variantes modificadas postraduccionalmente (tales como entrecruzamiento o glicosilación), fragmentos, y
- 25 productos de degradación, así como cualquier ácido nucleico multi-unidad, proteínas y estructuras de glicoproteína que comprenden cualquiera de los biomarcadores como subunidades constitutivas de la estructura completamente montada. Toda la expresión de biomarcadores en muestras de sangre ha sido validada mediante experimentación.

Tabla 1. Biomarcadores sanguíneos relacionados con tumores colorrectales

Símbolo del gen	Descripción del gen	Secuencia cebador directo (SEC ID Nº)	Secuencia cebador inverso (SEC ID Nº)	ID sonda UPL
BCL3	B-cell CLL/linfoma 3	ACAACAACCTACGGCAGACA (1)	CCACAGACGGTAATGTGGTG (2)	76
CACNB4	canal de calcio dependiente de voltaje, subunidad beta-4	TCCAAGCACAGCTATCTCCTT (3)	CCCTCTTCCACCAGCCTTC (4)	138
CCR1	receptor 1 de quimiocina (motivo C-C)	AGTGATTCCACAGTGACTCCA (5)	GGCAGATGCTGGCTACTGAT (6)	95
CD63	molécula CD-63	GAATGAAATGTGTGAAGTTCTTGC (7)	GCAATCAGTCCCAGTGCAC (8)	18
CES1	carboxilesterasa 1	CAGGAGITTTGGCTGGTTGAT (9)	CAGTTGCCCTTCGGAGAGT (10)	136
CXCL10	ligando 10 de quimiocina (motivo C-X-C)	AAAAGGTATGCAATCAAATCTGC (11)	AAGAATTTGGGCCCTTG (12)	86
CXCL11	ligando 11 de quimiocina (motivo C-X-C)	TTGTGTGCTACAGTTGTTCAAGG (13)	TCTGCCACTTTCAGTCTTTA (14)	81
CXCR3	receptor 3 de quimiocina (motivo C-X-C)	ACCACAAGCACCAAAGCAG (15)	GGCGTCATTTAGCACTTGGT (16)	27
EGR1	respuesta 1 del crecimiento temprano	AGCACCTGACCCAGAGT (17)	GGCAGTCGAGTGGTTTGG (18)	54
FXYS5	regulador 5 del transporte de iones que contiene dominio FXYS	ACCACGTCCAGTTCTTCAAGC (19)	GGGCTGGAGTTCTGTAGACT (20)	45
GATA2	proteína 2 de unión de GATA	CACAAGATGAATGGGCAGAA (21)	TGACAATTCACACACAGGTG (22)	117
IL1B	interleucina 1, beta	AGCTGATGGCCCTAACAGAG (23)	TCGGAGATTCGTAGCTGGAT (24)	85
IL8	interleucina 8	TAGCCAGGATCCACAAGTCC (25)	CTGTGAGGTAAGATGGTGGCTA (26)	98
ITGA2	integrina, alfa 2 (CD49B)	AACATGAGCCTCGGCTTG (27)	GCCCACAGAGGACCACAT (28)	154
ITGB5	integrina, beta 5	GCATGAGCACCAAGAGAG (29)	GCAGGTCTGGTTGTCAGGTT (30)	40
JUN	proto-oncogen jun	AGTCAGGCAGACAGACAGAC (31)	AAAATAAGATTTCAGTTCGGACTAT (32)	20
LTF	lactotransferrina	TAAGTGGAAAGCCTGAAAC (33)	CCATTCTCCAAATTTAGCC (34)	22
MAP2K3	proteína cinasa cinasa 3 activada por mitógeno	CGAGTTTGTGGACTTCACTGC (35)	AAGGTGAAGAAGGGGTGCTC (36)	1
MAPK6	proteína cinasa 6 activada por mitógeno	TGGATGAAACTCACAGTCACATT (37)	GGCCAATCATGCTCTGAAA (38)	48
MMP11	metaloproteína 11 de matriz (estromelina 3)	AAGAGGTTCTGCTTCTTGG (39)	CCATGGGAACCGAAGGAT (40)	14
MMP9	metaloproteína 9 de matriz (gelatinasa B)	ATCCGGCACCTCTATGGTC (41)	CAGACCGTCGGGGGAG (42)	77
MSL1	homólogo masculino-específico letal 1 (<i>Drosophila</i>)	CAGGCCAAGGAAAAGGAGAT (43)	CGTTCAATCCGAGCAAGG (44)	17
NME1	células no metastásicas 1, proteína (NM23A)	CCTAAGCAGCTGGAAGGAAC (45)	CGCTTGATAATCTCTCCACA (46)	100
PPARG	receptor gamma activado por el factor proliferador de peroxisomas	GACAGGAAAGACAACAGACAAATC (47)	GGGGTGATGTGTTGAACTTG (48)	7
PTGES	prostaglandina E sintasa	AGAAGGCCTTTGCCAACC (49)	GATGGTCTCCATGTCTTCC (50)	122
PTGS2	prostaglandina-endoperoxido sintasa 2	CGCTCAGCCATACAGCAA (51)	TCATACATACACCTCGGTTTGA (52)	150
RHOC	familia de genes homólogos de ras, miembro C	AGCACACCAGGAGAGAGCTG (53)	GTAGCCAAGGCAGCTGATCC (54)	92
S100A8	proteína A8 de unión al calcio S100	CAGCTGTCTTTCAGAAGACTG (55)	CTTCTCCAGTCCGGTCAAC (56)	105
TNFSF13B	superfamilia factor de necrosis tumoral (ligando), miembro 13b	CTCAAGACTGCTTCAACTGA (57)	AAGCTGAGAAGCCATGGAAC (58)	112

Tabla 2: genes de limpieza usados para normalización de la expresión génica

Símbolo del gen	Descripción del gen	Secuencia cebador directo (SEC ID N°)	Secuencia cebador inverso (SEC ID N°)	ID sonda UPL
NACA	subunidad alfa de complejo asociado a polipéptido naciente	TGCTACAGAGCAGGAGTTGC (59)	TCCTGTTCTCAAGCTCTGGT (60)	45
RPLP0	proteína ribosomal, grande, PO	TCGACAATGGCAGCATCTAC (61)	GCCAATCTGCAGACAGACAC (62)	6
TPT1	proteína tumoral, controlada traduccionalmente 1	CAATCAAAGGGAAACTTGAAGAA (63)	GATTCATGTTTTACCAATAAAGAAC (64)	54

5 Estos biomarcadores sanguíneos pueden medirse y utilizarse en combinación en un modelo de predicción que comprende tres o más biomarcadores. En algunos aspectos, los 29 biomarcadores enumerados en la Tabla 1 pueden ser medidos y utilizados. Los intervalos preferidos a partir de los que se mide y se utiliza el número de biomarcadores incluyen márgenes limitados por cualquier mínimo seleccionado entre 3 y 29.

En ciertas realizaciones, los al menos tres biomarcadores de la Tabla 1 incluyen (a) IL1B, PTGS2, LTF; (b) IL1B, PTGS2, S100A8; (c) IL1B, PTGS2, S100A8, LTF; o (d) IL1B, PTGS2, S100A8, LTF.

10 En particular, los al menos tres biomarcadores de la Tabla 1 incluyen al menos dos biomarcadores nucleares (también llamados biomarcadores indispensables) en combinación con al menos un biomarcador de alta prioridad, que se elige individualmente entre un panel de biomarcadores.

Un "biomarcador nuclear" usado en el presente texto se refiere a un biomarcador que tiene un nivel de importancia de 1 o 2, de acuerdo con la Tabla 6. Un biomarcador nuclear se elige entre el grupo que consiste en IL1B, PTGS2, S100A8, LTF, CXCL10 y CACNB4.

15 Un "biomarcador de alta prioridad" usado en el presente texto se refiere a un biomarcador que tiene un nivel de importancia de 2, 3 o 4, de acuerdo con la Tabla 6.

En algunas realizaciones, los al menos tres biomarcadores de la Tabla 1 utilizados en cualquier método o uso descrito en el presente texto incluyen al menos un biomarcador de alta prioridad y al menos dos biomarcadores nucleares seleccionados entre el grupo que consiste en IL-1B, PTGS2, S100A8, LTF, CXCL10 y CACNB4.

20 Por ejemplo los al menos dos biomarcadores nucleares son (a) IL1B y PTGS2; (b) IL1B, PTGS2 y S100A8; (c) IL1B, PTGS2, S100A8 y LTF; (d) IL1B, PTGS2, S100A8, LTF, y CXCL 10; o (e) IL1B, PTGS2, S100A8, LTF, CXCL10 y CACNB4.

25 Por ejemplo los dos biomarcadores nucleares son IL1B y PTGS2 y se combinan con al menos un biomarcador elegido entre el panel de biomarcadores de alta prioridad que comprende S100A8, LTF, CXCL10, CACNB4, MMP9, CXCL 11, EGR1, JUN, TNFSF 13B, GATA2, MMP 11, NME1, PTGES, CCR 1, CXCR3, FXD5, IL-8, ITGA2, 1TGB5, MAPK6, RHOC, BCL3, CD63, CES 1, MAP2K3, MSL1 y PPARG.

En particular, el panel de biomarcadores de alta prioridad comprende S100A8, LTF, CXCLIO, y/o CACNB4.

30 En cualquier método y uso descrito en el presente texto, pueden utilizarse los biomarcadores de prioridad 1, 2, 3, 4 o más alta. Por ejemplo el al menos un biomarcador de alta prioridad (1, 2, 3, 4, o más) se elige entre el grupo que consiste en S100A8, LTF, CXCL10, CACNB4, MMP9, CXCL 11, EGR 1, JUN, TNFSF13B, GATA2, MMP 11, NME1, PTGES, CCR1, CXCR3, FXD5, IL8, ITGA2, ITGB5, MAPK6, RHOC, BCL3, CD63, CES 1, MAP2K3, MSL1 y PPARG.

35 Por ejemplo el al menos un biomarcador de alta prioridad incluye dos biomarcadores seleccionados entre el grupo que consiste en S100A8, LTF, CXCL10 y CACNB4. Por ejemplo el al menos un biomarcador de alta prioridad incluye tres biomarcadores seleccionados entre el grupo que consiste en S100A8, LTF, CXCL 10 y CACNB4. Por ejemplo el al menos un biomarcador de alta prioridad incluye cuatro biomarcadores S100A8, LTF, CXCL10, y CACNB4.

40 Por ejemplo los biomarcadores de alta prioridad son (a) CXCL 10 y S100A8, (b) CXCL10 y LTF, (c) CXCL10 y CACNB4, (d) S100A8 y LTF, (e) S100A8 y CACNB4, (f) LTF y CACNB4, (g) CXCL10 y S100A8 y LTF, (h) CXCL10 y S100A8 y CACNB4, (i) CXCL10 y LTF y CACNB4, (j) S100A8 y LTF y CACNB4, o (k) CXCL10 y S100A8 y LTF y CACNB4.

En ciertas realizaciones, los biomarcadores utilizados en la presente invención son cualquier combinación de una combinación del Grupo A con una combinación del Grupo B, eliminando el duplicado si hay alguno (véase la Tabla más adelante). Por ejemplo los biomarcadores son combinación de (a) del Grupo A y combinación de (a) del grupo B.

