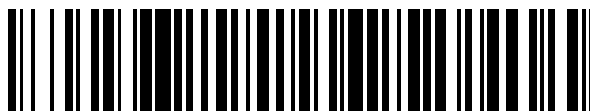


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 729 554**

51 Int. Cl.:

C12Q 1/6886 (2008.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **19.10.2012 PCT/US2012/061044**

87 Fecha y número de publicación internacional: **10.05.2013 WO13066641**

96 Fecha de presentación y número de la solicitud europea: **19.10.2012 E 12846333 (8)**

97 Fecha y número de publicación de la concesión europea: **20.03.2019 EP 2768985**

54 Título: **Biomarcadores de ácidos nucleicos en circulación asociados al cáncer colorrectal**

30 Prioridad:

21.10.2011 US 201161550098 P

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

04.11.2019

73 Titular/es:

**CHRONIX BIOMEDICAL (100.0%)
5941 Optical Court, Suite 203E
San Jose, California 95138, US**

72 Inventor/es:

**SCHUTZ, EKKEHARD;
BECK, JULIA y
URNOVITZ, HOWARD**

74 Agente/Representante:

PONS ARIÑO, Ángel

Observaciones :

Véase nota informativa (Remarks, Remarques o Bemerkungen) en el folleto original publicado por la Oficina Europea de Patentes

ES 2 729 554 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

Biomarcadores de ácidos nucleicos en circulación asociados al cáncer colorrectal

5 **Referencia cruzada a solicitudes relacionadas**

La presente solicitud reivindica el beneficio de prioridad de la solicitud provisional de Estados Unidos N.º 61/550.098, presentada el viernes 21 de octubre de 2011.

10 **Antecedentes de la invención**

El cáncer colorrectal es el tercer diagnóstico de cáncer más común en Estados Unidos y la segunda causa principal de muertes relacionadas con el cáncer. Existen métodos para detectar el cáncer colorrectal, incluidas las pruebas de colonoscopia y heces, sin embargo, estos métodos de prueba presentan varios inconvenientes (véase, por ejemplo, 15 McFarland et al., *Radiology* 248:717-720, 2008). El documento WO 2011/011426 A2 informa sobre métodos y biomarcadores para evaluar el riesgo de un sujeto de contraer una enfermedad, tal como cáncer, una enfermedad autoinmune o una enfermedad neurológica. El documento US 2010/136560 A1 informa sobre análisis integrados de 20 cáncer de mama y colorrectal. Boni et al (*Surgical Oncology* 16: 29-31 (2007)) informan que el ADN circulante libre es un posible marcador tumoral en el cáncer colorrectal. Lecomte et al (*International Journal of Cancer* 100(5): 542-548 (2002)) informan sobre la detección de ADN asociado a tumor en circulación libre en plasma de pacientes con cáncer colorrectal y su asociación con el pronóstico. Existe la necesidad de métodos de detección de cáncer colorrectal eficientes. Esta invención aborda esa necesidad.

25 **Breve resumen de la invención**

La invención reivindicada proporciona un método para el diagnóstico o el cribado del cáncer colorrectal en un paciente, que comprende detectar, en una muestra que es sangre, suero o plasma de dicho paciente,

30 (a) el nivel total de todos los ADN extracelulares en circulación, cada uno con una secuencia que está dentro de la misma región cromosómica única designada como "ASCENDENTE" en la Tabla 2; y correlacionar un nivel total aumentado con una probabilidad mayor de que dicho paciente tenga cáncer colorrectal cuando el nivel total del ADN extracelular que tiene una secuencia que está dentro de la misma región cromosómica única es al menos 1, 2, 3, 4, 5, 10, 15, o 20 desviaciones estándar mayor que un valor de índice de sujetos normales; o

35 (b) el nivel total de todos los ADN extracelulares en circulación, cada uno con una secuencia que está dentro de la misma región cromosómica única designada como "DESCENDENTE" en la Tabla 2; y correlacionar un nivel total disminuido con una probabilidad mayor de que dicho paciente tenga cáncer colorrectal cuando el nivel total del ADN extracelular que tiene una secuencia que está dentro de la misma región cromosómica única es al menos 1, 2, 3, 4, 5, 10, o 20 desviaciones estándar menor que un valor de índice de sujetos normales;

40 en el que las posiciones de los nucleótidos en las regiones cromosómicas en la Tabla 2 están numeradas según la versión del genoma del National Center for Biotechnology Information human genome hg18/versión 36.1 publicada en marzo de 2006. La invención se basa, en parte, en el descubrimiento de biomarcadores de ácidos nucleicos en circulación (CNA) asociados al cáncer colorrectal. En algunas realizaciones, los biomarcadores de CNA son fragmentos de polinucleótidos, por ejemplo, fragmentos de ADN, que están presentes a un nivel 45 elevado en la sangre, por ejemplo, en una muestra de suero o plasma, de un paciente con cáncer colorrectal en comparación con el nivel en sangre, por ejemplo, una muestra de suero o plasma, obtenida de un individuo normal que no tiene cáncer colorrectal. En algunas realizaciones, los biomarcadores de CNA son secuencias de polinucleótidos de ADN, es decir, fragmentos de ADN que están presentes en la sangre, por ejemplo, en una muestra de suero o plasma, a un nivel reducido de un paciente con cáncer colorrectal en comparación con el nivel en sangre, por ejemplo, suero o plasma, de un individuo normal que no tiene cáncer colorrectal.

50 Por consiguiente, en un aspecto, la divulgación proporciona un método para analizar CNA en una muestra (sangre, suero o plasma) de un paciente, que comprende detectar el nivel de al menos un ADN extracelular que tenga una secuencia de nucleótidos que se encuentre dentro de una región cromosómica expuesta en la Tabla 2 en la muestra. En algunos aspectos, detectar el nivel de al menos un biomarcador comprende detectar una molécula de ADN 55 extracelular que tiene entre al menos 20 y al menos 500 nucleótidos consecutivos, o, por ejemplo, entre al menos 50 y al menos 400 nucleótidos consecutivos de una única secuencia dentro de una región cromosómica como se expone en la Tabla 2.

60 En un aspecto, se describe un método para analizar ADN libre en circulación en una muestra de un paciente, que comprende determinar, en una muestra que es sangre, suero o plasma, el nivel de al menos 2, 3, 4, 5, 7, 8, 9, 10, 15, 20, 30, 40, 45, 50, 55, 60, 65, 70, 75, 80 u 81 moléculas de ADN extracelular, cada una de las cuales tiene una secuencia que se encuentra dentro de una región cromosómica diferente como se muestra en la Tabla 2, y preferiblemente las secuencias de las moléculas de ADN extracelular están libres de elementos repetitivos.

65 En otro aspecto, la presente divulgación proporciona un kit que incluye dos o más (por ejemplo, al menos 2, 3, 4, 5, 7, 8, 9, 10, 15, 20, 30, 40, 45, 50, 55, 60, 65, 70, 75, 80, u 81) conjuntos de oligonucleótidos. En algunos aspectos, el

kit incluye 82 o menos conjuntos de oligonucleótidos. Cada conjunto comprende uno o más oligonucleótidos con una secuencia de nucleótidos que se encuentra dentro de una única región cromosómica que se expone en la Tabla 2. Preferiblemente, los distintos conjuntos de oligonucleótidos corresponden a distintas regiones cromosómicas dentro la Tabla 2. Preferiblemente, los oligonucleótidos carecen de elementos repetitivos. Opcionalmente, los oligonucleótidos se unen a uno o más sustratos sólidos, tales como *microchips* y perlas.

En otro aspecto, la presente divulgación proporciona un método para el diagnóstico o el cribado del cáncer colorrectal en un paciente. El método incluye las etapas de: (a) detectar, en una muestra que es sangre, suero o plasma de un paciente, el nivel de al menos 2, 3, 4, 5, 7, 8, 9, 10, 15, 20, 30, 40, 45, 50, 55, 60, 65, 70, 75, 80 u 81 de las moléculas de ADN extracelular, teniendo cada una una secuencia que se encuentra dentro de una región cromosómica distinta, como se expone en la Tabla 2; y (b) correlacionar el nivel de dicho primer y segundo ADN extracelular con una mayor probabilidad de que el paciente tenga cáncer colorrectal. Preferiblemente, las secuencias de las moléculas de ADN extracelular carecen de elementos repetitivos.

En un aspecto, la divulgación proporciona un método para identificar a un paciente que tiene un biomarcador CNA asociado con cáncer colorrectal, comprendiendo el método detectar un aumento en el nivel, con respecto a lo normal, de al menos un biomarcador designado como "ASCENDENTE" en la Tabla 2 en una muestra de CNA obtenida de suero o plasma del paciente. Un biomarcador puede identificarse utilizando varios de métodos, incluyendo la secuenciación del CNA, así como el uso de una sonda o un conjunto de sondas para detectar la presencia del biomarcador.

En algunos aspectos, la divulgación proporciona un método para identificar a un paciente que tiene un biomarcador CNA asociado con cáncer colorrectal, comprendiendo el método detectar una disminución en el nivel, con respecto a lo normal, de al menos un biomarcador designado como "DESCENDENTE" en la Tabla 2 en una muestra de CNA de suero o plasma del paciente. Un biomarcador puede identificarse utilizando varios de métodos, incluyendo la secuenciación del CNA, así como el uso de una sonda o un conjunto de sondas para detectar la presencia del biomarcador.

En un aspecto adicional, la divulgación proporciona un kit para identificar a un paciente que tiene un biomarcador para cáncer colorrectal, en el que el kit comprende al menos una sonda polinucleotídica para un biomarcador expuesto en la Tabla 2. Preferiblemente, tal kit comprende sondas para múltiples biomarcadores, por ejemplo, al menos 2, 3, 4, 5, 10, 20, 30, 40, 50, 55, 60, 65, 70, 75, 80, o los 81, de los biomarcadores expuestos en la Tabla 2. En algunos aspectos, el kit también incluye un dispositivo electrónico o software informático para comparar los patrones de hibridación de la CNA en la muestra del paciente con un conjunto de datos de cáncer colorrectal que comprende una lista de los niveles de biomarcadores en pacientes con cáncer colorrectal en comparación con los individuos normales.

En algunas realizaciones, el nivel de al menos un biomarcador en el CNA se determina mediante secuenciación. En algunas realizaciones, el nivel de al menos un biomarcador en el CNA se determina mediante una matriz. En algunas realizaciones, el nivel de al menos un biomarcador en el CNA se determina utilizando un ensayo que comprende una reacción de amplificación, tal como la reacción en cadena de la polimerasa (PCR). En algunas realizaciones, se emplea una matriz de ácidos nucleicos que forma un conjunto de sondas que comprende sondas para dos o más regiones cromosómicas expuestas en las Tablas 2. En algunas realizaciones, se emplea una matriz de ácidos nucleicos que forma un conjunto de sondas que comprende 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, o las 81 regiones cromosómicas, expuestas en la Tabla 2.

En un aspecto adicional, la divulgación proporciona un método de detección de cáncer colorrectal en un paciente que tiene, o se sospecha que tiene, cáncer colorrectal, comprendiendo el método poner en contacto ADN de la muestra de suero o plasma con una sonda que hibrida de forma selectiva con una secuencia, por ejemplo, de al menos 15, 20, 25, 50, 100, o 500, o más, nucleótidos de longitud presente en una región cromosómica expuesta en la Tabla 2 en condiciones en las que la sonda se hibrida selectivamente con la secuencia; y detectar el nivel de hibridación de la sonda, en la que el nivel de hibridación con la secuencia es indicativo de cáncer colorrectal.

Breve descripción de los dibujos

La figura 1 proporciona un ejemplo de curvas ROC para las diversas combinaciones de regiones. Panel A: Normalización global de todas las regiones (46) AUC: 0,95 0,88 - 0,99; Panel B: Normalización local de todas las regiones (35) AUC: 0,93 0,86 - 0,98; Panel C: Dirección Ascendente (35) AUC: 0,93 0,87 - 0,97; Panel D: Dirección descendente (46) AUC: 0,92 0,84 - 0,97.

Descripción detallada de la invención

Como se usa en el presente documento, un "biomarcador" se refiere a una secuencia de ácido nucleico que corresponde a una región cromosómica, donde el nivel del ácido nucleico en CNA con respecto a lo normal está asociado con el cáncer colorrectal. En algunas realizaciones, en las que un biomarcador se indica como "ASCENDENTE" en la Tabla 2, el nivel en CNA de un paciente con cáncer colorrectal aumenta con respecto a lo

normal. En algunas realizaciones, en las que un biomarcador se indica como "DESCENDENTE" en la Tabla 2, el nivel en CNA de un paciente con cáncer colorrectal disminuye con respecto a lo normal.

