

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 812 753**

51 Int. Cl.:

C12Q 1/68 (2008.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **26.03.2015 PCT/US2015/022749**

87 Fecha y número de publicación internacional: **08.10.2015 WO15153283**

96 Fecha de presentación y número de la solicitud europea: **26.03.2015 E 15774156 (2)**

97 Fecha y número de publicación de la concesión europea: **27.05.2020 EP 3126529**

54 Título: **Detección de neoplasma colorectal**

30 Prioridad:

31.03.2014 US 201461972942 P
10.04.2014 US 201461977954 P

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
18.03.2021

73 Titular/es:

**MAYO FOUNDATION FOR MEDICAL EDUCATION
AND RESEARCH (50.0%)**
200 First Street S.W.
Rochester, MN 55905, US y
**EXACT SCIENCES DEVELOPMENT COMPANY,
LLC (50.0%)**

72 Inventor/es:

AHLQUIST, DAVID ALAN;
TAYLOR, WILLIAM RUSSELL;
MAHONEY, DOUGLAS W.;
LIDGARD, GRAHAM P. y
ALLAWI, HATIM T.

74 Agente/Representante:

IZQUIERDO BLANCO, María Alicia

ES 2 812 753 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

Detección de neoplasma colorectal

5 **CAMPO DE INVENCION**

[0001] Aquí se incluye tecnología relativa a la detección de neoplasia y en particular, pero no exclusivamente, a kits para la detección de lesiones premalignas y tumores malignos tales como cáncer colorrectal.

10 **ANTECEDENTES**

[0002] El cáncer colorrectal sigue siendo el segundo cáncer más común en hombres y mujeres estadounidenses combinados (Siegel R, et al., CA Cáncer J Clin 2013; 63: 11-30). La biología subyacente de la progresión de la lesión precursora al cáncer se presta favorablemente a la detección (Vogelstein B, et al., Science 2013; 339: 1546-58). La evidencia apoya y las pautas avalan cualquiera de varias pruebas y estrategias (Levin B, et al., Gastroenterology 2008; 134: 1570-95; Rex DK, et al., Am J Gastroenterol 2009; 104: 739-50; Karl J, et al., Clin Gastroenterol Hepatol 2008; 6: 1122-8). Desde una perspectiva social, el cribado se considera rentable (Karl J, et al., Clin Gastroenterol Hepatol 2008; 6: 1122-8; Heitman SJ, et al., PLoS Med 2010; 7: e1000370; Parekh M, et al., Aliment Pharmacol Ther 2008; 27: 697-712; Sharaf RN, et al., Am J Gastroenterol 2013; 108: 120-32).

[0003] Surge el cáncer colorrectal a partir de alteraciones genéticas y epigenéticas acumuladas, proporcionando una base para el análisis de las heces para los cambios específicos de tumor (Berger BM, et al, Pathology 2012; 44: 80-8). Estudios previos a gran escala de pruebas de ADN a base de heces de primera generación en el entorno de detección demostraron solo una sensibilidad justa para el cáncer colorrectal y baja sensibilidad para adenomas avanzados (Ahlquist DA, et al., Ann Intern Med 2008; 149: 441-50, W81; Imperiale TF, et al., N Engl J Med 2004; 351: 2704-14). Desde entonces se han incorporado importantes avances, incluido un tampón estabilizador (Boynton KA, et al., Clin Chem 2003; 49: 1058-65; Zou H, et al., Cancer Epidemiol Biomarkers Prev 2006; 15: 1115-9), más marcadores discriminantes (Ahlquist DA, et al., Gastroenterology 2012; 142: 248-56; Bardan E, et al., Israel journal of medical sciences 1997; 33: 777-80), plataformas con mayor sensibilidad analítica (Ahlquist DA, et al. al., Gastroenterology 2012; 142: 248-56; Aronchick CA, et al., Gastrointestinal endoscopy 2000; 52: 346-52), determinación de resultados utilizando un análisis de regresión logística en lugar de valores de marcadores individuales y automatización.

[0004] Aunque el cribado reduce la mortalidad del cáncer colorrectal (Mandel JS, et al, N Engl J Med 1993, 328: 1365-1371; Hardcastle JD, et al, Lancet 1996, 348: 1472-7; Kronborg O, et al., Scand J Gastroenterol. 2004, 39: 846-51; Winawer SJ, et al., J Natl Cancer Inst. 1993, 85: 1311-8; Singh H, et al., JAMA.2006, 295: 2366-73), las reducciones observadas han sido modestas (Singh H, et al., JAMA. 2006; 295: 2366-73; Heresbach D, et al., Eur J Gastroenterol Hepatol. 2006, 18: 427-33) y más de la mitad de los adultos en los Estados Unidos no se han hecho pruebas de detección (Meissner HI, Biomarcadores de Epidemiol de Cáncer Prev. 2006, 15: 389-94).