Combinaciones de biomarcadores nucleares (Grupo A)	Combinaciones de biomarcadores de alta prioridad (Grupo B)
(a) IL1B y PTGS2	(a) CXCL10 y S100A8
(b) IL 1B, PTGS2 y S100A8	(b) CXCL10 y LTF
(c) IL1B, PTGS2, S100A8 y LTF	(c) CXCL10 y CACNB4
(d) IL1B, PTGS2, S100A8, LTF, y CXCL10	(d) S100A8 y LTF
(e) IL1B, PTGS2, S100A8, LTF, CXCL10 y CACNB4	(e) S100A8 y CACNB4
	(f) LTF y CACNB4
	(g) CXCL10, S100A8 y LTF
	(h) CXCL10, S100A8 y CACNB4
	(i) CXCL10, LTF y CACNB4
	(j) S100A8, LTF y CACNB4
	(k) CXCL10, S100A8, LTF y CACNB4

5

En algunas realizaciones, los biomarcadores nucleares son IL1B, PTGS2 y S100A8 y los biomarcadores de alta prioridad son (i) BCL3, CACNB4, CCR1, CXCL10, ITGA2, ITGB5, LTF, MAP2K3, MAPK6, MMP11, PTGES, y TNFSF13B; o (ii) CACNB4, CXCL10, LTF, MMP11, y PTGES.

10 Un "biomarcador" utilizado en la presente invención se refiere a un indicador molecular de una propiedad biológica específica; una característica o faceta bioquímica que puede utilizarse para detectar el cáncer colorrectal. "Biomarcador" comprende, sin limitación, proteínas, ácidos nucleicos y metabolitos, junto con sus polimorfismos, mutantes, variantes de isoformas, metabolitos relacionados, derivados, precursores incluyendo ácidos nucleicos y pro-proteínas, productos de escisión, complejos de proteína-ligando, post-variantes traduccionalmente modificados (tales como la reticulación o glicosilación), fragmentos, y productos de degradación, así como cualquier ácido nucleico multi-unidad, proteínas y estructuras de glicoproteína que comprenden cualquiera de los biomarcadores como subunidades constitutivas de la estructura montada totalmente, y otros analitos o medidas derivadas de la muestra.

15 "Medir", "medición", "detección" y "detectar" significan la evaluación de la presencia, ausencia, cantidad (que puede ser una cantidad efectiva) de una sustancia dada en una muestra clínica o derivada de un sujeto, incluyendo los niveles de concentración cualitativos o cuantitativos de tales sustancias, o bien de otro modo la evaluación de los valores o la categorización de los parámetros clínicos de un sujeto.

20 "Alterado", "un aumento" o "una disminución" se refiere a un cambio o diferencia detectables entre el biomarcador medido y el valor de referencia de un estado, perfil, medición, o similares razonablemente comparables. Un experto en la técnica podrá determinar un cambio medible razonable. Tales cambios pueden ser todos o ninguno. Pueden ser graduales y no precisan ser lineales. Pueden ser de varios órdenes de magnitud. Un cambio puede ser un aumento o una disminución de 1%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 99%, 100%, o más, o cualquier valor comprendido entre 0% y 100%.

25 Alternativamente, el cambio puede ser 1 vez, 1,5 veces, 2 veces, 3 veces, 4 veces, 5 veces o más, o cualquier valor entre 1 vez y cinco veces. El cambio puede ser estadísticamente significativo con un valor p de 0,1, 0,05, 0,001, o 0,0001.

30 Se entiende que la expresión "tumor colorrectal" incluye un amplio espectro de tumores derivados del epitelio que oscilan desde crecimientos benignos hasta cáncer invasivo. Estos incluyen pólipos colorrectales, tales como adenomas y carcinomas colorrectales.

Las expresiones "pólipos adenomatosos" y "adenoma" se utilizan indistintamente.

Los términos "individuo", "huésped", "paciente" y "sujeto" se utilizan indistintamente. Como se usa en el presente texto, un "sujeto" incluye un mamífero. El mamífero puede ser, p. ej. un mamífero humano o no humano apropiado, tal como un primate, ratón, rata, perro, gato, vaca, caballo, cabra, camello, oveja o cerdo. El sujeto también puede ser un pájaro o un ave. En una realización, el mamífero es una persona. Un sujeto puede ser macho o hembra.

Un sujeto puede ser uno que no ha sido diagnosticado o identificado previamente como afectado por un tumor colorrectal. Un sujeto puede ser un sujeto sano que está clasificado como sujeto con bajo riesgo de desarrollar una enfermedad del colon (como pólipos colorrectales o cáncer colorrectal). Alternativamente, un sujeto puede ser una persona en riesgo de desarrollar un tumor colorrectal. Un factor de riesgo es todo aquello que afecte a las probabilidades del sujeto de contraer una enfermedad como un tumor colorrectal. Los factores de riesgo que pueden aumentar las posibilidades de una persona de desarrollar pólipos colorrectales o cáncer colorrectal incluyen, pero sin limitarse a ellos, la edad, historia de pólipos colorrectales o cáncer colorrectal (especialmente si los pólipos son grandes o si hay muchos), historia de enfermedad intestinal inflamatoria (tales como colitis ulcerosa y enfermedad de Crohn), antecedentes de cáncer colorrectal o pólipos adenomatosos, síndromes genéticos heredados (tales como la poliposis adenomatosa familiar (FAP), cáncer de colon hereditario no asociado a poliposis (HNPCC), síndrome de Turcot, síndrome de Peutz-Jegher, poliposis asociada a MUTYH), tipo diabetes, factores relacionados con el estilo de vida (dieta, peso y ejercicio), inactividad física, la obesidad, el tabaquismo y el consumo excesivo de alcohol.

Una "muestra" en el contexto de la presente invención es una muestra biológica aislada de un sujeto y puede incluir, a título de ejemplo y sin limitarse a ellas, sangre completa, suero, plasma, células de la sangre, células mononucleares de sangre periférica, células endoteliales, células tumorales circulantes, biopsias de tejido, fluido linfático, fluido ascítico, fluido intersticial, médula ósea, líquido cefalorraquídeo (LCR), saliva, mucosa, esputo, sudor, orina, o cualquier otra secreción, excreción, u otros fluidos corporales. En algunas realizaciones, la muestra se refiere a células mononucleares de sangre periférica o células de la sangre.

"Células mononucleares de sangre periférica" (PBMC) se refiere a cualquier célula presente en la sangre que tiene un núcleo redondo. Esta fracción se aísla convencionalmente por centrifugación de sangre entera en un gradiente de densidad de líquido. Contiene principalmente linfocitos y monocitos mientras que excluye los glóbulos rojos y los granulocitos (eosinófilos, basófilos y neutrófilos). También podrían estar presentes en esta fracción células raras con un núcleo redondo, tales como células progenitoras endoteliales o células tumorales circulantes.

El término "cebador" se refiere a una cadena de ácido nucleico que sirve como punto de partida para la replicación del ADN.

Los términos "sonda" y "sonda de hidrólisis" se refieren a una cadena corta de ácido nucleico diseñado para hibridarse con una región dentro del amplicón y es doblemente marcada con un colorante reportero y un colorante de apagamiento. La estrecha proximidad del agente de apagamiento suprime la fluorescencia del colorante reportero. La sonda se basa en la actividad de 5'-3' exonucleasa de la polimerasa Taq, que degrada una sonda de ADN no extensible hibridada durante la etapa de extensión de la PCR. Una vez que la polimerasa Taq ha degradado la sonda, la fluorescencia del reportero aumenta a una velocidad que es proporcional a la cantidad de plantilla presente.

La expresión "expresión génica" significa la producción de una proteína o un ARNm funcional a partir de sus genes.

Los términos "firma", "clasificador", "modelo" y "predicador" se usan indistintamente. Se refieren a un algoritmo que discrimina entre los estados de la enfermedad con un nivel de significación estadística predeterminado. Un clasificador de dos clases es un algoritmo que utiliza puntos de datos a partir de mediciones de una muestra y clasifica los datos en uno de dos grupos. En ciertas realizaciones, los datos usados en el clasificador son la expresión relativa de ácidos nucleicos o proteínas en una muestra biológica. Los niveles de expresión de proteína o ácido nucleico en un sujeto pueden ser comparados con los niveles en pacientes previamente diagnosticados como libres de la enfermedad o con una condición especificada.

Un "nivel/valor de referencia o de línea de base" como se usa en el presente texto se puede usar de forma indistinta y se entiende que es relativa a un número o valor derivado de los estudios de población, incluyendo, sin limitación, tales sujetos que tienen rango de edad similar, el estado de la enfermedad (p. ej. estadio), los sujetos en el mismo grupo étnico o similar, o con relación a la muestra de partida de un sujeto sometido a tratamiento para el cáncer. Tales valores de referencia se pueden derivar de los análisis estadísticos y/o datos de predicción de riesgo de las poblaciones obtenidos de algoritmos matemáticos y los índices calculados de cáncer colorrectal. También pueden construirse y utilizarse índices de referencia que emplean algoritmos y otros métodos de clasificación estadística y estructural.

En algunas realizaciones de la presente invención, el valor de referencia o línea de base es el nivel de expresión de un biomarcador de interés en particular en una muestra testigo derivada de uno o más sujetos sanos o sujetos que no han sido diagnosticados con ningún cáncer.

En algunas realizaciones de la presente invención, el valor de referencia o línea de base es el nivel de expresión de un biomarcador de interés en particular en una muestra obtenida del mismo sujeto antes de cualquier tratamiento de cáncer. En otras realizaciones de la presente invención, el valor de referencia o línea de base es el nivel de expresión de un biomarcador de interés en particular en una muestra obtenida del mismo sujeto durante un tratamiento de cáncer. Alternativamente, el valor de referencia o línea de base es una medición previa del nivel de expresión de un gen de interés en particular en una muestra obtenida previamente a partir de la misma materia o de un sujeto que tiene un margen de edad o estatus de la enfermedad (p. ej. el estadio) similares al del sujeto evaluado.

El término "descartar" como se usa en el presente texto significa que el sujeto se selecciona para no recibir un protocolo de tratamiento.

El término "admitir" tal como se utiliza en el presente texto significa que el sujeto se selecciona para recibir un protocolo de tratamiento.

"Alterado", "cambiado" o "significativamente diferente" se refieren a un cambio o diferencia detectable a partir de un estado, perfil, medida, o similares, razonablemente comparable. Un experto en la técnica deberá ser capaz de determinar un cambio medible razonable. Tales cambios pueden ser todos o ninguno. Pueden ser graduales y no necesitan ser lineales. Pueden ser de varios órdenes de magnitud. Un cambio puede ser un aumento o una disminución de 1%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 99%, 100%, o más, o cualquier valor entre 0% y 100%. Alternativamente, el cambio puede ser 1 vez, 1,5 veces, 2 veces, 3 veces, 4 veces, 5 veces o más, o cualquier valor entre 1 vez y cinco veces. El cambio puede ser estadísticamente significativo con un valor p de 0,1, 0,05, 0,001, o 0,0001.

El término "normalización" o "normalizador" como se usa aquí se refiere a la expresión de un valor diferencial en términos de un valor estándar para ajustar los efectos que surgen de la variación técnica debida a la manipulación de muestras, preparación de muestras y medición por espectrometría de masas en lugar de la variación biológica de la concentración de proteína en una muestra. Por ejemplo, cuando se mide la expresión de una proteína (ácido nucleico) expresada diferencialmente, el valor absoluto para la expresión de la proteína (ácido nucleico) puede expresarse en términos de un valor absoluto para la expresión de una proteína estándar (ácido nucleico) que es sustancialmente constante en expresión. Esto evita que la variación técnica de la preparación de la muestra y la medición de PCR impidan la medición de los niveles de concentración de proteína (ácido nucleico) en la muestra.