5 En la presente invención, una "región cromosómica" enumerada en la Tabla 2 se refiere a la región del cromosoma que corresponde a las posiciones de nucleótido indicadas en las tablas. Las posiciones de nucleótido en los cromosomas están numeradas de acuerdo con el genoma de *Homo sapiens* (humano), versión del genoma hg18/versión 36.1 publicada en marzo de 2006. Como se entiende en la técnica, en el genoma de los individuos existen polimorfismos de origen natural. Por lo tanto, cada región cromosómica enumerada en la Tabla 2 abarca
10 variantes alélicas, así como la secuencia particular en la base de datos. Una variante alélica normalmente tiene al menos el 95 % de identidad, a menudo al menos el 96 %, al menos el 97 %, al menos el 98 % o al menos el 99 % de identidad de secuencia con una región cromosómica que está presente en una base de datos particular, por ejemplo, el National Center for Biotechnology Information (*Homo sapiens* versión 36.1 en la dirección de Internet www.ncbi.nlm.nih.gov/mapview/). El porcentaje de identidad se puede determinar utilizando algoritmos bien conocidos, incluyendo el algoritmo BLAST, por ejemplo, configurado a los parámetros por defecto. Además, se
15 entiende que las secuencias de nucleótidos de los cromosomas pueden mejorarse cuando que se descubren y corrigen errores en la base de datos actual. La expresión "región cromosómica" abarca cualquier variante o versión corregida de la misma región, como se define en la Tabla 2. Dada la información proporcionada en la Tabla 2 en la presente divulgación y las bases de datos de genoma disponibles, un experto en la técnica podrá entender las regiones cromosómicas usadas para la presente invención, incluso después de descubrir nuevas variantes o de
20 corregir errores.

"Detectar una región cromosómica" en CNA en el contexto de la presente invención se refiere a detectar el nivel de cualquier secuencia de una región cromosómica mostrada en la Tabla 2, donde la secuencia detectada puede asignarse de forma inequívoca a esa región cromosómica. Por lo tanto, esta expresión se refiere a la detección de
25 secuencias singulares de las regiones cromosómicas. En la presente invención, el nivel de al menos una región, típicamente múltiples regiones utilizadas en combinación, en una muestra de CNA se compara con el rango encontrado para dicha región en un grupo de individuos "normales", es decir, en el contexto de la presente invención, individuos que no tienen cáncer o al menos no han sido diagnosticadas con cáncer. Para las regiones que aumentan de nivel en los pacientes con cáncer colorrectal, es decir, las regiones enumeradas como ASCENDENTE en la Tabla 2, un resultado típicamente se considera un aumento si el resultado para la muestra es mayor del 60, 70,
30 75, 80, 85, 90, 95, o 99 por ciento. Para las regiones que disminuyen de nivel en los pacientes con cáncer colorrectal, es decir, las regiones enumeradas como DESCENDENTE en la Tabla 2, un resultado típicamente se considera que disminuye si el resultado para la muestra es inferior al 40, 30, 25, 20, 15, 10, 5, o el 1 por ciento en individuos normales. Los métodos para eliminar del análisis las secuencias repetitivas son conocidos en la técnica e incluyen el
35 uso de ADN bloqueante, por ejemplo, cuando los ácidos nucleicos diana se identifican por hibridación. En algunas realizaciones, típicamente, cuando la presencia de un biomarcador de cáncer colorrectal se determina mediante secuenciación del CNA de un paciente, se pueden usar programas informáticos y manipulaciones que ya se conocen para eliminar secuencias repetitivas del análisis (véase, por ejemplo, la sección EJEMPLOS). Además, las secuencias que tienen un alineamiento múltiple igual de adecuado con la base de datos de referencia normalmente se omiten en análisis posteriores.
40

El término "detectar un biomarcador" como se usa en el presente documento se refiere a detectar un polinucleótido, por ejemplo, ADN, de una región cromosómica enumerada en la Tabla 2 en el CNA. Como se usa en el presente documento, "detectar el nivel" de un biomarcador abarca medidas cuantitativas, así como detectar la presencia, o
45 ausencia, del biomarcador. Por lo tanto, por ejemplo, el término "detectar un aumento en el nivel de" un biomarcador, con respecto a lo normal, incluye realizaciones cualitativas en las que el biomarcador se detecta en una muestra de paciente, pero no una muestra normal. De forma similar, el término "detectar una disminución en el nivel de" un biomarcador, con respecto a lo normal, incluye realizaciones en las que el biomarcador no se detecta en una muestra de paciente, pero se detecta en muestras normales. Se considera que un biomarcador está "presente" si
50 alguna secuencia de ácido nucleico en el CNA se asigna de forma inequívoca a la región cromosómica.

La expresión "asignado de forma inequívoca" en el contexto de la presente invención se refiere a determinar que un ADN detectado en el CNA de un paciente es de una región cromosómica particular. Por lo tanto, en los métodos de detección que emplean la hibridación, la sonda hibrida específicamente con esa región. En los métodos de detección
55 que emplean amplificación, el cebador (o cebadores) hibrida específicamente con esa región. En los métodos de detección que emplean secuenciación, la secuencia se asigna a esa región basándose en algoritmos para la identidad muy conocidos, tal como el algoritmo BLAST que utiliza parámetros altamente rigurosos, tales como $e < 0,0001$. Además, tal secuencia no tiene otro resultado positivo igual de adecuado en la base de datos utilizada.

60 La expresión "ácidos nucleicos en circulación" se refiere a los ácidos nucleicos acelulares que están presentes en la sangre.

La expresión "ADN extracelular en circulación", como se usa en el presente documento, significa moléculas de ADN libres de 25 nucleótidos o más que no están contenidas dentro de ninguna célula intacta en la sangre humana, y que
65 se pueden obtener a partir de suero o plasma humano.

El término "hibridación" se refiere a la formación de una estructura bicatenaria por parte de dos ácidos nucleicos monocatenarios debido al emparejamiento de bases complementarias. La hibridación puede producirse entre cadenas de ácido nucleico exactamente complementarias o entre cadenas de ácido nucleico que contienen regiones menores de desapareamiento. Como se usa en el presente documento, la expresión "sustancialmente complementarias" se refiere a secuencias que son complementarias, excepto por regiones menores de desapareamiento. Típicamente, el número total de nucleótidos desapareados a lo largo de una región de hibridación no es de más de 3 nucleótidos para secuencias de aproximadamente 15 nucleótidos de longitud. Las condiciones en las que solo se hibridarán cadenas de ácido nucleico exactamente complementarias se denominan condiciones de hibridación "rigurosas" o "específicas de secuencia". Se pueden lograr estructuras bicatenarias estables de ácidos nucleicos sustancialmente complementarios en condiciones de hibridación menos rigurosas. Los expertos en la técnica de la tecnología de ácidos nucleicos pueden determinar la estabilidad de la estructura bicatenaria de forma empírica, considerando una serie de variables que incluyen, por ejemplo, la longitud y la concentración de pares de bases de los oligonucleótidos, la fuerza iónica y la incidencia de pares de bases desapareadas. Por ejemplo, está disponible en el mercado un programa informático de National Biosciences para calcular la estabilidad de la estructura secundaria, Inc. (Plymouth, Minn.); por ejemplo, OLIGO versión 5, de DNA Software (Ann Arbor, Michigan), por ejemplo, Visual OMP 6.

Las condiciones de hibridación rigurosas, específicas de secuencia, en las que un oligonucleótido hibridará solo con la secuencia diana, se conocen bien en la técnica (véanse, por ejemplo, las referencias generales proporcionadas en la sección sobre detección de polimorfismos en secuencias de ácido nucleico). Las condiciones rigurosas dependen de la secuencia y serán distintas en circunstancias distintas. Generalmente, las condiciones rigurosas se seleccionan para que sean aproximadamente de 5 °C más bajas a 5 °C más altas que el punto de fusión térmica (T_m) para la secuencia específica a una fuerza iónica y pH definidos. La T_m es la temperatura (a una fuerza iónica y pH definidos) a la que se ha disociado el 50 % de las cadenas de la estructura bicatenaria. Suavizar la rigurosidad de las condiciones de hibridación permitirá tolerar desapareamientos de secuencia; el grado de desapareamiento tolerado puede controlarse mediante el ajuste adecuado de las condiciones de hibridación.

El término "cebador" se refiere a un oligonucleótido que actúa como un punto de inicio de la síntesis de ADN en condiciones en las que se induce la síntesis de un producto de extensión de cebador complementario a una cadena de ácido nucleico, es decir, en presencia de cuatro nucleósido trifosfatos distintos y un agente para la polimerización (es decir, ADN polimerasa o transcriptasa inversa) en un tampón apropiado y a una temperatura adecuada. Un cebador es preferiblemente un oligodesoxirribonucleótido monocatenario. El cebador incluye una "región de hibridación" exacta o sustancialmente complementaria a la secuencia diana, preferiblemente, de aproximadamente 15 o aproximadamente 35 nucleótidos de longitud. Un oligonucleótido cebador puede consistir completamente en la región de hibridación o puede contener características adicionales que permitan la detección, inmovilización o manipulación del producto amplificado, pero que no alteran la capacidad del cebador para servir como reactivo de partida para la síntesis de ADN. Por ejemplo, puede incluirse una cola de una secuencia de ácido nucleico en el extremo 5' del cebador que hibride con un oligonucleótido de captura.

El término "sonda" se refiere a un oligonucleótido que hibrida de forma selectiva con un ácido nucleico diana en condiciones adecuadas. Una sonda para la detección de las secuencias de biomarcador descritas en el presente documento puede ser de cualquier longitud, por ejemplo, de 15-500 pb de longitud. Típicamente, en ensayos basados en sondas, son preferentes las sondas de hibridación que son de menos de 50 pb.

La expresión "secuencia diana" o "región diana" se refiere a una región de un ácido nucleico que se debe analizar y comprende la secuencia de interés.

Como se usa en el presente documento, los términos "ácido nucleico", "polinucleótido" y "oligonucleótido" se refieren a cebadores, sondas y fragmentos oligoméricos. Los términos no están limitados por la longitud y son genéricos para los polímeros lineales de polidesoxirribonucleótidos (que contienen 2-desoxi-D-ribosa), polirribonucleótidos (que contienen D-ribosa) y cualquier otro N-glucósido de una base purínica o pirimidínica, o bases purínicas o pirimidínicas modificadas. Estos términos incluyen ADN bicatenario y monocatenario, así como ARN bicatenario y monocatenario. Los oligonucleótidos para su uso en la invención se pueden usar como cebadores y/o sondas.

Un ácido nucleico, polinucleótido u oligonucleótido puede comprender enlaces fosfodiéster o enlaces modificados que incluyen, pero sin limitación, un fosfotriéster, fosforamidato, siloxano, carbonato, carboximetiléster, acetamidato, carbamato, tioéter, fosforamidato puenteado, metilfosfonato puenteado, fosforitoato, metilfosfonato, fosforoditioato, fosforotioato puenteado o enlaces sulfona, y combinaciones de tales enlaces.

Un ácido nucleico, polinucleótido u oligonucleótido puede comprender las cinco bases de origen biológico (adenina, guanina, timina, citosina y uracilo) y/o bases distintas de las cinco bases de origen biológico. Estas bases pueden servir para varios propósitos, por ejemplo, para estabilizar o desestabilizar la hibridación; para facilitar o inhibir la degradación de la sonda; o como puntos de unión para fracciones detectables o fracciones de desactivación. Por ejemplo, un polinucleótido de la invención puede contener una o más fracciones de bases modificadas, no convencionales o derivatizadas, incluyendo, pero sin limitación, N6-metiladenina, N6-*terc*-butilbencil-adenina, imidazol, imidazoles sustituidos, 5-fluorouracilo, 5 bromouracilo, 5-clorouracilo, 5-yodouracilo, hipoxantina, xantina,

4-acetilcitosina, 5 (carboxihidroximetil)uracilo, 5 carboximetilaminometil-2-tiouridina, 5 carboximetilaminometiluracilo, dihidrouracilo, beta-D-galactosilqueosina, inosina, N6 isopenteniladenina, 1-metilguanina, 1-metilinosina, 2,2-dimetilguanina, 2-metiladenina, 2-metilguanina, 3-metilcitosina, 5-metilcitosina, N6-metiladenina, 7-metilguanina, 5-metilaminometiluracilo, 5-metoxiaminometil-2-tiouracilo, beta-D mannosilqueosina, 5'-metoxicarboximetiluracilo, 5-metoxiuracilo, 2-metil-N6-isopenteniladenina, ácido uracil-5-oxiacético (v), wybutoxosina, pseudouracilo, queosina, 2 tiocitosina, 5-metil-2-tiouracilo, 2-tiouracilo, 4-tiouracilo, 5-metiluracilo, éster metílico del ácido uracil-5-oxiacético, 3-(3-amino-3-N-2-carboxipropil) uracilo, (acp3)w, 2,6-diaminopurina y 5-propinil pirimidina. Otros ejemplos de fracciones de bases modificadas, no convencionales o derivatizadas, se pueden encontrar en las patentes de Estados Unidos N.º 6.001.611; 5.955.589; 5.844.106; 5.789.562; 5.750.343; 5.728.525; y 5.679.785. Además, un ácido nucleico, polinucleótido u oligonucleótido puede comprender una o más fracciones de azúcar modificadas que incluyen, pero sin limitación, arabinosa, 2-fluoroarabinosa, xilulosa y hexosa.

La expresión "elemento repetitivo", como se usa en el presente documento, se refiere a un tramo de secuencia de ADN de al menos 25 nucleótidos de longitud que está presente en el genoma humano en al menos 50 copias.

Los términos "matrices", "micromatrices" y los "chips de ADN" se usan en el presente documento de manera indistinta para referirse a una matriz de polinucleótidos distintos fijados a un sustrato, tales como vidrio, plástico, papel, nylon u otro tipo de membrana, filtro, chip, perla o cualquier otro soporte sólido adecuado. Los polinucleótidos se pueden sintetizar directamente sobre el sustrato o se pueden sintetizar por separado del sustrato y después fijarse al sustrato. Las matrices se preparan utilizando métodos conocidos.