El término "puntuación" o "puntuar" se refiere a calcular una probabilidad (o un valor de probabilidad) mediante el modelo (por ejemplo, un modelo de regresión logística) para una muestra. Para la presente invención, los valores más próximos a 1,0 se usan para representar la probabilidad de que una muestra se derive de un paciente con una lesión del colon (como pólipos, adenoma, carcinomas colorrectales o tumores colorrectales), valores más próximos cercanos a 0,0 representan la probabilidad de que una muestra se derive de un paciente sin una lesión de colon (como pólipos, adenoma, carcinomas colorrectales o tumores colorrectales).

Una "puntuación predeterminada" se refiere a un umbral de probabilidad que se ha determinado durante la fase de modelado/entrenamiento mediante, por ejemplo, regresión logística y análisis ROC, y que define la probabilidad de tumor colorrectal y/o el diagnóstico de tumor colorrectal. Un experto en la técnica puede determinar fácilmente dicha puntuación de acuerdo con cualquier método disponible en la técnica.

El método propuesto para analizar y usar un perfil de biomarcador para la detección, el diagnóstico y el seguimiento de tumores colorrectales es a) extraer el ARN de células mononucleares de sangre periférica, b) transcripción inversa de dicho ARN en ADNc, c) realización de una amplificación por PCR en tiempo real específica para cada biomarcador de interés y d) realización del análisis de datos estadísticos derivados de la composición y de los métodos descritos, utilizando, por ejemplo, la regresión logística penalizada para construir modelos de predicción.

A título de ejemplo y sin pretender limitar ningún aspecto de la presente invención, se pueden aplicar otras composiciones y métodos para analizar datos derivados de la medición de uno o más biomarcadores de la presente invención.

Todos los datos y análisis estadísticos se pueden realizar en el software R (entorno de software libre R-CRAN para computación y gráficos estadísticos), MATLAB (The Math Works), SPSS (IBM), SYSTAT (Systat Software Inc.) y otros soportes que permiten análisis numéricos.

Los métodos que pueden usarse para analizar datos derivados de la medición de dichos biomarcadores relacionados con tumores colorrectales implican cualquier análisis estadístico de datos reconocido por la técnica, tales como regresión logística logística y penalizada, máquina de vector soporte, bosque aleatorio, lógica difusa, red neuronal, agrupamiento de genes, herramientas de minería de datos y otros algoritmos o índices calculados conocidos en la técnica y descritos en este documento.

La regresión logística (McCullagh, P. y Nelder, J. A. (1983) *Generalized Linear Models, Monographs on Statistics and Applied Probability*) es uno de los métodos comunes para discriminar dos grupos. Si se define que y_i es 0 o 1 de

acuerdo con su grupo, se puede modelar a través de una regresión logística como

$$\log\left(\frac{\Pr(y_i=1)}{1-\Pr(y_i=1)}\right) = \beta_0 + \beta_1 x_{1,i} + \dots + \beta_m x_{m,i}$$

donde $x_{m,i}$ es un valor deltaCp para el biomarcador m y el sujeto i , y $(\beta_0, \beta_1, \dots, \beta_m)$ es un vector de coeficientes (parámetros a estimar) para una regresión logística multivariada. Para estimar estos parámetros se puede usar el método de máxima verosimilitud.

5

Por ejemplo, el adenoma puede determinarse mediante una ecuación de modelo predictivo:

$$\log\left(\frac{\Pr(y_i=1)}{1-\Pr(y_i=1)}\right) = -0.668 + 0.07 \times \text{BCL3} + 0.449 \times \text{CACNB4} - 0.274 \times \text{CCR1} + 0.174 \times \text{CXCL10} - 0.260 \times \text{IL1B} - 0.115 \times \text{ITGA2} - 0.083 \times \text{ITGB5} - 0.130 \times \text{LTF} - 0.024 \times \text{MAP2K3} - 0.213 \times \text{MAPK6} + 0.297 \times \text{MMP11} + 0.001 \times \text{PTGES} - 0.140 \times \text{PTGS2} - 0.145 \times \text{S100A8} - 0.212 \times \text{TNFSF13B}.$$

Por ejemplo, el carcinoma se puede determinar mediante una ecuación de modelo predictivo:

$$\log\left(\frac{\Pr(y_i=1)}{1-\Pr(y_i=1)}\right) = -8.544 + 0.707 \times \text{CACNB4} + 0.688 \times \text{CXCL10} - 0.592 \times \text{IL1B} - 0.234 \times \text{LTF} + 0.044 \times \text{MMP11} + 0.105 \times \text{PTGES} - 0.143 \times \text{PTGS2} - 1.605 \times \text{S100A8}.$$

10 Se observa que para el conjunto de datos altamente dimensionales con multi-co-linealidad, la regresión logística puede fallar. Dado que algunos de los biomarcadores seleccionados pueden estar altamente correlacionados, una solución es usar regresión logística penalizada.

15 La regresión logística penalizada se basa en la ecuación matemática derivada de la regresión logística. Más específicamente, la regresión logística penalizada es una regresión cresta (*ridge*) para el modelo logístico con una norma L2 o una penalización de norma L1. Para estimar los parámetros en este método, se añade una penalización cuadrática (L2) y/o norma L1 en la probabilidad logarítmica que debe maximizarse. Para elegir el mejor valor de λ_1 y λ_2 , se usa la validación cruzada con los criterios A1C. Para ajustarse al modelo logístico penalizado, se pueden utilizar los siguientes algoritmos (paquetes en R Cran, software estadístico): *glm*path (Park MY y Hastie T (2006) *An L1 Regularization-path Algorithm For Generalized Linear Models. A Generalization of the LARS algorithm for GLMs and the Cox proportional hazard model*, penalizado (Goeman, J. (2010) *L1 (lasso) and L2 (ridge) penalized estimation in GMLs and in the Cox model*) y *glmnet* (Hasti, T., Tibshirani and R. Friedman, J. (2010) *Lasso and elastic-net regularized generalized linear models*) con diferentes parámetros de ajuste.

25 La aplicación de la regresión logística a problemas biológicos es rutinaria en la técnica. Varios softwares de análisis estadístico, tales como los mencionados anteriormente, se pueden usar para construir modelos de regresión logística. Los modelos de regresión logística ajustados se prueban preguntando si el modelo puede predecir correctamente el resultado clínico utilizando datos de pacientes distintos de aquellos con los que fue ajustado el modelo de regresión logística, pero que tienen un resultado clínico conocido. Después del entrenamiento, el resultado del modelo de 0 (control) a 1 (cáncer) puede calcularse a ciegas por el error promedio de todas las predicciones de N (un grupo de validación). Basándose en los valores de salida, se puede construir la curva característica operativa del receptor (ROC) para calcular el resultado de la predicción clínica: especificidad y sensibilidad de la detección de cáncer CRC. Hay medidas estadísticas de la realización de una prueba de clasificación binaria. La sensibilidad mide la proporción de positivos reales que están correctamente identificados como tales (p. ej. el porcentaje de personas enfermas que se identifican correctamente como portadoras de la afección). La especificidad mide la proporción de negativos que están identificados correctamente (por ejemplo, el porcentaje de personas sanas que están identificadas correctamente como no portadoras de la condición). Un predictor perfecto se describiría como 100% sensible (es decir, que predice a todas las personas del grupo enfermo como enfermas) y 100% específica (es decir, no predice a nadie del grupo sano como enfermo). Sin embargo, cualquier predictor tendrá un límite de error mínimo.

40 Una realización de la presente invención es un modelo predictivo que comprende una combinación/perfil de biomarcadores de células mononucleares de sangre periférica que detecta tumores colorrectales preferiblemente con una sensibilidad igual o superior al 60% y una especificidad igual o superior al 84%.

La expresión "sensibilidad de una prueba" se refiere a la probabilidad de que un resultado de una prueba sea

positivo cuando la enfermedad está presente en el paciente (tasa positiva verdadera). Esto se deduce del número de pacientes con la enfermedad que tienen un resultado positivo de la prueba (positivo verdadero) dividido por el número total de pacientes con la enfermedad, incluidos aquellos con resultados positivos verdaderos y los pacientes con la enfermedad que tienen un resultado negativo, es decir, negativo falso.

5 La expresión "especificidad de una prueba" se refiere a la probabilidad de que el resultado de una prueba sea negativo cuando la enfermedad no está presente en el paciente (tasa negativa verdadera). Esto se deduce del número de pacientes sin la enfermedad que tienen un resultado de la prueba negativo (negativo verdadero) dividido por todos los pacientes sin la enfermedad, incluidos aquellos que tienen un resultado negativo verdadero y aquellos
10 pacientes sin la enfermedad que tienen un resultado positivo de la prueba, p. ej. positivo falso. Mientras que la sensibilidad, la especificidad, la tasa de positivo verdadero o falso y la tasa de negativo verdadero o falso de una prueba proporcionan una indicación del rendimiento de una prueba, p. ej. en relación con otras pruebas, para tomar una decisión clínica para un paciente en particular basándose en el resultado de la prueba el clínico requiere parámetros de rendimiento de la prueba con respecto a una población determinada.

15 La expresión "valor predictivo positivo" (PPV) se refiere a la probabilidad de que un resultado positivo identifique correctamente a un paciente que tiene la enfermedad, que es el número de positivos verdaderos dividido por la suma de positivos verdaderos y positivos falsos.

20 La expresión "valor predictivo negativo" o "NPV" se refiere a la probabilidad de que una prueba negativa identifique correctamente a un paciente sin la enfermedad, que es el número de verdaderos negativos dividido por la suma de verdaderos negativos y falsos negativos. Al igual que el PPV, también se ve afectado intrínsecamente por la prevalencia de la enfermedad y la probabilidad previa a la prueba de la población que se pretende ensayar. Un resultado positivo de una prueba con un PPV suficiente puede usarse para admitir la enfermedad en un paciente, mientras que un resultado negativo de una prueba con un NPV suficiente puede usarse para descartar la enfermedad, si es conocida la prevalencia de la enfermedad para la población dada, de la que el paciente puede considerarse parte.

25 Una "curva de Características Operativas del Receptor (ROC)", como se usa en el presente texto, se refiere a un gráfico de la tasa positiva verdadera (sensibilidad) frente a la tasa de falsos positivos (especificidad) para un sistema clasificador binario ya que su umbral de discriminación se varía. Una curva ROC puede ser representada de manera equivalente representando la fracción de verdaderos positivos de los positivos (TPR = tasa de verdaderos positivos) frente a la fracción de falsos positivos fuera de los negativos (FPR = tasa de falsos positivos). Cada punto de la
30 curva ROC representa un par sensibilidad/especificidad correspondiente a un umbral de decisión en particular.

El AUC representa el área bajo la curva ROC. El AUC es una indicación general de la precisión diagnóstica de 1) un biomarcador o un panel de biomarcadores y 2) una curva ROC. El AUC está determinado por la "regla trapezoidal". Para una curva dada, los puntos de datos se conectan por segmentos en línea recta, se erigen perpendiculares desde la abscisa a cada punto de datos, y se calcula la suma de las áreas de los triángulos y trapezoides así
35 construidos. En ciertas realizaciones de los métodos proporcionados en el presente texto, una proteína biomarcadora tiene un AUC en el intervalo de aproximadamente 0,75 a 1,0. En algunas de estas realizaciones, el AUC está en el intervalo de aproximadamente 0,8 a 0,85, 0,9 a 0,95 o 0,95 a 1,0.

Los métodos proporcionados en el presente texto son mínimamente invasivos y plantean poco o ningún riesgo de efectos adversos. Como tales, se pueden usar para diagnosticar, controlar y proporcionar tratamiento clínico a
40 sujetos que no muestran ningún síntoma de una enfermedad del colon (tumor colorrectal) y sujetos clasificados como de bajo riesgo para desarrollar una enfermedad del colon (tumor colorrectal). Por ejemplo, los métodos descritos en el presente texto se pueden usar para diagnosticar tumores colorrectales en un sujeto que no presenta un pólipo colorrectal y/o no ha presentado un pólipo colorrectal en el pasado, pero que, a pesar de ello, considera que corre el riesgo de desarrollar un pólipo colorrectal y/o una enfermedad de colon. De forma similar, los métodos
45 descritos en el presente texto se pueden usar como una medida estrictamente precautoria para diagnosticar sujetos sanos que están clasificados de bajo riesgo para desarrollar una enfermedad de colon.

La descripción proporciona además un kit para usar de acuerdo con el método anteriormente mencionado para detectar la presencia de tumores colorrectales a partir de una muestra de sangre periférica, en particular de una muestra de células mononucleares de sangre periférica (PBMC). El kit puede comprender uno o más de un par de cebadores para medir uno o más biomarcadores enumerados en la Tabla 1, en particular el panel de biomarcadores como se describe en el presente texto. Además, el kit puede comprender pares de cebadores específicos para uno o más genes de limpieza, por ejemplo para los genes TPT1, RPLP0 y NACA descritos en la Tabla 2. El kit puede comprender además una o más sondas, muestras de referencia para realizar controles de calidad de medición, recipientes de plástico y reactivos para realizar reacciones de ensayo e instrucciones para usar los reactivos en el
50 método de una cualquiera de las reivindicaciones precedentes. Opcionalmente, un kit puede comprender instrucciones de uso en forma de una etiqueta o un prospecto por separado. Los kits pueden contener reactivos que se unen específicamente a proteínas en los paneles descritos en el presente texto. Estos reactivos pueden incluir anticuerpos.

Se ha de entender que, aunque la invención se ha descrito junto con la descripción detallada de la misma, la descripción precedente pretende ilustrar y no limitar el alcance de la invención, que se define por el alcance de las reivindicaciones adjuntas. Otros aspectos, ventajas y modificaciones están dentro del alcance de la presente descripción.

- 5 Los ejemplos que siguen se proporcionan para ilustrar mejor la invención reivindicada y no deben interpretarse como limitantes del alcance de la invención. En la medida en que se mencionan materiales específicos, es meramente a título de ilustración y no pretende limitar la invención.

Ejemplo

Métodos para la detección del cáncer colorrectal a partir de una muestra de sangre.

10 *Pacientes y muestras.*

Se inscribieron prospectivamente 181 sujetos mayores de 50 años en un estudio de control de casos que incluía seis centros. Al hacer la colonoscopia, se diagnosticó que eran sujetos control (n = 75), pacientes con adenoma \geq 1 cm (n = 61) o pacientes con cáncer colorrectal (CRC) en estadio I - IV (n = 45). Se obtuvo el consentimiento informado por escrito de todos los participantes del estudio que se adhirieron a las directrices éticas locales. Ninguno de los sujetos tenía antecedentes familiares de primer grado de CRC o una predisposición al CRC conocida, historia previa de cáncer, enfermedades autoinmunitarias u otros trastornos inflamatorios, fiebre ($>$ 38 °C) o infecciones en las últimas 4 semanas antes de la colonoscopia, ni ninguna otra enfermedad definida en el estudio. Se extrajo sangre de todos los sujetos hasta 30 días antes o hasta 12 semanas después de la colonoscopia y antes de cualquier resección de pólipos o de cualquier tratamiento específico del cáncer. El diagnóstico de adenoma y cáncer se confirmó histológicamente a partir de una biopsia o una muestra quirúrgica.

20 *Recogida de sangre y extracción de ARN.*

Se sacó una muestra de sangre a todos los sujetos inscritos. Se recogieron muestras de sangre periférica para la prueba de Colox® en tubos Becton Dickinson (BD) Vacutainer® CPT™ (4 x 4 ml). Los tubos rellenos de CPT™ se mantuvieron a temperatura ambiente y se realizó la separación de células sanguíneas mononucleares (PBMC) dentro de las 6 horas, de acuerdo con las instrucciones del fabricante. Los pélets o sedimentos de PBMC se resuspendieron en RNAlater® Solution (Life Technologies) y se almacenaron a -20°C.

La purificación automatizada del ARN total se llevó a cabo en QIAcube mediante el kit RNeasy Mini (QIAGEN). Esto incluía un tratamiento con DNasa. La concentración de ARN se determinó mediante un espectrofotómetro Nanodrop y el control de calidad del ARN se efectuó mediante un Agilent 2100 Bioanalyzer (Agilent Technologies). Las muestras con un RIN $<$ 5 se consideraron de mala calidad y se desecharon. El ARN total aislado se dividió en partes alícuotas y se almacenó a -80°C.

30 *Cebadores y sondas.*

Se adquirieron de Roche ensayos de PCR en tiempo real (ensayos RealTime ready Custom RT-qPCR) y se basaron en sondas cortas de hidrólisis Universal ProbeLibrary (UPL). La UPL se basa en solamente 165 sondas cortas de hidrólisis (8 - 9 nucleótidos). Se marcaron en el extremo 5' con fluoresceína (FAM) y en el extremo 3' con un colorante de apagado oscuro. Con el fin de mantener la especificidad y la temperatura de fusión (Tm) que requieren las sondas qPCR de hibridación, se incorporan Ácidos Nucleicos Bloqueados (LNA: Locked Nucleic Acids) a la secuencia de cada sonda UPL. Los LNAs son análogos de nucleótidos de ADN con una fuerza de unión incrementada en comparación con los nucleótidos de ADN estándar.

40 Las secuencias del cebador directo e inverso, así como el número de identificación de la sonda UPL, se enumeran en la Tabla 1. Los ensayos de PCR en tiempo real se precargaron en RealTime Ready™ Custom panel 384-32, placas LC480 de 384 pocillos (Roche).

RT-PCR cuantitativa.

45 Se sometieron a transcripción inversa 200 ng de ARN total en ADNc usando el kit de síntesis de ADNc VILO SuperScript® (Invitrogen) de acuerdo con las instrucciones del fabricante.

El análisis de PCR en tiempo real se realizó en el instrumento Lightcycler 480. Las reacciones PCR se llevaron a cabo por duplicado en placas de 384 pocillos en 10 μ l de volumen total. Cada pocillo se cargó con 5 μ l de RealTime Ready™ DNA Probes Master Mix (Roche) y el equivalente de cDNA de 2,5 ng de RNA total mediante el robot de pipeteo MICROLAB® STARLet (Hamilton). La amplificación se realizó después de 1 minuto a 95°C para la activación de Taq ADN polimerasa, que fue seguida por 40 ciclos de 2 segundos a 95°C y 30 segundos a 60°C. Se generaron muestras de control positivo y negativo con cada lote RT y se incluyeron en cada placa y para cada ensayo. El control negativo fue una mezcla RT-PCR sin ARN y ADNc para confirmar que no se había producido contaminación

5 durante el ensayo. El control positivo se hizo con una cantidad estandarizada de ARN Human Universal Reference (Clontech), y fue distribuido en partes alícuotas y almacenado a -80°C. Para la validación de la ejecución de la PCR, el control negativo no debe dar amplificación o un valor del punto de cruce (Cp) (el análogo Lightcycler de Ct) igual o mayor que 35, y el control positivo un valor Cp, para cada gen objetivo, que cae dentro de un intervalo predeterminado. Los valores Cp se calculan automáticamente mediante el método Abs Quant/2nd Derivative Max del software de análisis LightCycler 480. Los valores de expresión génica (Cp) se normalizaron mediante el método delta Ct de acuerdo con la fórmula: $\text{deltaCp} = \text{Cp}_{\text{target}} - \text{Cp}_{\text{ref}}$. En el caso presente, Cp_{ref} es el valor Cp medio de 3 genes de referencia (RPLP0, NACA, TPT1).

Análisis estadístico.

10 Se usaron valores de expresión génica normalizada (deltaCp) para todos los análisis estadísticos, que se realizaron con el software R (entorno de software libre R-CRAN para computación estadística y gráficos). Todos los análisis de laboratorio se realizaron con la fórmula: deltaCp = Cp_{target} - Cp_{ref}. Una vez revisadas médicamente las muestras y bloqueados los datos de acuerdo con el Diagnoplex Data Management Manual, el estadístico de la prueba quedó sin cegamiento para el análisis.

15 Se agruparon 181 sujetos de acuerdo con el diagnóstico, el sexo y el país de origen, y se asignaron aleatoriamente a un conjunto de entrenamiento y validación, con la proporción de dos tercios (n = 120) y un tercio (n = 61), respectivamente. La distribución de muestra entre los tres grupos bajo investigación se expone en la Tabla 3. Este tamaño de muestra permitió un nivel de significación $\alpha = 0,05$ y una potencia $1 - \beta = 0,90$.

20 **Tabla 3.** Distribución de las muestras en los grupos de control, adenoma y carcinoma del conjunto de Entrenamiento y Validación

	Conjunto de entrenamiento	Conjunto de validación
Controles	50	25
Adenoma ≥ 1 cm	40	21
Adenoma 1 cm - 2 cm	24	10
Adenoma ≥ 2 cm	16	11
CRC	30	15
Estadio I	9	3
Estadio II	4	5
Estadio III	10	3
Estadio IV	7	4

25 Los sujetos en el conjunto de entrenamiento se usaron para ajustarse a modelos de regresión logística penalizada en relación con la hipótesis de control frente al CRC y control frente a adenoma ≥ 1 cm. El error de predicción para los modelos ajustados se estimó en el conjunto de validación. Se seleccionaron clasificadores predictivos de acuerdo con las actuaciones en el conjunto de entrenamiento y validación.

Las predicciones del conjunto de entrenamiento y validación se utilizaron para determinar los rendimientos de la prueba, como la especificidad (negativo verdadero/control total) y la sensibilidad (positiva verdadera/enfermedad total) para la detección de CCR y adenoma ≥ 1 cm.

Resultados.

30 **Análisis descriptivo.**

La edad, el género, el sitio de recogida de muestra no tuvo influencia sobre el análisis de expresión génica. La expresión génica normalizada de 29 biomarcadores se ha comparado entre muestras y los niveles de expresión eran en general homogéneos.

35 El análisis de la distribución de quintiles en cada biomarcador se realizó a través de la gráfica quintil-quintil (gráfica Q-Q) frente a una distribución normal teórica. En general, los valores de expresión de biomarcadores se distribuyeron normalmente y solo unos pocos biomarcadores muestran una desviación de la distribución de referencia en las colas. Se realizaron análisis de correlación (de Pearson) y agrupamiento jerárquico de 29 biomarcadores. Sólo unas cuantas variables parecían estar fuertemente correlacionadas (CXCL10 y CXCL11, IL1B y PTGS2, EGR1 y PTGS2: coeficiente de correlación: 0,8; S110A8 y TNFSF13B, ITGB5 y ITGA2, JUN y coeficiente

de correlación IL8: 0,7); los genes restantes muestran solo una correlación débil o no muestran correlación.