Introducción

La invención se basa, al menos en parte, en la identificación de biomarcadores de ácido nucleico en CNA que tienen secuencias de regiones cromosómicas particulares que están presentes en un nivel aumentado, con respecto a lo normal, en la sangre de pacientes que tienen cáncer colorrectal. La invención también se basa, en parte, en la identificación de biomarcadores en el CNA que están presentes en un nivel disminuido, con respecto a lo normal, en la sangre de pacientes que tienen cáncer colorrectal. Por lo tanto, la invención proporciona métodos para analizar la presencia y el nivel en CNA de moléculas de polinucleótidos de una región cromosómica correspondiente al menos a una de las regiones cromosómicas expuestas en la Tabla 2. La divulgación proporciona además dispositivos para analizar la presencia y el nivel en CNA de moléculas de polinucleótidos de una región cromosómica correspondiente al menos a una de las regiones cromosómicas expuestas en la Tabla 2.

Por consiguiente, en un aspecto, la divulgación proporciona un método para analizar CNA en una muestra (sangre, suero o plasma) de un paciente, que comprende detectar un nivel de al menos un ADN extracelular en circulación que tenga una secuencia de al menos 25 nucleótidos que se encuentre dentro de una región cromosómica expuesta en la Tabla 2. Preferiblemente, el ADN extracelular en circulación carece de elementos repetitivos. En un aspecto, el paciente es un individuo que se sospecha que tiene o diagnosticado de cáncer, por ejemplo, cáncer colorrectal.

Por "encontrarse dentro" se entiende en el presente documento que la secuencia de nucleótidos de un ADN extracelular en circulación es sustancialmente idéntica (por ejemplo, mayor que el 95 % idéntica) a una parte de la secuencia de nucleótidos de una región cromosómica y puede asignarse inequívocamente a la región cromosómica. En otras palabras, el ADN extracelular en circulación puede hibridar en condiciones rigurosas con, u obtenerse de, la región cromosómica.

En un aspecto, se proporciona un método para analizar ADN extracelular en circulación en una muestra de un paciente, que comprende determinar, en una muestra que es sangre, suero o plasma, un nivel de una pluralidad de moléculas de ADN extracelular en circulación, teniendo cada una una secuencia de al menos 25 nucleótidos consecutivos de longitud, o al menos 40, 50, 60, 75, o 100 o más nucleótidos consecutivos, que se encuentra dentro de la misma una región cromosómica única expuesta en la Tabla 2. Puede haber dos o más, o cualquier número de distintas moléculas de ADN extracelular en circulación que proceden todas de la misma una región cromosómica expuesta en la Tabla 2, y, en algunas realizaciones, se detectan todas tales moléculas de ADN extracelular en circulación y se determinan los niveles de las mismas.

Preferiblemente, las secuencias de las moléculas de ADN extracelular en circulación carecen de elementos repetitivos.

En un aspecto, se proporciona un método para analizar ADN extracelular en circulación en una muestra de un paciente, que comprende determinar, en una muestra que es sangre, suero o plasma, un nivel de al menos 2, 3, 4, 5, 7, 8, 9, 10, 15, 20, 30, 40, 50, 55, 60, 65, 70, 75, o al menos 80 o de 81 moléculas de ADN extracelular en circulación, teniendo cada una una secuencia de al menos 25 nucleótidos consecutivos, o al menos 40, 50, 60, 75, o 100, o más nucleótidos consecutivos, que se encuentra dentro de una región cromosómica distinta, como se expone en la Tabla 2. Preferiblemente, las secuencias de las moléculas de ADN extracelular en circulación carecen de elementos repetitivos. En aspectos preferidos, las moléculas de ADN extracelular tienen secuencias que se encuentran dentro de distintas regiones cromosómicas en la Tabla 2. En un aspecto específico, se determinan los niveles de al menos 2, 3, 4, 5, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, o al menos 80, o de 81,

moléculas de ADN extracelular en circulación, estando la secuencia de cada una dentro de una región cromosómica diferente expuesta en la Tabla 2.

5 En una realización específica, el método para analizar ADN extracelular en circulación incluye las etapas de: aislar, de una muestra de sangre, suero o plasma de un paciente, sustancialmente todas las moléculas de ADN extracelular en circulación que tengan una longitud de al menos 20, 25, 30, 40, 50, 75 o 100 nucleótidos consecutivos de longitud, o entre 50 y 400 nucleótidos de longitud, obtener la secuencia de cada una de las moléculas de ADN extracelular en circulación, determinar si la secuencia se encuentra dentro de un conjunto de regiones cromosómicas expuestas en la Tabla 2 y el nivel de dicha secuencia.

10 En otra realización específica, el método para analizar ADN extracelular en circulación incluye las etapas de: aislar, de una muestra de sangre, suero o plasma de un paciente, sustancialmente todas las moléculas de ADN extracelular en circulación que tengan una longitud de al menos 20, 25, 30, 40, 50, 75 o 100 nucleótidos consecutivos de longitud, o entre 50 y 400 nucleótidos de longitud, y poner en contacto las moléculas de ADN extracelular en circulación con una pluralidad de oligonucleótidos (por ejemplo, en un chip de ADN o micromatriz) para determinar si una o más de las moléculas de ADN extracelular en circulación hibrida con una cualquiera de la pluralidad de sondas de oligonucleótido en condiciones rigurosas. Cada una de las sondas de oligonucleótido tiene una secuencia de nucleótidos idéntica a una parte de la secuencia de una región cromosómica expuesta en la Tabla 2. Por lo tanto, si una molécula de ADN en circulación hibrida en condiciones rigurosas con una de las sondas de oligonucleótido, indica que la molécula de ADN en circulación tiene una secuencia de nucleótidos que se encuentra dentro de una región cromosómica expuesta en la Tabla 2 e indica la presencia de la molécula de ADN en circulación. El nivel de la molécula de ADN en circulación se puede determinar determinando la cantidad de sondas hibridadas.

25 En las diversas realizaciones anteriores, preferiblemente, las moléculas de ADN extracelular en circulación tienen al menos 25 nucleótidos consecutivos de longitud (preferiblemente al menos 50, 70, 80, 100, 120 o 200 nucleótidos consecutivos de longitud). Más preferiblemente, las moléculas de ADN extracelular en circulación tienen entre aproximadamente 50 y aproximadamente 300 o 400, preferiblemente entre aproximadamente 75 y aproximadamente 300 o 400, más preferiblemente de aproximadamente 100 a aproximadamente 200 nucleótidos consecutivos de una única secuencia dentro de una región cromosómica como se expone en la Tabla 2.

30 En otro aspecto, la presente invención proporciona un método para el diagnóstico o el cribado del cáncer colorrectal en un paciente. El método incluye las etapas de: (a) determinar, en una muestra que es sangre, suero o plasma de un paciente, el nivel de al menos 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, al menos 30 o más, o de 35, moléculas de ADN extracelular en circulación, teniendo cada una una secuencia de al menos 25 nucleótidos de longitud que se encuentra dentro de una región cromosómica diferente designada como "ASCENDENTE" en la Tabla 2; y (b) correlacionar la presencia de un nivel aumentado del ADN extracelular en circulación, con respecto a lo normal, con una mayor probabilidad de que el paciente tenga cáncer colorrectal.

40 En otra realización, el método de la invención incluye las etapas de: (a) determinar, en una muestra que es sangre, suero o plasma de un paciente, el nivel de al menos 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, de al menos 45, o de 46, moléculas de ADN extracelular en circulación, teniendo cada una una secuencia de al menos 25 nucleótidos de longitud que se encuentra dentro de una región cromosómica diferente designada como "DESCENDENTE" en la Tabla 2; y (b) correlacionar la presencia de un nivel disminuido del ADN extracelular en circulación, con respecto a lo normal, con una mayor probabilidad de que el paciente tenga cáncer colorrectal.

45 Cuando las etapas de los métodos anteriores se aplican a un paciente diagnosticado de cáncer colorrectal, puede seguirse al paciente en cuanto al estado del cáncer colorrectal, o para determinar el efecto del tratamiento de un régimen de tratamiento particular, o para detectar la recaída o recidiva del cáncer.

50 En el método de diagnóstico/seguimiento de la presente invención, preferiblemente, las secuencias de las moléculas de ADN extracelular en circulación carecen de elementos repetitivos. En realizaciones preferidas, las moléculas de ADN extracelular tienen secuencias que se encuentran dentro de distintas regiones cromosómicas expuestas en la Tabla 2.

55 En una realización, se proporciona un método de diagnóstico del cáncer colorrectal en un individuo, que comprende (a) determinar los niveles de al menos 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, al menos 30 o más, o de 35, moléculas de ADN extracelular en circulación, teniendo cada una una secuencia de al menos 25 nucleótidos de longitud que se encuentra dentro de una región cromosómica diferente designada como "ASCENDENTE" en la Tabla 2; y (b) correlacionar la presencia de un nivel aumentado, con respecto a lo normal, una o más de las moléculas de ADN extracelular en circulación con una probabilidad aumentada de que el individuo tenga cáncer colorrectal o una recaída del cáncer colorrectal, o una ineficacia del tratamiento para el cáncer colorrectal.

60 En una realización, se proporciona un método de diagnóstico/seguimiento del cáncer colorrectal en un individuo, que comprende (a) determinar los niveles de al menos 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, de al menos 45, o de 46, moléculas de ADN extracelular en circulación, teniendo cada una una secuencia de al menos 25 nucleótidos de longitud que se encuentra dentro de una región cromosómica diferente designada como "DESCENDENTE" en la

Tabla 2; y (b) correlacionar la presencia de un nivel disminuido, con respecto a lo normal, una o más de las moléculas de ADN extracelular en circulación con una probabilidad aumentada de que el individuo tenga cáncer colorrectal o una recaída del cáncer colorrectal, o una ineficacia del tratamiento para el cáncer colorrectal.

- 5 En otra realización más, el método de diagnóstico, seguimiento o cribado del cáncer colorrectal en un paciente, incluye determinar, en una muestra que es sangre, suero o plasma del paciente, el nivel de cada uno y todos los ADN extracelulares en circulación, teniendo cada uno una secuencia que está dentro de la misma región cromosómica única designada como "ASCENDENTE" en la Tabla 2; y correlacionar un nivel total aumentado de dichos ADN extracelulares en circulación, con una probabilidad aumentada de que dicho paciente tenga cáncer colorrectal o una recaída del cáncer colorrectal. En otras palabras, puede haber cualquier cantidad de, y normalmente muchas, distintas moléculas de ADN extracelular en circulación procedentes de una misma única región cromosómica expuesta en la Tabla 2, y todas tales distintas moléculas de ADN extracelular en circulación se detectan y se determinan los niveles, y se realiza la correlación con el estado del cáncer colorrectal.
- 10
- 15 En otra realización, el método de diagnóstico, seguimiento o cribado del cáncer colorrectal en un paciente, incluye determinar, en una muestra que es sangre, suero o plasma del paciente, el nivel de cada uno y todos los ADN extracelulares en circulación, teniendo cada uno una secuencia que está dentro de la misma región cromosómica única designada como "DESCENDENTE" en la Tabla 2; y correlacionar una disminución del nivel de dichos ADN extracelulares en circulación con una mayor probabilidad de que dicho paciente tenga cáncer colorrectal o una recaída del cáncer colorrectal. En otras palabras, puede haber cualquier cantidad de, y normalmente muchas, distintas moléculas de ADN extracelular en circulación procedentes de una misma única región cromosómica expuesta en la Tabla 2, y todas tales distintas moléculas de ADN extracelular en circulación se detectan y se determina el nivel, y se realiza la correlación con el estado del cáncer colorrectal.
- 20
- 25 En una realización específica, sustancialmente todas las moléculas de ADN extracelular en circulación que tengan una longitud de al menos 20, 25, 30, 40, 50, 75 o 100 nucleótidos consecutivos de longitud, o entre 50 y 400 nucleótidos de longitud, se aíslan de una muestra de sangre, suero o plasma de un paciente. Se determina la secuencia de al menos alguna porción representativa de cada una de las moléculas de ADN extracelular en circulación aisladas y se compara con una o más de las secuencias de las regiones cromosómicas expuestas en la Tabla 2 para determinar si la secuencia de un ADN extracelular en circulación está dentro de una región cromosómica designada como "ASCENDENTE" en la Tabla 2 y el nivel del ADN en circulación que tiene dicha secuencia. Si el nivel aumenta con respecto a lo normal, se realiza un diagnóstico de cáncer colorrectal. En el caso de un paciente tratado con una terapia para el cáncer colorrectal, se indica una recaída si se detecta un aumento, con respecto a lo normal, en el nivel de un ADN extracelular en circulación que se encuentra dentro de una región cromosómica designada como "ASCENDENTE" en la Tabla 2. En realizaciones preferidas, se indica un diagnóstico de cáncer colorrectal o ineficacia del tratamiento o recaída del cáncer colorrectal, si aumentan dos o más moléculas de ADN extracelular en circulación que se encuentran dentro de 2, 3, 4, 5, 6, 7, 8, 9, 10, o más regiones cromosómicas designadas como "ASCENDENTE" en la Tabla 2.
- 30
- 35
- 40 En otra realización específica, sustancialmente todas las moléculas de ADN extracelular en circulación que tengan una longitud de al menos 20, 25, 30, 40, 50, 75 o 100 nucleótidos consecutivos de longitud, o entre 50 y 400 nucleótidos de longitud, se aíslan de una muestra de sangre, suero o plasma de un paciente. Estas moléculas de ADN extracelular en circulación, o una porción representativa de las mismas, se hibridan con una micromatriz que se describe anteriormente en el contexto del kit de la divulgación para determinar si una de las moléculas de ADN extracelular en circulación hibrida con una cualquiera de una pluralidad de sondas de oligonucleótido en condiciones rigurosas. Cada una de las sondas de oligonucleótido tiene una secuencia de nucleótidos idéntica a una parte de la secuencia de una región cromosómica designada como "ASCENDENTE" en la Tabla 2. Por lo tanto, si una molécula de ADN en circulación hibrida en condiciones rigurosas con una de las sondas de oligonucleótido, indica que la molécula de ADN en circulación tiene una secuencia de nucleótidos que se encuentra dentro de una región cromosómica expuesta en la Tabla 2 y se determina el nivel. Si el nivel aumenta, con respecto a lo normal, se realiza un diagnóstico de cáncer colorrectal. En el caso de un paciente tratado con una terapia para el cáncer colorrectal, se indica una recaída si hay un aumento del nivel de un ADN extracelular en circulación que se encuentra dentro de una región cromosómica designada como "ASCENDENTE" en la Tabla 2. En realizaciones preferidas, se indica un diagnóstico de cáncer colorrectal o ineficacia del tratamiento o recaída del cáncer colorrectal, si aumentan dos o más moléculas de ADN extracelular en circulación que se encuentran dentro de 2, 3, 4, 5, 6, 7, 8, 9, 10, o más regiones cromosómicas designadas como "ASCENDENTE" en la Tabla 2.
- 45
- 50
- 55