Tabla 4. Características demográficas de la cohorte de estudio

	Controles	Adenoma ≥ 1 cm	CRC	Estadio I	Estadio II	Estadio III	Estadio IV	Estadio desconocido
Nº total	124	100	74	20	15	21	18	8
Edad (media ± S.D.)	60,7±7,7	67,4±8,1	69,5±9,8	70,7±9,1	70,3±8,06	68,0±12,5	69,3±7,3	70,4±11,1
Varones (%)	45 %	64 %	62,2 %	65 %	60 %	57,1 %	66,6 %	87,5 %
Hembras (%)	55 %	36 %	37,8 %	35 %	40 %	42,9 %	34,4 %	12,5 %

Análisis y clasificación de biomarcadores.

5 El conjunto de datos fue objeto de una serie de pruebas estadísticas para determinar la significación estadística de cada uno de los 29 biomarcadores en la discriminación de controles de muestras de carcinoma o adenoma. Al dibujar con reemplazo a partir del conjunto de entrenamiento (método de *bootstrap* o arranque), se crearon conjuntos de muestras del mismo tamaño que el conjunto original. Esto se repitió independientemente 1000 veces. La prueba de la *t* de Student, la regresión logística univariada (Dobson, A. J. (2002) *An introduction to generalized linear models. 2nd ed., Chapman & Hall/CRC Texts in Statistical Science Series, McCullagh, P. y Nelder, J. A. (1983) Generalized linear models, Monographs on Statistics and Applied Probability*), y la prueba de clasificación de Wilcoxon se aplicaron al conjunto de entrenamiento (Tabla 5) y a cada uno de los conjuntos derivados del *bootstrap*. Además, se calculó el incremento en veces (FC: fold change) de la expresión génica entre el control y el CRC o adenomas grandes, para cada biomarcador en los 1000 conjuntos. Los resultados obtenidos se resumen para cada biomarcador por la frecuencia de valores *p* significativos (< 0,01 o 0,05) de los 1000 resultados y por el incremento en veces de la expresión génica medio. Todos los resultados de las pruebas se categorizaron por magnitud y se dio una puntuación a cada categoría. Se obtuvo una puntuación final por la suma de puntuaciones parciales que resultan de la clasificación de los 29 biomarcadores (Tabla 5).

20 Basándose en la puntuación obtenida, seis biomarcadores, IL1B, CCR1, PTGS2, S100A8, PPARG y LTF, parecían ser muy fuertes en el control discriminante de las muestras de carcinoma mediante análisis univariado. Todos esos genes estaban regulados al alza en las PBMC de pacientes con cáncer. Los tres mejores genes regulados a la baja fueron: CACNB4, MMP11 y CXCL10.

25 **Tabla 5.** Los biomarcadores fueron clasificados de acuerdo con su capacidad para separar los sujetos de control del grupo CRC: Esta capacidad se resume por una puntuación derivada de una serie de análisis estadísticos descritos anteriormente. Como ejemplo, se enumeran el valor *p* del test de la *t* y el incremento en veces (*fold change*: FC) de la expresión génica

	Función biológica	valor <i>p</i> de Wilcoxon	Frec. valor <i>p</i> <0,01/1000	FC CRC/Con	Dirección
IL1B	Respuesta inmunitaria / Inflamación / Quimiotaxis	4,19 E-04	847	2,14	Arriba
CCR1	Adhesión celular / Quimiotaxis	4,42 E-04	860	1,65	Arriba
PTGS2	Metabolismo de lípidos	7,68 E-04	804	2,11	Arriba
S100AB	Respuesta inmunitaria / Inflamación / Quimiotaxis	5,07 E-06	992	1,65	Arriba
PPARG	Transcripción / Ciclo celular / Regulación	3,59 E-03	656	1,41	Arriba
LTF	Transporte de iones	2,14 E-02	419	2,36	Arriba
EGR1	Transcripción / Ciclo celular / Regulación	3,79 E-01	64	1,42	Arriba
MAPK6	Transcripción / Ciclo celular / Regulación	3,95 E-03	661	1,15	Arriba
CACNB4	Transporte de iones	1,31 E-02	452	1,30	Abajo
MMP11	Degradación del colágeno	1,66 E-02	442	1,30	Abajo
TNFSF13B	Respuesta inmunitaria / Inflamación / Quimiotaxis	1,03 E-02	524	1,21	Arriba
CXCL10	Respuesta inmunitaria / Inflamación / Quimiotaxis	7,13 E-02	220	1,29	Abajo
CD63	Diferenciación / Estructura	3,14 E-02	375	1,14	Arriba

	Función biológica	valor p de Wilcoxon	Frec. valor p <0,01/1000	FC CRC/Con	Dirección
CES1	Respuesta inmunitaria / Inflamación / Quimiotaxis	5,70 E-02	263	1,18	Arriba
MMP9	Degradación del colágeno	1,21 E-01	140	1,35	Arriba
PTGES	Metabolismo de lípidos	3,28 E-01	47	1,27	Abajo
BCL3	Transcripción / Ciclo celular / Regulación	1,67 E-01	110	1,12	Arriba
CXCR3	Respuesta inmunitaria / Inflamación / Quimiotaxis	8,04 E-01	14	1,04	Abajo
FXD5	Adhesión celular / Quimiotaxis	9,48 E-01	7	1,00	Arriba
GATA2	Transcripción / Ciclo celular / Regulación	6,40 E-01	16	1,09	Abajo
IL8	Transcripción / Ciclo celular / Regulación	9,76 E-01	18	1,04	Arriba
ITGA2	Transcripción / Ciclo celular / Regulación	3,90 E-01	42	1,21	Arriba
ITGB5	Adhesión celular / Quimiotaxis	7,85 E-01	17	1,12	Arriba
JUN	Adhesión celular / Quimiotaxis	5,61 E-01	23	1,10	Abajo
MAP2K3	Diferenciación / Estructura	9,89 E-02	170	1,09	Arriba
MSL1	Diferenciación / Estructura	9,60 E-01	8	1,01	Abajo
NME1	Respuesta inmunitaria / Inflamación / Quimiotaxis	7,39 E-01	12	1,02	Arriba
RHOC	Transporte de iones	2,02 E-01	91	1,11	Abajo
CXCL11	Respuesta inmunitaria / Inflamación / Quimiotaxis	3,20 E-01	48	1,23	Abajo

Los niveles de expresión génica fueron analizados también a través de los siguientes subgrupos de muestras: control, adenoma entre 1 y 2 cm, adenoma > 2cm y 4 estadios de carcinoma (estadio I, II, III y IV).

5 Se observó una clara tendencia a la sobreexpresión durante la evolución de la enfermedad para: IL1B, PTGS2, LTF, MMP9, S100A8, CXCL10 y CCR1, (figura 1), confirmando su potencial como biomarcadores capaces de discriminar entre pacientes portadores de carcinoma y de adenoma y sujetos de control.

Como se esperaba, los genes de referencia no mostraron ninguna tendencia durante la evolución de la enfermedad.

10 Estos análisis junto con los resultados del análisis de regresión logística permitieron a los autores dar prioridad a los 29 biomarcadores (Tabla 6) y definir un grupo de biomarcadores nucleares "indispensables" compuestos por PTGS2 y IL1B, y un grupo de marcadores de alta prioridad compuesto por S100A8, LTF, CXCL10 y CACNB4.

Tabla 6. Lista priorizada de los 29 biomarcadores CRC

Gen	Nivel de importancia
IL1B	1
PTGS2	1
S100AB	2
LTF	2
CXCL10	2
CACNB4	2
MMP9	3
CXCL11	3
EGR1	3
JUN	3
TNFSF13B	3
GATA2	3

Gen	Nivel de importancia
MMP11	3
NME1	3
PTGES	3
CCR1	3
CXCR3	3
FXYD5	3
IL8	3
ITGA2	3
ITGB5	3
MAPK6	3
RHOC	3
BCL3	4
CD63	4
CES1	4
MAP2K3	4
MSL1	4
PPARG	4

Clasificadores predictivos para la detección de tumores colorrectales.

El conjunto de entrenamiento que contiene datos derivados de los 29 biomarcadores se usó para ajustar modelos de regresión logística penalizada independientemente para cada una de las siguientes hipótesis:

- 5
 - Controles frente a adenomas ≥ 1 cm y CRC (subconjunto POLCRC)
 - Controles frente CRC (subconjunto CRC)
 - Controles frente a adenomas ≥ 1 cm (subconjunto POL)
 - Controles frente a CRC (estadio I, II) (subconjunto CRCI-II)

10 Los modelos de regresión logística penalizada fueron validados directamente en el conjunto de entrenamiento o por el método de arranque (*bootstrap*) no superpuesto: 1000 conjuntos de datos aleatorios fueron extraídos con reemplazo del conjunto de entrenamiento; cada conjunto de datos tenía el mismo tamaño que el conjunto de entrenamiento. El modelo fue re-ajustado en cada arranque y validado con las muestras *out-of-bag* (oob). Los valores promedio de la especificidad y la sensibilidad sobre 1000 arranques fueron calculados al valor de corte de la puntuación de la probabilidad indicada y se expone en la Tabla 7. Los modelos diferentes se definen por diferentes combinaciones de biomarcadores.

15 Se calculó la especificidad y sensibilidad a diferentes valores de corte de la puntuación de probabilidad y se generaron curvas de Características de Funcionamiento del Receptor (ROC) (Figura 2).

Tabla 7. Tabla compendio de la especificidad y sensibilidad de diferentes modelos estadísticos obtenidos por arranque (*bootstrap*) en el conjunto de entrenamiento. El modelado y los comportamientos se calcularon con el subconjunto de datos indicado y al valor de corte de la puntuación de probabilidad dada

Validación del arranque (<i>bootstrap</i>)				
Modelo	Subconjunto	Valor de corte	Sens.	Esp.
<i>GLMpath NF</i>	CRCI-II	0.30	0.70	0.93
	CRC	0.50	0.69	0.90
	POLCRC	0.75	0.54	0.85
	POL	0.65	0.51	0.84
<i>GLMnet Alpha0.5</i>	CRCI-II	0.30	0.76	0.92
	CRC	0.5	0.73	0.89
	POLCRC	0.75	0.49	0.88
	POL	0.6	0.48	0.85
<i>GLMnet Alpha0.6</i>	CRCI-II	0.30	0.75	0.92
	CRC	0.5	0.72	0.89
	POLCRC	0.75	0.49	0.88
	POL	0.6	0.48	0.85
<i>GLMnet Alpha0.8</i>	CRCI-II	0.30	0.74	0.92
	CRC	0.5	0.71	0.88
	POLCRC	0.75	0.49	0.87
	POL	0.6	0.47	0.84
<i>PenalizedL1</i>	CRCI-II	0.30	0.73	0.91
	CRC	0.5	0.72	0.89
	POLCRC	0.75	0.49	0.89
	POL	0.6	0.40	0.87
<i>PenalizedL1L2</i>	CRCI-II	0.30	0.72	0.92
	CRC	0.5	0.68	0.90
	POLCRC	0.70	0.53	0.87
	POL	0.6	0.36	0.91

- 5 Los modelos de mejor comportamiento fueron seleccionados de acuerdo con los siguientes criterios:
1. Comportamiento. Los clasificadores fueron seleccionados de acuerdo con el mejor comportamiento en el conjunto de entrenamiento y en el conjunto de validación. Para evaluar el comportamiento del clasificador se utilizó la suma de especificidad y sensibilidad como parámetro de clasificación.
 2. Estabilidad. Se definieron clasificadores estables a través del conjunto de entrenamiento y del conjunto de validación como aquellos que muestran el mínimo de distancia euclidiana de dos dimensiones calculada entre sensibilidad y especificidad del conjunto de entrenamiento y el conjunto de validación.

La distancia euclidiana en 2 dimensiones está dada por:

$$\begin{aligned}
 p &= (x_1, y_1) \\
 q &= (x_2, y_2) \\
 \text{Distancia euclidiana } (p, q) &= \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}
 \end{aligned}$$

- 15 Los parámetros x e y se sustituyen por la sensibilidad y especificidad del modelo en el conjunto de entrenamiento y el conjunto de validación.

$$\begin{aligned}
 p &= (\text{SENSI}_{TS}, \text{SPECI}_{TS}) \\
 q &= (\text{SENSI}_{VS}, \text{SENSI}_{VS})
 \end{aligned}$$

- 20 Basándose en el criterio expuesto anteriormente, se eligieron dos modelos de regresión logística penalizada como los de mejor comportamiento: uno para la detección del adenoma ≥ 1 cm y otro para la detección del carcinoma colorrectal.

La ecuación del modelo predictivo de adenoma es:

$$\log\left(\frac{\Pr(y_i=1)}{1-\Pr(y_i=1)}\right) = -0.668 + 0.07 \times \text{BCL3} + 0.449 \times \text{CACNB4} - 0.274 \times \text{CCR1} + 0.174 \times \text{CXCL10} - 0.260 \times \text{IL1B} - 0.115 \times \text{ITGA2} - 0.083 \times \text{ITGB5} - 0.130 \times \text{LTF} - 0.024 \times \text{MAP2K3} - 0.213 \times \text{MAPK6} + 0.297 \times \text{MMP11} + 0.001 \times \text{PTGES} - 0.140 \times \text{PTGS2} - 0.145 \times \text{S100A8} - 0.212 \times \text{TNFSF13B}$$

La ecuación del modelo predictivo de carcinoma es:

$$\log\left(\frac{\Pr(y_i=1)}{1-\Pr(y_i=1)}\right) = -8.544 + 0.707 \times \text{CACNB4} + 0.688 \times \text{CXCL10} - 0.592 \times \text{IL1B} - 0.234 \times \text{LTF} + 0.044 \times \text{MMP11} + 0.105 \times \text{PTGES} - 0.143 \times \text{PTGS2} - 1.605 \times \text{S100A8}$$

- 5 El modelo predictivo de CRC mostró, para la detección de CRC, una especificidad del 92% y la sensibilidad de 67% cuando fue aplicado al propio conjunto de entrenamiento sin *bootstrap* y una especificidad del 84% y sensibilidad del 60% cuando se aplicó al conjunto de validación independiente. Por otro lado, el modelo predictivo de adenoma mostró, para la detección de adenoma ≥ 1 cm, una especificidad del 88% y una sensibilidad del 50% cuando se aplicó al propio conjunto de entrenamiento sin *bootstrap*, y una especificidad del 76% y una sensibilidad de 47% cuando se aplicó al conjunto de validación independiente.

Otros modelos predictivos, definidos por diferentes combinaciones de biomarcadores, junto con su precisión diagnóstica, se exponen en la Tabla 8 y en la Tabla 9.

Tabla 8. Modelos de regresión logística penalizada

Ecuación nº	1	2	3	4	5
Intercept. β_0	-0.393	4.353	2.195	5.722	4.571
IL1B	-0.068	-0.386	-0.238	-0.339	-0.379
PTGS2	0	-0.061	-0.224	-0.182	-0.080
S100A8	-0.668	0	-1.447	-1.445	-0.037
LTF	0.000	-0.120	0	-0.292	-0.131

- 15 La tabla expone los coeficientes β correspondientes que definen las ecuaciones logísticas ajustadas. La ecuación logística tiene la forma:

$$\log\left(\frac{\Pr(y_i=1)}{1-\Pr(y_i=1)}\right) = \beta_0 + \beta_1 x_{1,i} + \dots + \beta_m x_{m,i}$$

en donde $x_{m,i}$ es un valor medido para el biomarcador m y el sujeto i , y $(\beta_0, \beta_1, \dots, \beta_m)$ es un vector de coeficientes. Un coeficiente igual a 0 significa que el biomarcador no es considerado por el modelo.

Tabla 9. Sensibilidad y especificidad para la detección de CRC y adenoma (POL) de los modelos predictivos expuestos en la Tabla 8.

Ecuación n°	Combinación de biomarcadores	Conjunto de entrenamiento			Conjunto de validación		
		Sp	Sn CRC	Sn POL	Sp	Sn CRC	Sn POL
1	IL1B, S100A8	0.80	0.60	0.43	0.88	0.53	0.33
2	IL1B, PTGS2, LTF	0.94	0.47	0.35	0.80	0.27	0.33
3	IL1B, PTGS2, S100A8	0.88	0.57	0.33	0.92	0.47	0.29
4	IL1B, PTGS2, S100A8, LTF	0.92	0.60	0.30	0.80	0.47	0.33
5	IL1B, PTGS2, S100A8, LTF	0.94	0.47	0.40	0.80	0.27	0.33

5 **Listado de secuencias**

<110> Therianos , Stavros
 Ruegg , Curzio
 Monnier-Benoit , Sylvain
 Ciaroni, Laura
 Hosseinian , Sahar