En una realización específica, sustancialmente todas las moléculas de ADN extracelular en circulación que tengan una longitud de al menos 20, 25, 30, 40, 50, 75 o 100 nucleótidos consecutivos de longitud, o entre 50 y 400 nucleótidos de longitud, se aíslan de una muestra de sangre, suero o plasma de un paciente. Se determina la secuencia de al menos alguna porción representativa de cada una de las moléculas de ADN extracelular en circulación aisladas y se compara con una o más de las secuencias de las regiones cromosómicas expuestas en la Tabla 2 para determinar si la secuencia de un ADN extracelular en circulación está dentro de una región cromosómica designada como "DESCENDENTE" en la Tabla 2 y el nivel del polinucleótido que tiene dicha secuencia. Si el nivel disminuye con respecto a lo normal, se realiza un diagnóstico de cáncer colorrectal. En el caso de un paciente tratado con una terapia para el cáncer colorrectal, se indica una recaída si se detecta una

60

65

disminución, con respecto a lo normal, en el nivel de un ADN extracelular en circulación que se encuentra dentro de una región cromosómica designada como "DESCENDENTE" en la Tabla 2. En realizaciones preferidas, se indica un diagnóstico de cáncer colorrectal o ineficacia del tratamiento o recaída del cáncer colorrectal, si disminuyen dos o más moléculas de ADN extracelular en circulación que se encuentran dentro de 2, 3, 4, 5, 6, 7, 8, 9, 10, o más regiones cromosómicas designadas como "DESCENDENTE" en la Tabla 2.

En otra realización específica, sustancialmente todas las moléculas de ADN extracelular en circulación que tengan una longitud de al menos 20, 25, 30, 40, 50, 75 o 100 nucleótidos consecutivos de longitud, o entre 50 y 400 nucleótidos de longitud, se aíslan de una muestra de sangre, suero o plasma de un paciente. Estas moléculas de ADN extracelular en circulación, o una porción representativa de las mismas, se hibridan con una micromatriz que se describe anteriormente en el contexto del kit de la divulgación para determinar si una de las moléculas de ADN extracelular en circulación hibrida con una cualquiera de una pluralidad de sondas de oligonucleótido en condiciones rigurosas. Cada una de las sondas de oligonucleótido tiene una secuencia de nucleótidos idéntica a una parte de la secuencia de una región cromosómica designada como "DESCENDENTE" en la Tabla 2. Por lo tanto, si una molécula de ADN en circulación hibrida en condiciones rigurosas con una de las sondas de oligonucleótido, indica que la molécula de ADN en circulación tiene una secuencia de nucleótidos que se encuentra dentro de una región cromosómica expuesta en la Tabla 2 y se determina el nivel. Si el nivel disminuye, con respecto a lo normal, se realiza un diagnóstico de cáncer colorrectal. En el caso de un paciente tratado con una terapia para el cáncer colorrectal, se indica una recaída si hay una disminución del nivel de un ADN extracelular en circulación que se encuentra dentro de una región cromosómica designada como "DESCENDENTE" en la Tabla 2. En realizaciones preferidas, se indica un diagnóstico de cáncer colorrectal o ineficacia del tratamiento o recaída del cáncer colorrectal, si disminuyen dos o más moléculas de ADN extracelular en circulación que se encuentran dentro de 2, 3, 4, 5, 6, 7, 8, 9, 10, o más regiones cromosómicas designadas como "ASCENDENTE" en la Tabla 2.

En las diversas realizaciones anteriores, preferiblemente, las moléculas de ADN extracelular en circulación tienen al menos 25 nucleótidos consecutivos de longitud (preferiblemente al menos 50, 70, 80, 100, 120 o 200 nucleótidos consecutivos de longitud). Más preferiblemente, las moléculas de ADN extracelular en circulación tienen entre aproximadamente 50 y aproximadamente 300 o 400, preferiblemente entre aproximadamente 75 y aproximadamente 300 o 400, más preferiblemente de aproximadamente 100 a aproximadamente 200 nucleótidos consecutivos de una única secuencia dentro de una región cromosómica como se expone en la Tabla 2.

Detección de ácidos nucleicos en circulación en la sangre

Para detectar los ácidos nucleicos en circulación en la sangre de pacientes que pueden tener, o se sospecha que tienen, cáncer colorrectal, se obtiene una muestra de sangre del paciente. Después, se analiza el suero o plasma de la muestra de sangre en cuanto a la presencia y el nivel de un ADN extracelular en circulación o biomarcador, como se describe en el presente documento. Los ácidos nucleicos pueden aislarse del suero o plasma usando técnicas muy conocidas, véase, por ejemplo, las secciones de los ejemplos. En el contexto de la actual invención, las secuencias de ácido nucleico que se analizan son secuencias de ADN. Por lo tanto, en esta sección, los métodos descritos para la evaluación de "ácidos nucleicos" se refieren a la evaluación de ADN.

Las técnicas de detección para evaluar ácidos nucleicos en cuanto a la presencia y el nivel de un biomarcador implican procedimientos muy conocidos en el campo de la genética molecular. Además, muchos de los métodos implican la amplificación de ácidos nucleicos. Se proporciona en la técnica una amplia guía para llevarlos a cabo. Las referencias ejemplares incluyen manuales tales como PCR Technology: Principles and Applications for DNA Amplification (ed. H. A. Erlich, Freeman Press, NY, N.Y., 1992); PCR Protocols: A Guide to Methods and Applications (eds. Innis, *et al.*, Academic Press, San Diego, Calif., 1990); Current Protocols in Molecular Biology, Ausubel, 1994-1999, incluyendo actualizaciones suplementarias hasta abril de 2004; Sambrook & Russell, Molecular Cloning, A Laboratory Manual (3ª Ed, 2001).

Aunque los métodos pueden emplear etapas de PCR, también se pueden usar otros protocolos de amplificación. Los métodos de amplificación adecuados incluyen la reacción en cadena de la ligasa (véase, por ejemplo, Wu y Wallace, Genomics 4:560-569, 1988); ensayo de desplazamiento de cadena (véase, porejemplo, Walker *et al.*, Proc. Natl. Acad. Sci. USA 89:392-396, 1992; Patente de EE.UU N.º 5.455.166); y varios sistemas de amplificación basados en transcripción, incluyendo los métodos descritos en las patentes de EE.UU. N.º 5.437.990; 5.409.818; y 5.399.491; el sistema de amplificación basado en transcripción (TAS) (Kwoh *et al.*, Proc. Natl. Acad. Sci. USA 86:1173-1177, 1989); y replicación de secuencia autosostenida (3SR) (Guatelli *et al.*, Proc. Natl. Acad. Sci. USA 87:1874-1878, 1990; documento WO 92/08800). Como alternativa, se pueden usar métodos que amplifican la sonda hasta niveles detectables, tal como la amplificación por Q β -replicasa (Kramer y Lizardi, Nature 339:401-402, 1989; Lomeli *et al.*, Clin. Chem. 35: 1826-1831, 1989). Se proporciona una revisión de los métodos de amplificación conocidos, por ejemplo, por Abramson y Myers en Current Opinion in Biotechnology 4:41-47, 1993.

En algunas realizaciones, la detección de un biomarcador en el CNA de un paciente se realiza utilizando cebadores oligonucleotídicos y/o sondas para detectar una secuencia diana, en la que la secuencia diana está presente en (por ejemplo, comprende alguna porción asignada de forma inequívoca de) cualquiera de las regiones cromosómicas enumeradas en la Tabla 2). Los oligonucleótidos pueden prepararse mediante cualquier método adecuado,

habitualmente síntesis química, y también se pueden adquirir a través de fuentes comerciales. Los oligonucleótidos pueden incluir enlaces fosfodiéster modificados (por ejemplo, fosforotioato, metilfosfonatos, fosfoamidato o boranofosfato) o pueden usarse en un oligonucleótido enlaces distintos de un derivado de ácido fosforoso, para impedir la escisión en un sitio seleccionado. Además, el uso de azúcares modificados en 2'-amino tiende a favorecer el desplazamiento sobre la digestión del oligonucleótido cuando se hibrida con un ácido nucleico que también es el molde para la síntesis de una nueva cadena de ácido nucleico.

En una realización, el biomarcador se identifica por hibridación con una sonda que se dirige a una región cromosómica en condiciones de hibridación específicas de secuencia, por ejemplo, se dirige a alguna porción asignada de forma inequívoca de, cualquiera de las regiones cromosómicas enumeradas en la Tabla 2) descritas en el presente documento. La sonda utilizada para este análisis puede ser una sonda larga o conjuntos para sondas de oligonucleótido cortas, por ejemplo, pueden emplearse de aproximadamente 20 o aproximadamente 150 nucleótidos de longitud.

Los formatos de hibridación adecuados son muy conocidos en la técnica, incluyendo, pero sin limitación, de fase en solución, fase sólida, formatos de matrices de oligonucleótidos, de fase mixta o ensayos de hibridación *in situ*. En las hibridaciones en fase de solución (o líquida), tanto el ácido nucleico diana como la sonda o los cebadores están libres para interactuar en la mezcla de reacción. Además se han desarrollado técnicas, tales como los sistemas de PCR en tiempo real, que permiten el análisis, por ejemplo, la cuantificación de productos amplificados durante una reacción de PCR. En este tipo de reacción, la hibridación con una sonda de oligonucleótido específica se produce durante el programa de amplificación, para identificar la presencia de un ácido nucleico diana. La hibridación de sondas oligonucleotídica asegura la mayor especificidad debido a la transición de dos estados controlada termodinámicamente. Los ejemplos de este formato de ensayo son las sondas de hibridación de transferencia de energía de resonancia de fluorescencia, balizas moleculares, escorpiones moleculares y sondas de hibridación con exonucleasas (por ejemplo, revisado en Bustin, *J. Mol. Endocrin.* 25:169-93, 2000).

Los formatos de ensayo adecuados incluyen formatos basados en matrices, descritos con mayor detalle a continuación en la sección "Dispositivo", donde la sonda normalmente está inmovilizada. Como alternativa, puede inmovilizarse la diana.

En un formato donde la diana está inmovilizada, el ADN diana amplificado se inmoviliza en un soporte sólido y el complejo diana se incuba con la sonda en condiciones de hibridación adecuadas, la sonda no hibridada se elimina lavando en condiciones rigurosas de forma adecuada y el soporte sólido se controla en cuanto a la presencia de sonda unida. En formatos en donde las sondas se inmovilizan en un soporte sólido, el ADN diana típicamente se marca, habitualmente durante la amplificación. La sonda inmovilizada se incuba con el ADN diana amplificado en condiciones de hibridación adecuadas, el ADN diana no hibridado se elimina lavando en condiciones rigurosas de forma adecuada y el soporte sólido/sonda se controla en cuanto a la presencia de ADN diana unido.

En realizaciones típicas, se inmovilizan múltiples sondas en un soporte sólido y se analizan las regiones cromosómicas diana en el CNA de un paciente utilizando las múltiples sondas de forma simultánea. El documento WO 95/11995 describe ejemplos de matrices de ácidos nucleicos.

En un método alternativo sin sonda, el ácido nucleico amplificado correspondiente a un ácido nucleico diana presente en una región cromosómica se realiza utilizando cebadores de ácido nucleico para la región cromosómica y se detecta controlando el aumento del nivel total de ADN bicatenario en la mezcla de reacción; se describe, por ejemplo, en la patente de EE.UU. N.º 5.994.056; y en las publicaciones de patente europea N.º 487.218 y 512.334. La detección de ADN diana bicatenario está basada en la fluorescencia aumentada de diversos colorantes de unión a ADN, por ejemplo, SYBR Green, que se presenta cuando se unen a ADN bicatenario.