10

<120> COMBINACIONES DE BIOMARCADORES PARA TUMORES COLORECTALES

<130> 36522-502001WO

<150> EP12191236.4

<151> 2012-11-05

15

<160> 64

<170> PatentIn version 3.5

<210> 1

<211> 20

<212> DNA

20

<213> Secuencia Artificial

<220>

<223> Cebador

<400> 1

acaacaacct acggcagaca 20

25

<210> 2

<211> 20

<212> DNA

<213> Secuencia Artificial

<220>

30

<223> Cebador

<400> 2

ccacagacgg taatgtggtg 20

<210> 3

<211> 21

35

<212> DNA

<213> Secuencia Artificial

<220>
 <223> Cebador

 <400> 3
 tccaagcaca gctatctct t 21

 5 <210> 4
 <211> 19
 <212> DNA
 <213> Secuencia Artificial

 <220>
 <223> Cebador

 10 <400> 4
 ccctctttca ccagccttc 19

 <210> 5
 <211> 22
 15 <212> DNA
 <213> Secuencia Artificial

 <220>
 <223> Cebador

 20 <400> 5
 agtgatttcc acagtgactc ca 22

 <210> 6
 <211> 20
 <212> DNA
 <213> Secuencia Artificial

 25 <220>
 <223> Cebador

 <400> 6
 ggcagatgct ggctactgat 20

 30 <210> 7
 <211> 24
 <212> DNA
 <213> Secuencia Artificial

 <220>
 <223> Cebador

 35 <400> 7
 gaatgaaatg tgtgaagttc ttgc 24

 <210> 8
 <211> 19
 <212> DNA
 40 <213> Secuencia Artificial

 <220>
 <223> Cebador

 <400> 8
 gcaatcagtc ccaactgcac 19

 45 <210> 9
 <211> 20
 <212> DNA
 <213> Secuencia Artificial

<220>
 <223> Cebador

 <400> 9
 caggagtttg gctggtgat 20

 5 <210> 10
 <211> 19
 <212> DNA
 <213> Secuencia Artificial

 <220>
 <223> Cebador

 10 <400> 10
 cagttgccct tcggagagt 19

 <210> 11
 <211> 23
 15 <212> DNA
 <213> Secuencia Artificial

 <220>
 <223> Cebador

 20 <400> 11
 aaaaggtatg caatcaatc tgc 23

 <210> 12
 <211> 18
 <212> DNA
 <213> Secuencia Artificial

 25 <220>
 <223> Cebador

 <400> 12
 aagaatttgg gcccttg 18

 30 <210> 13
 <211> 23
 <212> DNA
 <213> Secuencia Artificial

 <220>
 <223> Cebador

 35 <400> 13
 ttgtgtgcta cagttgtca agg 23

 <210> 14
 <211> 22
 <212> DNA
 40 <213> Secuencia Artificial

 <220>
 <223> Cebador

 <400> 14
 tctgccactt tcactgcttt ta 22

 45 <210> 15
 <211> 19
 <212> DNA
 <213> Secuencia Artificial

<220>
 <223> Cebador

 <400> 15
 accacaagca ccaaagcag 19

 5 <210> 16
 <211> 20
 <212> DNA
 <213> Secuencia Artificial

 <220>
 <223> Cebador

 10 <400> 16
 ggcgtcattt agcacttgt 20

 <210> 17
 <211> 18
 15 <212> DNA
 <213> Secuencia Artificial

 <220>
 <223> Cebador

 20 <400> 17
 agcacctgac cgcagagt 18

 <210> 18
 <211> 18
 <212> DNA
 <213> Secuencia Artificial

 25 <220>
 <223> Cebador

 <400> 18
 ggcagtcgag tggtttgg 18

 30 <210> 19
 <211> 20
 <212> DNA
 <213> Secuencia Artificial

 <220>
 <223> Cebador

 35 <400> 19
 accacgtcca gttcttcagc 20

 <210> 20
 <211> 22
 <212> DNA
 40 <213> Secuencia Artificial

 <220>
 <223> Cebador

 <400> 20
 gggctggagt tctgtgtaga ct 22

 45 <210> 21
 <211> 20
 <212> DNA
 <213> Secuencia Artificial

<220>
 <223> Cebador

 <400> 21
 cacaagatga atgggcagaa 20

 5 <210> 22
 <211> 21
 <212> DNA
 <213> Secuencia Artificial

 <220>
 <223> Cebador

 10 <400> 22
 tgacaatttg cacaacaggt g 21

 <210> 23
 <211> 20
 <212> DNA
 <213> Secuencia Artificial

 15 <220>
 <223> Cebador

 <400> 23
 agctgatggc cctaacaga 20

 <210> 24
 <211> 20
 <212> DNA
 <213> Secuencia Artificial

 20 <220>
 <223> Cebador

 <400> 24
 tcggagattc gtagctggat 20

 <210> 25
 <211> 20
 <212> DNA
 <213> Secuencia Artificial

 25 <220>
 <223> Cebador

 <400> 25
 tagccaggat ccacaagtcc 20

 <210> 26
 <211> 22
 <212> DNA
 <213> Secuencia Artificial

 30 <220>
 <223> Cebador

 <400> 26
 ctgtgaggta agatggtggc ta 22

 <210> 27
 <211> 18
 <212> DNA
 <213> Secuencia Artificial

 35 <220>
 <223> Cebador

 <400> 26
 ctgtgaggta agatggtggc ta 22

 <210> 27
 <211> 18
 <212> DNA
 <213> Secuencia Artificial

 40 <220>
 <223> Cebador

 <400> 26
 ctgtgaggta agatggtggc ta 22

 <210> 27
 <211> 18
 <212> DNA
 <213> Secuencia Artificial

 45 <220>
 <223> Cebador

 <400> 26
 ctgtgaggta agatggtggc ta 22

 <210> 27
 <211> 18
 <212> DNA
 <213> Secuencia Artificial

<220>
 <223> Cebador

 <400> 27
 aacatgagcc tcggcttg 18

 5 <210> 28
 <211> 18
 <212> DNA
 <213> Secuencia Artificial

 <220>
 10 <223> Cebador

 <400> 28
 gcccacagag gaccacat 18

 <210> 29
 <211> 19
 15 <212> DNA
 <213> Secuencia Artificial

 <220>
 <223> Cebador

 20 <400> 29
 gcatgcagca ccaagagag 19

 <210> 30
 <211> 20
 <212> DNA
 <213> Secuencia Artificial

 25 <220>
 <223> Cebador

 <400> 30
 gcaggtctgg ttgtcaggt 20

 30 <210> 31
 <211> 22
 <212> DNA
 <213> Secuencia Artificial

 <220>
 <223> Cebador

 35 <400> 31
 agtcaggcag acagacagac ac 22

 <210> 32
 <211> 26
 <212> DNA
 40 <213> Secuencia Artificial

 <220>
 <223> Cebador

 <400> 32
 aaaataagat ttgcagttcg gactat 26

 45 <210> 33
 <211> 20
 <212> DNA
 <213> Secuencia Artificial

<220>
 <223> Cebador

 <400> 33
 taaggtgaa cgctgaaac 20

 5 <210> 34
 <211> 21
 <212> DNA
 <213> Secuencia Artificial

 <220>
 <223> Cebador

 10 <400> 34
 ccatttctcc caaathtagc c 21

 <210> 35
 <211> 21
 <212> DNA
 <213> Secuencia Artificial

 15 <220>
 <223> Cebador

 <400> 35
 cgagttgtg gacttcactg c 21

 <210> 36
 <211> 20
 <212> DNA
 <213> Secuencia Artificial

 20 <220>
 <223> Cebador

 <400> 36
 aaggtgaaga aggggtgctc 20

 <210> 37
 <211> 23
 <212> DNA
 <213> Secuencia Artificial

 25 <220>
 <223> Cebador

 <400> 37
 tggatgaaac tcacagtcac att 23

 <210> 38
 <211> 19
 <212> DNA
 <213> Secuencia Artificial

 30 <220>
 <223> Cebador

 <400> 38
 ggccaatcat gctctgaaa 19

 <210> 39
 <211> 20
 <212> DNA
 <213> Secuencia Artificial

 35 <220>
 <223> Cebador

 <400> 38
 ggccaatcat gctctgaaa 19

 <210> 39
 <211> 20
 <212> DNA
 <213> Secuencia Artificial

 40 <220>
 <223> Cebador

 <400> 38
 ggccaatcat gctctgaaa 19

 <210> 39
 <211> 20
 <212> DNA
 <213> Secuencia Artificial

 45 <220>
 <223> Cebador

 <400> 38
 ggccaatcat gctctgaaa 19

 <210> 39
 <211> 20
 <212> DNA
 <213> Secuencia Artificial

<220>
 <223> Cebador

 <400> 39
 aagaggttcg tgctttctgg 20

 5 <210> 40
 <211> 18
 <212> DNA
 <213> Secuencia Artificial

 <220>
 <223> Cebador

 10 <400> 40
 ccatgggaac cgaaggat 18

 <210> 41
 <211> 19
 15 <212> DNA
 <213> Secuencia Artificial

 <220>
 <223> Cebador

 20 <400> 41
 atccggcacc tctatggtc 19

 <210> 42
 <211> 16
 <212> DNA
 <213> Secuencia Artificial

 25 <220>
 <223> Cebador

 <400> 42
 cagaccgtcg ggggag 16

 30 <210> 43
 <211> 20
 <212> DNA
 <213> Secuencia Artificial

 <220>
 <223> Cebador

 35 <400> 43
 caggccaagg aaaaggagat 20

 <210> 44
 <211> 18
 <212> DNA
 40 <213> Secuencia Artificial

 <220>
 <223> Cebador

 <400> 44
 cgttcaatcc gagcaagg 18

 45 <210> 45
 <211> 20
 <212> DNA
 <213> Secuencia Artificial

<220>
 <223> Cebador

 <400> 45
 cctaagcagc tggaaggaac 20

 5 <210> 46
 <211> 21
 <212> DNA
 <213> Secuencia Artificial

 <220>
 <223> Cebador

 10 <400> 46
 cgcttgataa tctctccac a 21

 <210> 47
 <211> 24
 15 <212> DNA
 <213> Secuencia Artificial

 <220>
 <223> Cebador

 20 <400> 47
 gacaggaag acaacagaca aatc 24

 <210> 48
 <211> 21
 <212> DNA
 <213> Secuencia Artificial

 25 <220>
 <223> Cebador

 <400> 48
 ggggtgatgt gtttgaactt g 21

 30 <210> 49
 <211> 18
 <212> DNA
 <213> Secuencia Artificial

 <220>
 <223> Cebador

 35 <400> 49
 agaaggcctt tgccaacc 18

 <210> 50
 <211> 20
 <212> DNA
 40 <213> Secuencia Artificial

 <220>
 <223> Cebador

 <400> 50
 gatggtctcc atgctgtcc 20

 45 <210> 51
 <211> 18
 <212> DNA
 <213> Secuencia Artificial

<220>
 <223> Cebador

 <400> 51
 cgctcagcca tacagcaa 18

 5 <210> 52
 <211> 23
 <212> DNA
 <213> Secuencia Artificial

 <220>
 <223> Cebador

 10 <400> 52
 tcatacatat acctcggttt tga 23

 <210> 53
 <211> 20
 <212> DNA
 <213> Secuencia Artificial

 15 <220>
 <223> Cebador

 <400> 53
 agcacaccag gagagagctg 20

 <210> 54
 <211> 20
 <212> DNA
 <213> Secuencia Artificial

 20 <220>
 <223> Cebador

 <400> 54
 gtagccaaag gcactgatcc 20

 <210> 55
 <211> 22
 <212> DNA
 <213> Secuencia Artificial

 25 <220>
 <223> Cebador

 <400> 55
 cagctgtctt tcagaagacc tg 22

 <210> 56
 <211> 20
 <212> DNA
 <213> Secuencia Artificial

 30 <220>
 <223> Cebador

 <400> 56
 ctttctccag ctcggtcaac 20

 <210> 57
 <211> 21
 <212> DNA
 <213> Secuencia Artificial

 35 <220>
 <223> Cebador

 <400> 57
 ctttctccag ctcggtcaac 20

 <210> 58
 <211> 21
 <212> DNA
 <213> Secuencia Artificial

 40 <220>
 <223> Cebador

 <400> 58
 ctttctccag ctcggtcaac 20

 <210> 59
 <211> 21
 <212> DNA
 <213> Secuencia Artificial

 45 <220>
 <223> Cebador

 <400> 59
 ctttctccag ctcggtcaac 20

 <210> 60
 <211> 21
 <212> DNA
 <213> Secuencia Artificial

<220>
 <223> Cebador

 <400> 57
 ctcaagactg cttgcaactg a 21

 5 <210> 58
 <211> 20
 <212> DNA
 <213> Secuencia Artificial

 <220>
 <223> Cebador

 10 <400> 58
 aagctgagaa gccatggaac 20

 <210> 59
 <211> 20
 <212> DNA
 <213> Secuencia Artificial

 15 <220>
 <223> Cebador

 <400> 59
 20 tgctacagag caggagtgc 20

 <210> 60
 <211> 21
 <212> DNA
 <213> Secuencia Artificial

 25 <220>
 <223> Cebador

 <400> 60
 tcctgttctt caagctctgg t 21

 30 <210> 61
 <211> 20
 <212> DNA
 <213> Secuencia Artificial

 <220>
 <223> Cebador

 35 <400> 61
 tcgacaatgg cagcatctac 20

 <210> 62
 <211> 20
 <212> DNA
 40 <213> Secuencia Artificial

 <220>
 <223> Cebador

 <400> 62
 gccaatctgc agacagacac 20

 45 <210> 63
 <211> 23
 <212> DNA
 <213> Secuencia Artificial

<220>

<223> Cebador

<400> 63

caatcaaagg gaaactgaa gaa 23

5

<210> 64

<211> 26

<212> DNA

<213> Secuencia Artificial

10

<220>

<223> Cebador

<400> 64

gattcatggt ttcaccaata aagaac 26

REIVINDICACIONES

1. Un método para determinar la probabilidad de un tumor colorrectal en un sujeto, que comprende:
 - (a) medir en una muestra obtenida del sujeto la cantidad de cada biomarcador de un panel de biomarcadores que comprende IL1B, PTGS2, y al menos un biomarcador nuclear que se elige entre el grupo que consiste en S100A8, LTF y CACNB4;
 - (b) calcular una puntuación de probabilidad basándose en la medición de la etapa (a); y
 - (c) descartar un tumor colorrectal para el sujeto si la puntuación en la etapa (b) es inferior a una puntuación predeterminada; o
 - (d) admitir la probabilidad de tumor colorrectal para el sujeto si la puntuación en la etapa (b) es superior a una puntuación predeterminada.
2. Un método para clasificar a un sujeto como más probable de tener un tumor colorrectal, que comprende:
 - (a) medir en una muestra obtenida del sujeto la cantidad de cada biomarcador de un panel de biomarcadores que comprende IL-1B, PTGS2 y al menos un biomarcador nuclear que se elige entre el grupo que consiste en S100A8, LTF y CACNB4;
 - (b) comparar la cantidad medida en la etapa (a) con un valor de referencia; y
 - (c) clasificar al sujeto como más probable de tener tumor colorrectal cuando se detecta en la etapa (b) un aumento o una disminución en la cantidad de cada biomarcador del panel de biomarcadores respecto al valor de referencia.
3. El método según la reivindicación 2, en el que al sujeto clasificado por la etapa (c) se le administra además una cantidad terapéuticamente eficaz de al menos un agente modulador colorrectal.
4. El método según la reivindicación 1 o 2, en el que dicho panel de biomarcadores además comprende al menos un biomarcador que se elige entre el grupo que consiste en S100A8, LTF, CXCL10, CACNB4, MMP9, CXCL11, EGR1, JUN, TNFSF13B, GATA2, MMP11, NME1, PTGES, CCR1, CXCR3, FXYD5, IL8, ITGA2, ITGB5, MAPK6, RHOC, BCL3, CD63, CES1, MAP2K3, MSL1 y PPARg.
5. El método según la reivindicación 1 o 2, en el que dicho al menos un biomarcador nuclear es:
 - (a) S100A8;
 - (b) S100A8 y LTF;
 - (c) o
 - (d) S100A8, LTF y CACNB4.
6. El método según la reivindicación 1, en el que cuando se descarta el tumor colorrectal el sujeto no recibe un protocolo de tratamiento.
7. El método según la reivindicación 1, en el que cuando se admite un tumor colorrectal el sujeto recibe un protocolo de tratamiento.
8. El método según las reivindicaciones 6 o 7, en el que dicho protocolo de tratamiento es una colonoscopia, una biopsia, una cirugía, una quimioterapia, una radioterapia o cualquier combinación de las mismas.
9. El método según la reivindicación 1, en el que dicha puntuación de probabilidad se calcula a partir de un modelo de predicción de regresión logística aplicado a la medición.
10. El método según las reivindicaciones 1 o 2, en el que dicha muestra se elige entre el grupo que consiste en células mononucleares de sangre periférica, células sanguíneas, sangre entera, suero, plasma, células endoteliales, células tumorales circulantes, biopsias de tejido, fluido linfático, líquido ascítico, fluido intersticial, médula ósea, líquido cefalorraquídeo (CSF), saliva, mucosa, esputo, sudor y orina.
11. El método según las reivindicaciones 1 o 2, en el que dicho tumor colorrectal es adenoma o carcinoma.
12. El método según la reivindicación 1, en el que dicha probabilidad de tumor colorrectal se determina además por la sensibilidad, especificidad, valor predictivo negativo (NPV) o valor predictivo positivo (PPV) asociado con la puntuación.
13. El método según las reivindicaciones 1 o 2, en el que dicho sujeto está en riesgo de desarrollar un tumor colorrectal.