Como aprecia un experto en la técnica, se pueden realizar en la reacción métodos específicos de amplificación que emplean múltiples cebadores para dirigirse a las regiones cromosómicas, de forma que el biomarcador pueda abarcarse adecuadamente.

Secuenciación de ADN

En realizaciones preferidas, una secuencia de una región cromosómica expuesta en la Tabla 2 en el CNA de un paciente en evaluación se detecta mediante secuenciación directa. Dicha secuenciación, especialmente utilizando los sistemas de secuenciación de Roche 454, Illumina y de Applied Biosystems mencionados a continuación o sistemas de secuenciación avanzada similares, puede incluir la cuantificación de ácidos nucleicos que tienen una secuencia particular para determinar el nivel de un biomarcador. En realizaciones típicas, se secuencia el CNA de un paciente utilizando un método de secuenciación a gran escala que proporciona la capacidad de obtener información de las secuencias a partir de muchas lecturas. Dichas plataformas de secuenciación incluyen las comercializadas por Roche 454 Life Sciences (sistemas GS), Illumina (por ejemplo, HiSeq, MiSeq) y Applied Biosystems (por ejemplo, sistemas SOLiD).

La plataforma de secuenciación de Roche 454 Life Sciences implica el uso de PCR en emulsión y la inmovilización

de fragmentos de ADN en perlas. La incorporación de nucleótidos durante la síntesis se detecta midiendo la luz que se genera cuando se incorpora un nucleótido.

5 La tecnología Illumina implica la unión de ADN genómico fragmentado aleatoriamente a una superficie plana, ópticamente transparente. Los fragmentos de ADN unidos se extienden y se amplifican en puente para crear una celda de flujo de secuenciación de densidad ultra alta con conglomerados que contienen copias del mismo molde. Estos moldes se secuencian utilizando una tecnología de secuenciación por síntesis que emplea terminadores reversibles con colorantes fluorescentes que se pueden eliminar.

10 Además, se pueden usar métodos que emplean secuenciación por hibridación. Dichos métodos, por ejemplo, como se utilizan en la tecnología SOLiD4+ de Applied Biosystems, implican una PCR en emulsión que inmoviliza fragmentos de ADN en perlas, seguido del uso de un conjunto de todos los oligonucleótidos posibles de una longitud fija, marcados de acuerdo con la posición secuenciada. Los oligonucleótidos son emparejados y unidos; el ligamiento preferencial mediante la ADN ligasa para aparear secuencias da como resultado una señal informativa del
15 nucleótido en esa posición.

La secuencia se puede determinar utilizando cualquier otro método de secuenciación de ADN, incluyendo, por ejemplo, métodos que utilizan tecnología de semiconductores para detectar nucleótidos que se incorporan en un
20 cebador extendido, midiendo los cambios en la corriente que se producen cuando se incorpora un nucleótido (véase, por ejemplo, las Publicaciones de Solicitud de Patente de EE.UU. N.º 20090127589 y 20100035252). Otras técnicas incluyen la secuenciación directa de exonucleasas sin etiqueta en la que los nucleótidos escindidos del ácido nucleico se detectan al pasar a través de un nanoporo (Oxford Nanopore) (Clark et al., Nature Nanotechnology 4: 265 - 270, 2009); y tecnología de secuenciación de ADN en tiempo real de una única molécula (SMRT™) (Pacific Biosciences), que es una técnica de secuenciación por síntesis.

25

Dispositivos y kits

En un aspecto adicional, la divulgación proporciona dispositivos de diagnóstico y kits útiles para identificar y
30 determinar el nivel de uno o más biomarcadores asociados con cáncer colorrectal en el CNA de un paciente en el que el uno o más biomarcadores tienen una secuencia asignada de forma inequívoca a cualquiera de las regiones cromosómicas expuestas en la Tabla 2. Como será evidente para los expertos en la técnica, el kit de la presente divulgación es útil en el método discutido anteriormente para analizar el ADN extracelular en circulación en una muestra del paciente y en el diagnóstico, seguimiento o cribado del cáncer colorrectal, como se describe
35 anteriormente.

Por lo tanto, en un aspecto, la presente divulgación proporciona el uso de al menos un oligonucleótido para la
40 fabricación de un kit de diagnóstico útil en el diagnóstico, seguimiento o cribado del cáncer colorrectal. La secuencia de nucleótidos del oligonucleótido que se encuentra dentro de una región cromosómica expuesta en la Tabla 2.

Preferiblemente, el kit de la presente divulgación incluye uno, dos o más (por ejemplo, al menos 1, 2, 3, 4, 5, 6, 7, 8,
45 9, 10, 12, 15, 20, 25, 30, 40 o al menos 50, pero preferiblemente menos de 81, preferiblemente de uno a aproximadamente 50, más preferiblemente de 2 a preferiblemente 50, o de 3 a preferiblemente 50 conjuntos de oligonucleótidos. Cada conjunto comprende uno o más oligonucleótidos (por ejemplo, de aproximadamente uno a aproximadamente 10.000, preferiblemente de 50, 100, 200 o 300 a aproximadamente 10.000). Todas las secuencias
50 de nucleótidos de tales uno o más oligonucleótidos en cada conjunto se encuentran dentro de la misma una única región cromosómica que se expone en la Tabla 2 (o que coinciden con una parte de la misma secuencia una única expuesta en la Tabla A). Cada oligonucleótido debe tener de aproximadamente 18 a 100 nucleótidos, o de 20 a aproximadamente 50 nucleótidos, y tiene la capacidad de hibridar, en condiciones de hibridación rigurosas, con la región cromosómica en la que se encuentra su secuencia. Los oligonucleótidos son útiles como sondas para
55 detectar moléculas de ADN extracelulares en circulación procedentes de las regiones cromosómicas. Preferiblemente, cada conjunto incluye un número suficiente de oligonucleótidos con secuencias que mapean en una región cromosómica, de forma que cualquier molécula de ADN extracelular en circulación procedente de la región cromosómica puede detectarse con el conjunto de oligonucleótidos. Por lo tanto, el número de oligonucleótidos necesarios en cada conjunto está determinado por la longitud total de la secuencia de nucleótidos singular de una
60 región cromosómica particular, como será evidente para los expertos en la técnica. Dichas longitudes totales se indican en la Tabla 2.

Preferiblemente, en el kit de la presente divulgación, los distintos conjuntos de oligonucleótidos corresponden a
65 distintas regiones cromosómicas dentro de la misma tabla. Preferiblemente, los oligonucleótidos carecen de un elemento repetitivo. Opcionalmente, los oligonucleótidos se unen a uno o más sustratos sólidos, tales como *microchips* y perlas. En aspectos preferidos, el kit es una micromatriz con los oligonucleótidos anteriores.

También se contempla el uso de los oligonucleótidos incluidos en el kit descrito para la fabricación del kit útil para el
diagnóstico, el cribado o el seguimiento del cáncer colorrectal. La fabricación de tal kit debe ser evidente para un
65 experto en la técnica.

En algunos aspectos, un dispositivo de diagnóstico comprende sondas para detectar al menos 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 30, 40, 50, 60, 75, 80, o las 81 regiones cromosómicas expuestas en la Tabla 2. En algunos aspectos, la presente divulgación proporciona sondas unidas a un soporte sólido, tal como un portaobjetos de matriz o chip, por ejemplo, como se describe en DNA Microarrays: A Molecular Cloning Manual, 2003, Eds. Bowtell y Sambrook, Cold Spring Harbor Laboratory Press. La construcción de tales dispositivos es muy conocida en la técnica, como se describe, por ejemplo, en las patentes de Estados Unidos y las publicaciones de patentes de EE.UU. N.º 5.837.832; solicitud PCT WO95/11995; Patente de EE.UU. N.º 5.807.522; Patentes de EE.UU. N.º 7.157.229, 7.083.975, 6.444.175, 6.375.903, 6.315.958, 6.295.153, y 5.143.854, 2007/0037274, 2007/0140906, 2004/0126757, 2004/0110212, 2004/0110211, 2003/0143550, 2003/0003032, y 2002/0041420. Las matrices de ácidos nucleicos también se revisan en las siguientes referencias: Biotechnol Annu Rev 8:85-101 (2002); Sosnowski *et al.*, Psychiatr Genet 12(4):181-92 (diciembre de 2002); Heller, Annu Rev Biomed Eng 4: 129-53 (2002); Kolchinsky *et al.*, Hum. Mutat 19(4):343-60 (abril de 2002); y McGail *et al.*, Adv Biochem Eng Biotechnol 77:21-42 (2002).

Puede implementarse en una matriz cualquier número de sondas. Se puede usar un conjunto de sondas que hibrida con distintos segmentos, preferentemente singulares, de una región cromosómica, donde el conjunto de sondas detecta cualquier parte de la región cromosómica. Como alternativa, se puede inmovilizar sobre una superficie sólida una única sonda para una región cromosómica. La sonda polinucleotídica puede sintetizarse en áreas designadas (o sintetizarse por separado y después fijarse en áreas designadas) en un sustrato, por ejemplo, utilizando un procedimiento químico dirigido por luz. Los polinucleótidos sintéticos típicos pueden ser de aproximadamente 15-200 nucleótidos de longitud.

El kit puede incluir múltiples reactivos de detección de biomarcadores, o uno o más reactivos de detección de biomarcadores en combinación con uno o más de otros tipos de elementos o componentes (por ejemplo, otros tipos de reactivos bioquímicos, recipientes, embalajes tales como el embalaje para la venta comercial, sustratos a los que se están unidos los reactivos de detección de biomarcadores, componentes electrónicos de hardware, etc.). Por consiguiente, la presente divulgación proporciona adicionalmente kits y sistemas de detección de biomarcadores, que incluyen, pero sin limitación, matrices/micromatrices de moléculas de ácido nucleico y perlas que contienen una o más sondas u otros reactivos de detección para detectar uno o más biomarcadores de la presente divulgación. Los kits pueden incluir opcionalmente diversos componentes electrónicos de hardware; por ejemplo, las matrices ("chips de ADN") y los sistemas microfluidos (sistemas de "laboratorio en un chip") proporcionados por diversos fabricantes normalmente comprenden componentes de hardware. Otros kits pueden no incluir componentes electrónicos de hardware, pero pueden estar compuesto de, por ejemplo, uno o más reactivos de detección de biomarcadores (junto con, opcionalmente, otros reactivos bioquímicos) envasados en uno o más recipientes.

Los kits/sistemas de detección de biomarcadores pueden contener, por ejemplo, una o más sondas, o conjuntos de sondas, que hibridan con una molécula de ácido nucleico presente en una región cromosómica expuesta en la Tabla 2.

Un kit de detección de biomarcadores de la presente divulgación puede incluir componentes que se usan para preparar CNA a partir de una muestra de sangre de un paciente, para la posterior amplificación y/o detección de un biomarcador.

Correlación de la presencia de biomarcadores con cáncer colorrectal

La presente invención proporciona métodos para detectar el nivel de un biomarcador en CNA de un paciente que tiene cáncer colorrectal o que se está evaluando para determinar si el paciente puede tener cáncer colorrectal. La divulgación proporciona además reactivos para detectar el nivel de un biomarcador en CNA de un paciente que tiene cáncer colorrectal o que se está evaluando para determinar si el paciente puede tener cáncer colorrectal. En el contexto de la invención, "detección" o "identificación" o "identificación de la presencia" o "detección de la presencia" de un biomarcador asociado con cáncer colorrectal en una muestra de CNA de un paciente, se refiere a la determinación de cualquier nivel del biomarcador en el CNA del paciente, donde el nivel es mayor que un valor umbral que distingue entre muestras de CNA de cáncer colorrectal y sin cáncer colorrectal para un ensayo dado.

En la presente invención, por ejemplo, un aumento en el nivel de cualquiera de las regiones cromosómicas (es decir, biomarcadores) designadas como "ASCENDENTE" en la Tabla 2 es indicativo de cáncer colorrectal. En algunas realizaciones, se puede observar que un biomarcador está presente con poca frecuencia en el CNA obtenido de individuos normales; sin embargo, dada la baja frecuencia de aparición en muestras normales con respecto a la frecuencia de aparición más alta en cáncer colorrectal, la presencia del biomarcador en un paciente indica que el paciente tiene una probabilidad mayor, por ejemplo, 95 % o más de probabilidad, de tener cáncer colorrectal.

Los biomarcadores designados como "ASCENDENTE" en la Tabla 2 están asociados con el cáncer colorrectal, es decir, están sobrerrepresentados en pacientes con cáncer colorrectal, en comparación con individuos no diagnosticados de cáncer colorrectal. Por lo tanto, la detección de un aumento, con respecto a los pacientes con cáncer no colorrectal, en el nivel de uno o más de los biomarcadores designados como "ASCENDENTE" en la Tabla 2, es indicativo de cáncer colorrectal, es decir, el paciente tiene una mayor probabilidad de tener cáncer colorrectal comparado con un paciente que no tiene un aumento en el nivel del biomarcador. En algunas realizaciones, la

5 detección y el aumento en el nivel de dos o más biomarcadores designados como "ASCENDENTE" en la Tabla 2 en el CNA de un paciente es indicativo de una mayor probabilidad de cáncer colorrectal. Como se entiende en la técnica, otros criterios, por ejemplo, criterios clínicos, etc., también se emplean para diagnosticar el cáncer colorrectal en el paciente. Por consiguiente, los pacientes que tienen un biomarcador asociado con el cáncer colorrectal también se someten a otros procedimientos de diagnóstico. En algunos aspectos de la divulgación, al paciente se le administra un agente terapéutico para el cáncer colorrectal, tal como uno o más agentes quimioterapéuticos, por ejemplo, 5-fluorouracilo, leucovorina u oxaliplatino o capecitabina; y/o un anticuerpo monoclonal, tal como bevacizumab, cetuximab o panitumumab, o un anticuerpo monoclonal alternativo.

10 Los biomarcadores designados como "DESCENDENTE" en la Tabla 2 están asociados con el cáncer colorrectal, es decir, están subrepresentados en los pacientes con cáncer colorrectal en comparación con los individuos que no están diagnosticados con cáncer colorrectal. Por lo tanto, la detección de una disminución, con respecto a los pacientes con cáncer no colorrectal, en el nivel de uno o más de los biomarcadores designados como "DESCENDENTE" en la Tabla 2, es indicativo de cáncer colorrectal, es decir, el paciente tiene una mayor
15 probabilidad de tener cáncer colorrectal comparado con un paciente que no tiene una disminución en el nivel del biomarcador. En algunas realizaciones, se puede observar que un biomarcador está presente con poca frecuencia en el CNA obtenido de pacientes con cáncer; sin embargo, dada la baja frecuencia de aparición en muestras de cáncer con respecto a la frecuencia de aparición más alta en individuos normales, la presencia del biomarcador en un paciente indica que el paciente tiene una probabilidad menor, por ejemplo, 5 % o menos de probabilidad, de tener
20 cáncer colorrectal. Como se entiende en la técnica, otros criterios, por ejemplo, criterios clínicos, etc., también se emplean para diagnosticar el cáncer colorrectal en el paciente. Por consiguiente, los pacientes que tienen un biomarcador asociado con el cáncer colorrectal también se someten a otros procedimientos de diagnóstico. En algunos aspectos de la divulgación, al paciente se le administra un agente terapéutico para el cáncer colorrectal, tal como uno o más agentes quimioterapéuticos, por ejemplo, 5-fluorouracilo, leucovorina u oxaliplatino o capecitabina;
25 y/o un anticuerpo monoclonal, tal como bevacizumab, cetuximab o panitumumab, o un anticuerpo monoclonal alternativo.

"Sobrerrepresentado" o "nivel aumentado" significa que el nivel de uno o más los ADN extracelulares en circulación es más alto que los niveles normales. En general, esto significa un aumento del nivel en comparación con un valor de índice. A la inversa, "subrepresentado" o "nivel disminuido" significa que el nivel de una o más moléculas de ADN extracelular en circulación es más bajo que los niveles normales. En general, esto significa una disminución del nivel en comparación con un valor de índice.

35 En realizaciones preferidas, el valor de prueba que representa el nivel de un ADN extracelular en circulación particular se compara con uno o más valores de referencia (o valores de índice), y se correlaciona opcionalmente con el cáncer colorrectal o la recaída del cáncer. Opcionalmente, se indica una mayor probabilidad de cáncer colorrectal si el valor de prueba es mayor que el valor de referencia para el CNA enumerado como "ASCENDENTE" en la Tabla 2 o menor que el valor de referencia para el CNA enumerado como "DESCENDENTE" en la Tabla 2.

40 Los expertos en la técnica están familiarizados con diversas formas de obtener y usar valores de índice. Por ejemplo, el valor de índice puede representar el número de copias o la concentración de un ADN extracelular particular enumerado como "ASCENDENTE" en la Tabla 2 en una muestra de sangre de un paciente de interés en estado saludable, en cuyo caso un número de copias o la concentración en una muestra del paciente, en un momento o estado distinto, significativamente mayor (por ejemplo, 1,01 veces, 1,05 veces, 1,10 veces, 1,2 veces,
45 1,3 veces, 1,4 veces, 1,5 veces, 1,6 veces, 1,7 veces, 1,8 veces, 1,9 veces, 2 veces, 3 veces, 4 veces, 5 veces, 10 veces, 20 veces, 30 veces, 40 veces, 50 veces, 100 veces o más alto) que este valor de índice indicaría, por ejemplo, cáncer colorrectal o un probabilidad aumentada de recaída del cáncer colorrectal. En algunas realizaciones, el nivel de CNA se "aumenta" si es de al menos 1, 2, 3, 4, 5, 10, 15, 20 o más desviaciones estándar mayores que el valor de índice en sujetos normales. En algunas realizaciones, un valor de índice puede representar el número de
50 copias o la concentración de un ADN extracelular particular enumerado como "DESCENDENTE" en la Tabla 2 en una muestra de sangre de un paciente de interés en estado saludable, en cuyo caso un número de copias o la concentración en una muestra del paciente, en un momento o estado distinto, significativamente menor (por ejemplo, 1,01 veces, 1,05 veces, 1,10 veces, 1,2 veces, 1,3 veces, 1,4 veces, 1,5 veces, 1,6 veces, 1,7 veces, 1,8 veces, 1,9 veces, 2 veces, 3 veces, 4 veces, 5 veces, 10 veces, 20 veces, 30 veces, 40 veces, 50 veces, 100 veces o más
55 bajo) que este valor de índice indicaría, por ejemplo, cáncer colorrectal o un probabilidad aumentada de recaída del cáncer colorrectal. En algunas realizaciones, el nivel de CNA se "disminuye" si es de al menos 1, 2, 3, 4, 5, 10, 15, 20 o más desviaciones estándar menores que el valor de índice en sujetos normales.

60 Como alternativa, el valor de índice puede representar la concentración promedio o el número de copias de un ADN extracelular en circulación particular para un conjunto de individuos de una población de cáncer diversa o un subconjunto de la población. Por ejemplo, se puede determinar el número promedio de copias o la concentración de un ADN extracelular en circulación en un muestreo aleatorio de pacientes con cáncer colorrectal. Por lo tanto, los pacientes que tienen un número de copias o concentración (valor de prueba) comparable o superior que este valor, se identifican como que tienen una mayor probabilidad de tener cáncer colorrectal o recaída del cáncer colorrectal
65 que los que tienen un valor de prueba más bajo que este valor.

Un valor de índice útil puede representar el número de copias o la concentración de un ADN extracelular en circulación particular o de una combinación (adición ponderada o directa) de dos o más de los ADN extracelulares en circulación correspondientes a la misma región cromosómica o a regiones cromosómicas distintas. Cuando se usan en el método de diagnóstico/seguimiento dos o más biomarcadores o moléculas de ADN extracelular en circulación, puede ponderarse y combinarse el nivel de cada biomarcador o ADN extracelular en circulación. Por lo tanto, se puede proporcionar un valor de prueba (a) ponderando el nivel determinado de cada molécula de ADN extracelular en circulación con un coeficiente predefinido, y (b) combinando el nivel ponderado para proporcionar un valor de prueba. La etapa de combinación puede ser por adición directa o promediando (es decir, ponderado por igual) o por un coeficiente predefinido distinto.

La información obtenida del análisis de biomarcadores se puede almacenar en una forma legible por ordenador. Tal sistema informático típicamente comprende los subsistemas principales, tales como un procesador central, una memoria de sistema (generalmente RAM), un controlador de entrada/salida (E/S), un dispositivo externo tal como una pantalla de visualización a través de un adaptador de pantalla, puertos serie, un teclado, una unidad de disco fija a través de una interfaz de almacenamiento y una unidad de disquete operativa para recibir un disquete, y un dispositivo de CD-ROM (o DVD-ROM) operativo para recibir un CD-ROM. Se pueden conectar muchos otros dispositivos, tal como una interfaz de red conectada a través de un puerto serie.

El sistema informático también puede estar vinculado a una red, que comprenda una pluralidad de dispositivos informáticos vinculados a través de un enlace de datos, tal como un cable Ethernet (coaxial o 10BaseT), línea telefónica, línea ISDN, red inalámbrica, fibra óptica u otro medio de transmisión de señales adecuado, mediante lo cual al menos un dispositivo de red (por ejemplo, ordenador, conjunto de discos, etc.) comprende un patrón de dominios magnéticos (por ejemplo, disco magnético) y/o dominios de carga (por ejemplo, una matriz de celdas DRAM) que componen un patrón de bits que codifican datos adquiridos a partir de un ensayo de la invención.

El sistema informático puede comprender un código para interpretar los resultados de un estudio que evalúa la presencia de uno o más de los biomarcadores. Por lo tanto, en una realización ejemplar, los resultados del análisis de biomarcadores se proporcionan a un ordenador donde un procesador central ejecuta un programa informático para determinar la probabilidad de que un paciente tenga cáncer colorrectal.

La divulgación también proporciona el uso de sistema informático, tal como el que se describe anteriormente, que comprende: (1) un ordenador; (2) un patrón de bits almacenado que codifica los resultados de las pruebas de biomarcadores obtenidos por los métodos de la invención, que puede almacenarse en el ordenador; (3) y, opcionalmente, (4) un programa para determinar la probabilidad de que un paciente tenga cáncer colorrectal.

La divulgación proporciona adicionalmente métodos para generar un informe basado en la detección de uno o más biomarcadores expuestos en la Tabla 2.

Por lo tanto, la presente divulgación proporciona sistemas relacionados con los métodos de la invención anteriores. En un aspecto, la divulgación proporciona un sistema para analizar ADN extracelular en circulación, que comprende: (1) un analizador de muestras para ejecutar el método de análisis de ADN extracelular en circulación en la sangre, suero o plasma de un paciente, como se describe en las diversas realizaciones anteriores; (2) un sistema informático para recibir y analizar automáticamente los datos obtenidos en la etapa (1) para proporcionar un valor de prueba que represente el estado (concentración o número de copias) de una o más moléculas de ADN extracelular en circulación con una secuencia de nucleótidos de al menos 25 nucleótidos que se encuentra dentro de una región cromosómica expuesta en la Tabla 2, y opcionalmente para comparar el valor de prueba con uno o más valores de referencia, cada uno asociado con un estado predeterminado de cáncer colorrectal. En algunos aspectos, el sistema comprende adicionalmente un módulo de visualización que presenta la comparación entre el valor de prueba y el uno o más valores de referencia, o que presenta un resultado de la etapa de comparación.

Por lo tanto, como será evidente para los expertos en la técnica, el analizador de muestras puede ser, por ejemplo, una máquina de secuenciación (por ejemplo, Illumina HiSeq™, Ion Torrent PGM, secuenciador de Applied Biosystems SOLiD™, PacBio RS, Helicos Heliscope™, etc.), una máquina de PCR (por ejemplo, Applied Biosystems 7900, Fluidigm BioMark™, etc.), un instrumento de micromatrices, etc.

En un aspecto, el analizador de muestras es un instrumento de secuenciación, por ejemplo, un instrumento de secuenciación de última generación, tal como los sistemas GS de Roche, los sistemas HiSeq y MiSeq de Illumina, y SOLiD de Applied Biosystems. Las moléculas de ADN extracelular en circulación se aíslan de la sangre o suero o plasma de un paciente, y las secuencias de todas las moléculas de ADN extracelular en circulación se obtienen mediante el analizador de muestras. El instrumento de secuenciación se usa para secuenciar las moléculas de ADN extracelular en circulación y obtener las secuencias de estas moléculas. Después se emplea un sistema informático para analizar automáticamente las secuencias para determinar el nivel de una molécula de ADN extracelular en circulación que tiene una secuencia de nucleótidos de al menos 25 nucleótidos que está dentro de una región cromosómica expuesta en la Tabla 2 en la muestra. Por ejemplo, el sistema informático puede comparar la secuencia de cada molécula de ADN extracelular en circulación en la muestra con la secuencia, disponible en la base de datos de secuencia humana, de la región cromosómica para determinar si existe una coincidencia, es decir,

si la secuencia de una molécula de ADN extracelular en circulación está dentro de una región cromosómica expuesta en la Tabla 2. El sistema informático también determina automáticamente el número de copias de una molécula de ADN extracelular en circulación particular. Opcionalmente, el sistema informático correlaciona automáticamente el resultado del análisis de secuencia con un diagnóstico con respecto al cáncer colorrectal. Por ejemplo, si se identifica que una, y preferiblemente dos o más, moléculas de ADN extracelular en circulación se derivan de regiones cromosómicas designadas como "ASCENDENTE" en la Tabla 2 y están presentes en un nivel aumentado, entonces el sistema informático correlaciona automáticamente este resultado del análisis con un diagnóstico de cáncer colorrectal. Si se identifica que una, y preferiblemente dos o más, moléculas de ADN extracelular en circulación se derivan de regiones cromosómicas designadas como "DESCENDENTE" en la Tabla 2 y están presentes en un nivel disminuido, entonces el sistema informático correlaciona automáticamente este resultado del análisis con un diagnóstico de cáncer colorrectal. Opcionalmente, el sistema informático comprende adicionalmente un módulo de visualización que presenta los resultados del análisis de secuencia y/o el resultado de la etapa de correlación. El módulo de visualización puede ser, por ejemplo, una pantalla de visualización, tal como un monitor de ordenador, monitor de TV o pantalla táctil, una impresora y altavoces de audio.