14. El método según las reivindicaciones 1 o 2, en el que dicho panel de biomarcadores comprende además CXCL10.

5 15. El método según la reivindicación 14, en el que dicho panel de biomarcadores comprende además al menos un biomarcador que se elige entre el grupo que consiste en MMP9, CXCL10, EGR1, JUN, TNFSF13B, GATA2, MMP11, NME1, PTGES, CCR1, CXCR3, FXD5, IL8, ITGA2, ITGB5, MAPK6, RHOC, BCL3, CD63, CES1, MAP2K3, MSL1 y PPARg.

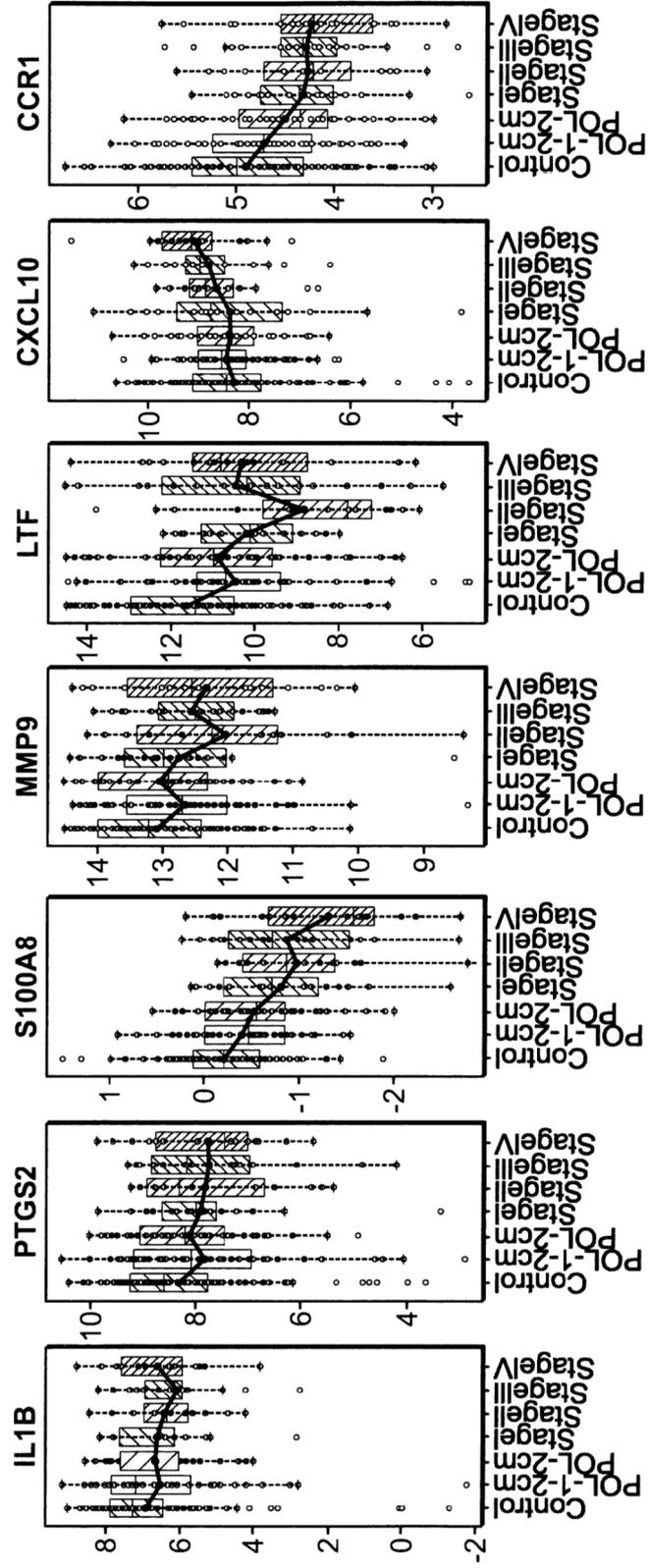


FIG. 1

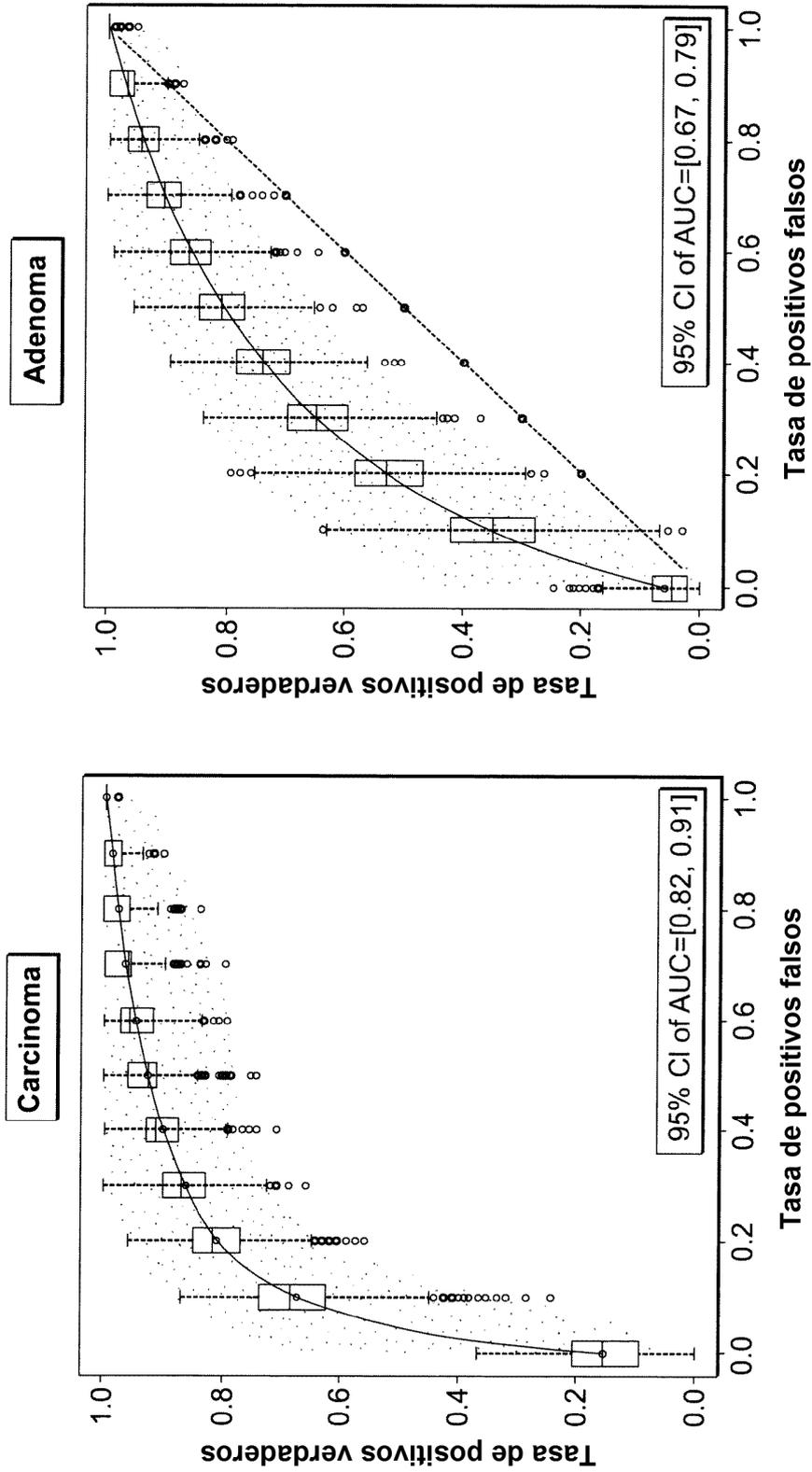


FIG. 2

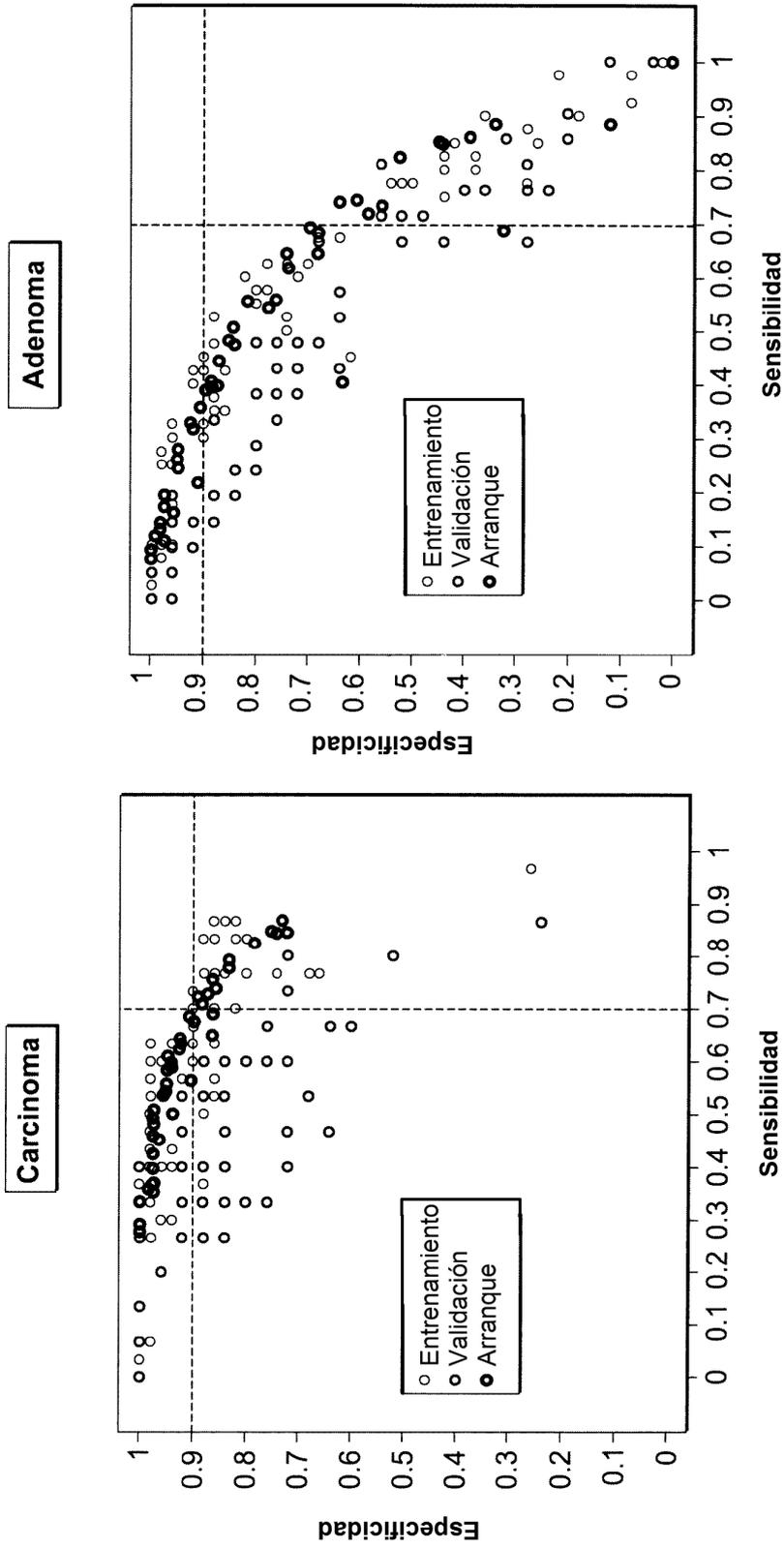


FIG. 3