La función de análisis basada en ordenador se puede implementar en cualquier lenguaje y/o exploradores adecuados. Por ejemplo, puede implementarse con lenguaje C y, preferiblemente, utilizando lenguajes de programación de alto nivel orientados a objetos, tal como Visual Basic, SmallTalk, C++ y similares. La aplicación se puede escribir para adaptarse a entornos tales como el entorno de Microsoft Windows™, incluyendo Windows™ 98, Windows™ 2000, Windows™ NT y similares. Además, la aplicación también puede escribirse para entorno Macintosh™, SUN™, UNIX o LINUX. Además, las etapas funcionales también se pueden implementar utilizando un lenguaje de programación universal o independiente de una plataforma. Los ejemplos de tales lenguajes de programación multiplataforma incluyen, pero sin limitación, lenguaje de marcado de hipertexto (HTLM), JAVA™, JavaScript™, lenguaje de programación Flash, interfaz común de pasarela/lenguaje de consulta estructurada (CGI/SQL), lenguaje práctico de extracción e informes (PERL), AppleScript™ y otros lenguajes de *script* del sistema, lenguaje de programación/lenguaje de consulta estructurado (PL/SQL) y similares. Pueden usarse navegadores compatibles con Java™ o JavaScript™ tales como HotJava™, Microsoft™ Explorer™, o Netscape™. Cuando se usan páginas de Internet de contenido activo, pueden incluir *applets* de Java™ o controles ActiveX™, u otras tecnologías de contenido activo.

La función de análisis también puede incorporarse en productos de programas informáticos y utilizarse en los sistemas descritos anteriormente, u otros sistemas informáticos o basados en Internet. Por consiguiente, otro aspecto de la presente divulgación se refiere a un producto de programa informático que comprende un medio utilizable por ordenador que tiene códigos de programa legibles por ordenador o instrucciones incorporadas en el, para permitir que un procesador lleve a cabo las funciones de análisis y correlación como se describe anteriormente. Estas instrucciones de programa informático se pueden cargar en un ordenador u otro aparato programable para producir una máquina, de forma que las instrucciones que se ejecutan en el ordenador u otro aparato programable creen medios para implementar las funciones o etapas descritos anteriormente. Estas instrucciones de programa informático también pueden almacenarse en una memoria o medio legible por ordenador que puede dirigir un ordenador u otro aparato programable para que funcione de una manera en particular, de forma que las instrucciones almacenadas en la memoria o el medio legible por ordenador produzcan un artículo de fabricación que incluye medios de instrucción que implementen el análisis. Las instrucciones del programa informático también se pueden cargar en un ordenador u otro aparato programable para provocar que se realice una serie de etapas operativas en el ordenador u otro aparato programable, para producir un proceso implementado en el ordenador de forma que las instrucciones que se ejecutan en el ordenador u otro aparato programable proporcionen etapas para implementar las funciones o etapas descritas anteriormente.

Los siguientes ejemplos se proporcionan solo como ilustración y no como limitación. Los expertos en la técnica reconocerán fácilmente una diversidad de parámetros no críticos que podrían cambiarse o modificarse para producir resultados esencialmente similares.

Ejemplos

Ejemplo 1. Identificación de CNA asociado a cáncer colorrectal Muestras de estudio

El estudio evaluó 68 muestras de suero obtenidas de pacientes con cáncer colorrectal y 72 muestras de suero de controles sanos. Las muestras de suero de pacientes se obtuvieron de dos sitios diferentes: la instalación satelital Cleveland Clinic en Florida, EE.UU. (n = 16) y el Ryazan Central Oblast Hospital, Rusia (n = 47). La sangre se extrajo antes de la operación de pacientes sin tratamiento previo con aprobación del IRB local y se procesó como se describe anteriormente (Beck et al., Clin. Chem. 55:730-738, 2009). Las muestras normales se obtuvieron del departamento de Medicina Transfusional de la Georg-August University de Gotingen (n = 12), el Ryazan Central Oblast Hospital (n = 50), Asterand plc., Detroit, MI, EE.UU., (n = 8), y dos voluntarios adicionales.

Construcción de bibliotecas de secuenciación

Después de la extracción de ADN a partir de suero o plasma, utilizando un método estándar basado en sílice, se

realizó una amplificación del genoma completo por duplicado. Los productos de las dos reacciones se agruparon y se utilizaron para un análisis adicional. El adaptador P2 utilizado para la secuenciación y una secuencia de nucleótidos específica de muestra de 10 pb (también denominada código de barras molecular) se añaden mediante PCR utilizando cebadores de fusión. Se realizaron dos PCR consecutivas con diferentes cebadores de fusión; el número total de ciclos fue de cuatro. Tras las PCR, se agrupó el ADN marcado de 43 muestras (grupo 1) o 49 muestras (grupos 2 y 3) y se realizaron todas las preparaciones adicionales sobre este material de ADN agrupado. Otras etapas de preparación de la biblioteca fueron las siguientes:

- i) Restricción del ADN con endonucleasa NlaIII;
- ii) Eliminación de los salientes 3' creados por NlaIII usando el fragmento Klenow grande;
- iii) Ligación de P1 (segundo adaptador de secuenciación) a los extremos romos;
- iv) Amplificación de la biblioteca utilizando cebadores complementarios a los adaptadores P1/P2 de los fragmentos; y
- v) Selección por tamaño utilizando el sistema de electroforesis iBase y geles de agarosa con selección por tamaño de E-Gel al 2 % (Invitrogen) para obtener fragmentos en el rango de 150-250 pb.

Secuenciación

La secuenciación de las bibliotecas se realizó en un instrumento SOLiD4+ (Applied Biosystems) equipado con un sistema EZBead (Applied Biosystems) para realizar las PCR en emulsión. Todos los reactivos necesarios se adquirieron en Applied Biosystems. Las PCR en emulsión y la secuenciación se realizaron según lo recomendado por el fabricante. Para cada fragmento, se secuenciaron 50 pb y 10 pb de código de barras molecular.

Análisis de datos

Las lecturas de secuencia se asignaron a las diferentes muestras de acuerdo con la secuencia del código de barras molecular.

Las secuencias se mapearon en el genoma humano (Versión 36.1/Hg18) usando el paquete de software BioScope (Applied Biosystems) usando la parametrización predeterminada para lecturas de 50 pb. Brevemente, se utilizó el algoritmo de mapeo local del software que emplea un esquema de siembra de 25 bases con dos siembras en serie a partir de la base 1 y la base 16. Durante la extensión de las semillas, un emparejamiento recibió una puntuación de 1 y un desapareamiento recibió una puntuación de -2. Para las lecturas que se asignaron a más de una posición dentro del genoma, la mejor posición de mapeo se registró cuando su puntuación de calidad fue cinco veces mejor que la puntuación de calidad del segundo mejor mapeo (procedimiento de zona clara). Todos los resultados del mapeo se registraron para cada muestra individual. Se determinó el número de lecturas mapeadas en ventanas genómicas de 100.000 pb. Las ventanas (cada una de 100.000 pb de tamaño) se movieron a lo largo de los cromosomas en intervalos de 50.000 pb, comenzando en una posición de 200.000 de cada cromosoma para excluir las regiones de telómeros. Se produjo un archivo de texto tabulado para cada uno de los cromosomas humanos y cada muestra. El archivo de texto tabulado contenía la siguiente información:

- i) Identificación del cromosoma
- ii) Inicio de ventana
- iii) Detención de ventana
- iv) Número de lecturas mapeadas

Cada línea contenía información para una ventana. Estos datos se utilizaron para una búsqueda de aglomeraciones sin supervisión en 300 rondas independientes de selección aleatoria de conjuntos de entrenamiento, que consisten en el 60 % de cada uno de los grupos de control y de la enfermedad.

Selección de conglomerados genómicos

La primera etapa de la búsqueda de aglomeración sin supervisión (UCS) fue:

- 1) Normalización de las lecturas (por muestra)
 - a. Global -> lecturas totales como base
 - b. Local -> lectura por cromosoma como base

Para 300 rondas, los datos se asignaron al azar al entrenamiento (60 %) y al conjunto de validación (40 %). Los conjuntos de entrenamiento se utilizaron para:

- 1) Optimizar las aglomeraciones que segregaron la enfermedad del grupo de control
 - a. Combinando conglomeraciones consecutivas (añadir lecturas)
 - b. Deteniendo al máximo:

- i. n.º de enfermedad < control más pequeño
- ii. n.º de enfermedad > control más grande

2) Registrar cuándo se encontraron óptimos y n.º de enfermedad >12, de otro modo ir a 3):

- 5 a. Normalización (Global/Local)
- b. Cromosoma
- c. Región optimizada (inicio - parada)
- 10 d. N.º de muestras de enfermedad positivas en conjunto de entrenamiento
- e. N.º de muestras de enfermedad positivas en el conjunto de validación usando:
 - i. delimitador del conjunto de entrenamiento
 - ii. delimitador del conjunto de validación (según 1(b.))
- 15 f. valores para cada muestra en (enfermedad segregada/control)
 - i. conjunto de entrenamiento
 - ii. conjunto de validación

3) Realizar análisis en la siguiente ventana

La siguiente asignación aleatoria se realizó y los datos se registraron en una nueva tabla.

Para cada una de las 300 realizaciones, el rendimiento en el conjunto de validación se probó llamando a cada lectura normalizada para cualquier región significativa en ese conjunto positivo si es mayor que los controles o menor que los controles, respectivamente. Una región llamada positivamente se estableció en "1", una no positiva se estableció en "0" para cada muestra y región.

Definición de aglomeraciones finales segregando los controles de cáncer colorrectal:

Todas las regiones identificadas a partir de la UCS anterior se clasificaron según su número de apariciones en las 300 rondas. La superposición o las regiones se combinaron y se eliminaron las duplicaciones.

En tres realizaciones de secuenciación con SOLiD4+, se generaron 1.170.174.163 lecturas. Para el grupo de control, pudieron mapearse un promedio de $6,3 \times 10^6$ (DE: $2,2 \times 10^6$) lecturas por muestra con respecto a la base de datos del genoma humano versión HG18. En el grupo de cáncer colorrectal, el promedio fue de $5,2 \times 10^6$ (DE: $1,6 \times 10^6$).

Las 300 rondas de conjuntos de entrenamiento/validación aleatorios, muestran una separación de los grupos en el conjunto de validación como se indica en la tabla. Las AUC de las curvas ROC para cada ronda se construyeron utilizando la suma de llamadas de lectura en diferentes condiciones (por ejemplo, normalización local y global y ascendente o descendente en la enfermedad).

Los datos en la Tabla 1 muestran AUC de curvas ROC con desviaciones estándar.

Tabla 1

	Todos	Ascendente global	Descendente	Todos	Ascendente local	Descendente
AUC media	88,5 %	88,9 %	87,6 %	88,1 %	86,7 %	90,1 %
Desv. Est.	5,9 %	5,8 %	7,4 %	5,8 %	6,0 %	5,2 %

Se construyó un modelo final a partir de las 300 rondas y se aplicó a todas las muestras. Las regiones de biomarcadores para los cánceres colorrectales definidas de esta manera se proporcionan en la Tabla 2. Estas regiones se pueden utilizar en diferentes combinaciones para la detección del estado de la muestra. El "Orden" se calcula a partir del número de aleatorizaciones (véase anteriormente) en las que se identificó una región. Los gráficos presentados en la Figura 1 con los valores de AUC se basan en la combinación de dichas regiones, denominados positivos al nivel de especificidad del 95 %.

Tabla 2

Dirección	Normalización	HS	Región	Orden
ASCENDENTE	GLOBAL	1	69800001-70200000	31
ASCENDENTE	GLOBAL	1	196550001-196800000	74
ASCENDENTE	GLOBAL	2	34550001-34950000	74
ASCENDENTE	GLOBAL	3	154600001-155050000	38
ASCENDENTE	GLOBAL	3	34350001-34550000	48
ASCENDENTE	GLOBAL	3	133900001-134350000	55

ES 2 729 554 T3

(continuación)

Dirección	Normalización	HS	Región	Orden
ASCENDENTE	GLOBAL	4	27550001-27800000	7
ASCENDENTE	GLOBAL	5	18650001-18950000	47
ASCENDENTE	GLOBAL	5	85650001-85950000	50
ASCENDENTE	GLOBAL	5	90850001-91100000	66
ASCENDENTE	GLOBAL	6	114250001-114550000	25
ASCENDENTE	GLOBAL	7	87000001-87300000	14
ASCENDENTE	GLOBAL	7	11350001-11700000	16
ASCENDENTE	GLOBAL	7	19600001-20100000	38
ASCENDENTE	GLOBAL	7	95100001-95400000	71
ASCENDENTE	GLOBAL	8	51450001-52000000	22
ASCENDENTE	GLOBAL	8	61100001-61450000	34
ASCENDENTE	GLOBAL	8	82850001-83200000	78
ASCENDENTE	GLOBAL	9	75350001-75600000	50
ASCENDENTE	GLOBAL	12	44700001-45050000	31
ASCENDENTE	GLOBAL	14	21350001-22050000	20
DESCENDENTE	GLOBAL	1	180850001-181150000	23
DESCENDENTE	GLOBAL	2	234900001-235400000	8
DESCENDENTE	GLOBAL	2	26950001-27450000	12
DESCENDENTE	GLOBAL	2	95200001-95550000	53
DESCENDENTE	GLOBAL	2	105100001-105400000	55
DESCENDENTE	GLOBAL	3	53950001-54200000	11
DESCENDENTE	GLOBAL	3	140050001-140200000	74
DESCENDENTE	GLOBAL	4	183950001-184250000	18
DESCENDENTE	GLOBAL	5	2400001-2800000	29
DESCENDENTE	GLOBAL	5	134800001-135050000	59
DESCENDENTE	GLOBAL	7	65150001-65350000	66
DESCENDENTE	GLOBAL	8	30200001-30600000	34
DESCENDENTE	GLOBAL	8	10200001-11250000	2
DESCENDENTE	GLOBAL	9	100200001-100550000	61
DESCENDENTE	GLOBAL	10	500001-800000	31
DESCENDENTE	GLOBAL	10	114450001-114750000	36
DESCENDENTE	GLOBAL	10	123650001-124100000	19
DESCENDENTE	GLOBAL	12	127350001-127950000	5
DESCENDENTE	GLOBAL	15	72150001-72400000	64
DESCENDENTE	GLOBAL	16	68250001-68800000	10
DESCENDENTE	GLOBAL	16	19350001-19800000	30
DESCENDENTE	GLOBAL	16	49650001-49950000	37
DESCENDENTE	GLOBAL	16	13050001-13500000	45
DESCENDENTE	GLOBAL	20	47500001-47900000	54
DESCENDENTE	GLOBAL	22	31000001-31200000	42
ASCENDENTE	LOCAL	1	86600001-87150000	12
ASCENDENTE	LOCAL	1	69650001-70250000	9
ASCENDENTE	LOCAL	2	34550001-35100000	71
ASCENDENTE	LOCAL	3	154600001-154950000	42
ASCENDENTE	LOCAL	3	107600001-107850000	74
ASCENDENTE	LOCAL	5	85650001-85950000	41
ASCENDENTE	LOCAL	6	142250001-142450000	50
ASCENDENTE	LOCAL	6	106850001-107000000	55
ASCENDENTE	LOCAL	7	87000001-87350000	26
ASCENDENTE	LOCAL	9	75350001-75600000	48

ES 2 729 554 T3

(continuación)

Dirección	Normalización	HS	Región	Orden
ASCENDENTE	LOCAL	10	68500001-69050000	61
ASCENDENTE	LOCAL	17	42750001-43100000	78
ASCENDENTE	LOCAL	19	19850001-20300000	66
ASCENDENTE	LOCAL	20	8000001-8250000	78
DESCENDENTE	LOCAL	1	180850001-181100000	61
DESCENDENTE	LOCAL	2	234900001-235400000	6
DESCENDENTE	LOCAL	2	105100001-105400000	45
DESCENDENTE	LOCAL	3	53950001-54200000	16
DESCENDENTE	LOCAL	4	183900001-184300000	4
DESCENDENTE	LOCAL	5	173400001-173700000	58
DESCENDENTE	LOCAL	6	163600001-163850000	66
DESCENDENTE	LOCAL	7	129100001-129550000	42
DESCENDENTE	LOCAL	7	65150001-65350000	59
DESCENDENTE	LOCAL	7	98000001-98650000	64
DESCENDENTE	LOCAL	7	153600001-153950000	78
DESCENDENTE	LOCAL	8	10200001-11400000	1
DESCENDENTE	LOCAL	8	30200001-30700000	23
DESCENDENTE	LOCAL	8	124600001-124950000	40
DESCENDENTE	LOCAL	9	100150001-100650000	15
DESCENDENTE	LOCAL	10	500001-800000	28
DESCENDENTE	LOCAL	10	123650001-124200000	20
DESCENDENTE	LOCAL	10	114450001-114750000	66
DESCENDENTE	LOCAL	12	127350001-128100000	3
DESCENDENTE	LOCAL	15	72150001-72400000	71
DESCENDENTE	LOCAL	16	78050001-78250000	26

REIVINDICACIONES

1. Un método para el diagnóstico o el cribado del cáncer colorrectal en un paciente, que comprende detectar, en una muestra que es sangre, suero o plasma de dicho paciente,
- 5 (a) el nivel total de todos los ADN extracelulares en circulación, cada uno con una secuencia que está dentro de la misma región cromosómica única designada como "ASCENDENTE" en la Tabla 2; y correlacionar un nivel total aumentado con una probabilidad mayor de que dicho paciente tenga cáncer colorrectal cuando el nivel total del ADN extracelular que tiene una secuencia que está dentro de la misma región cromosómica única es al menos 1, 2, 3, 4, 5, 10, 15, o 20 desviaciones estándar mayor que un valor de índice de sujetos normales; o
- 10 (b) el nivel total de todos los ADN extracelulares en circulación, cada uno con una secuencia que está dentro de la misma región cromosómica única designada como "DESCENDENTE" en la Tabla 2; y correlacionar un nivel total disminuido con una probabilidad mayor de que dicho paciente tenga cáncer colorrectal cuando el nivel total del ADN extracelular que tiene una secuencia que está dentro de la misma región cromosómica única es al menos 1, 2, 3, 4, 5, 10, o 20 desviaciones estándar menor que un valor de índice de sujetos normales;
- 15 en el que las posiciones de los nucleótidos en las regiones cromosómicas en la Tabla 2 están numeradas según la versión del genoma del National Center for Biotechnology Information human genome hg18/versión 36.1 publicada en marzo de 2006.
- 20 2. El método de la reivindicación 1, en el que la etapa de detección comprende la secuenciación del CNA a partir de la muestra del paciente.
3. El método de la reivindicación 1, en el que la etapa de detección comprende poner en contacto al menos una sonda que es selectiva para una región cromosómica expuesta en la Tabla 2 con una muestra de ADN obtenida de la muestra del paciente en condiciones en las que la sonda se hibrida selectivamente con una secuencia diana presente en la región cromosómica; y detectar la hibridación de la sonda.
- 25 4. El método de la reivindicación 3, en el que la sonda está unida a una superficie sólida.
- 30 5. El método de la reivindicación 3 o 4, que comprende además poner en contacto la muestra de ADN con al menos 20, 25, 30, 50, 55, 60, 65, 70, 75, 80 u 81 sondas, en el que cada sonda es selectiva para una región cromosómica expuesta en la Tabla 2.

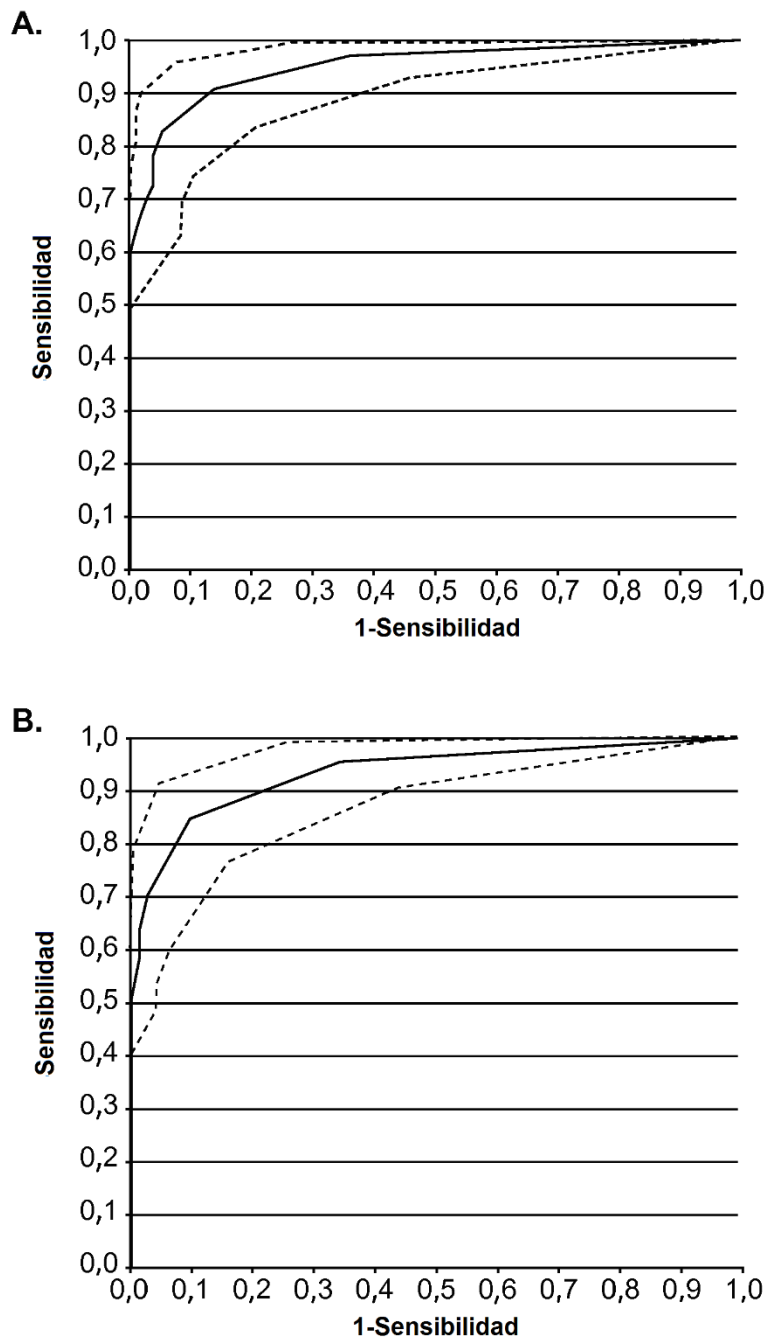


FIG. 1

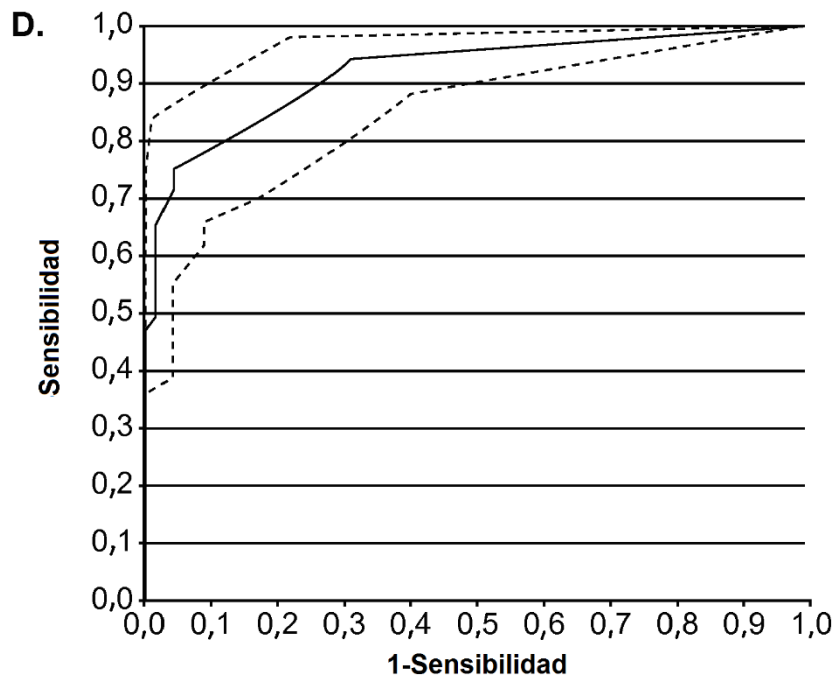
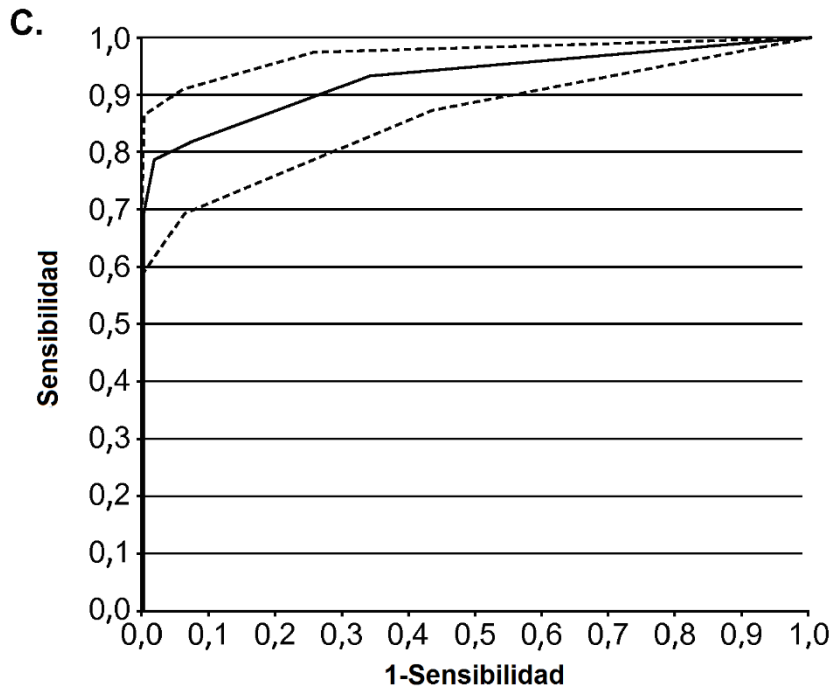


FIG. 1 (Cont.)