[0005] Un enfoque para la detección del cáncer emergente implica el ensayo de alteraciones del ADN específicos de tumores en muestras corporales de pacientes con cáncer, tales como heces, suero y orina (Osborn NK, Ahlquist DA Gastroenterology. 2005; 128: 192-206; Ahlquist DA, et al., Gastroenterology 2000; 119: 1219-27; Ahlquist DA, et al., Gastroenterology 2002; 122: Suppl A40; Chen WD, et al., J Natl Cancer Inst 2005; 97: 1124-32; Zou H, et al., Cancer Epidemiol Biomarkers Prev 2006; 15: 1115-9; Zou HZ, Clin Cancer Res 2002; 8: 188-91; Hoque MO, J Clin Oncol 2005; 23: 6569-75; Belinsky SA, et al., Cancer Res 2006; 66: 3338-44; Itzkowitz SH, et al., Clin Gastroenterol Hepatol 2007; 5: 111-7; Kann L, et al., Clin Chem 2006; 52: 2299-302). Es importante seleccionar marcadores con alta precisión si se quiere lograr eficiencia y efectividad en una aplicación de detección de cáncer. Debido a la heterogeneidad molecular de la neoplasia colorrectal, las altas tasas de detección a menudo requieren un panel de marcadores.

[0006] Varios genes metilados se han detectado en las heces y las muestras de suero/plasma de pacientes con cáncer colorrectal (Ahlquist DA, Gastroenterology 2002; 122: Suppl A40; Chen WD, et al, J Natl Cancer Inst 2005; 97: 1124-32; Zou HZ, et al., Clin Cancer Res 2002; 8: 188-91; Itzkowitz SH, et al., Clin Gastroenterol Hepatol 2007; 5: 111-7; Petko Z, et al., Clin Cancer Res 2005; 11: 1203-9; Muller HM, et al., Lancet 2004; 363: 1283-5; Leung WK, et al., Clin Chem 2004; 50: 2179-82; Ebert MP, et al., Gastroenterology 2006; 131: 1418-30; Grady WM, et al., Cancer Res 2001; 61: 900-2). Mientras que se han encontrado algunos genes metilados en la mayoría de los cánceres colorrectales, el rendimiento de los ensayos basados en fluidos corporales sigue siendo subóptimo (Ahlquist DA, et al., Gastroenterology 2002; 122: Suppl A40; Chen WD, et al., J Natl Cancer Inst 2005; 97: 1124-32; Zou H, et al., Cancer Epidemiol Biomarkers Prev 2006; 15: 1115-9; Zou HZ, Clin Cancer Res 2002; 8: 188-91; Belinsky SA, et al., Cancer Res 2006; 66: 3338-44; Itzkowitz SH, et al., Clin Gastroenterol Hepatol 2007; 5: 111-7; Kann L, et al., Clin Chem 2006; 52: 2299-302; Petko Z, et al., Clin Cancer Res 2005; 11: 1203-9; Muller HM, et al., Lancet 2004; 363: 1283-5; Leung WK, et al., Clin Chem 2004; 50: 2179-82; Ebert MP, et al., Gastroenterology 2006; 131: 1418-30; Grady WM, et al., Cancer Res 2001; 61: 900-2).

[0007] Se necesitan herramientas más precisas, fáciles de usar, y ampliamente distribuibles para mejorar la eficacia de detección, la aceptabilidad, y el acceso.

[0008] El documento WO 2012/088298 A2 describe métodos y kits para evaluar el riesgo de metástasis en un paciente con cáncer usando un clasificador para el fenotipo de metilador de isla CpG.

5 **[0009]** Naumov et al., Epigenetics 2013, vol,8, nº 9, páginas 921-934 describe el análisis a escala del genoma de la metilación del ADN en cáncer colorrectal utilizando matrices Infinium HumanMethylation450 BeadChip.

SUMARIO

10 **[0010]** El ADN metilado se ha estudiado como una clase potencial de biomarcadores en los tejidos de la mayoría de los tipos de tumores. En muchos casos, las metiltransferasas de ADN agregan un grupo metilo al ADN en los sitios de la isla de citosina-fosfato-guanina (CpG) como un control epigenético de la expresión génica. En un mecanismo biológicamente atractivo, se cree que los eventos de metilación adquiridos en regiones promotoras de genes supresores de tumores silencian la expresión, contribuyendo así a la oncogénesis. La metilación del ADN puede ser una herramienta de diagnóstico más estable química y biológicamente que la expresión de ARN o proteína (Laird (2010) Nat Rev Genet 11: 191-203). Además, en cánceres como el cáncer de colon esporádico, los marcadores de metilación ofrecen una excelente especificidad y son más ampliamente informativos y sensibles que las mutaciones de ADN individuales (Zou et al (2007) Cancer Epidemiol Biomarkers Prev 16: 2686-96).

20 **[0011]** El análisis de las islas CpG ha producido hallazgos importantes cuando se aplica a modelos animales y líneas celulares humanas. Por ejemplo, Zhang y sus colegas encontraron que los amplicones de diferentes partes de la misma isla de CpG pueden tener diferentes niveles de metilación (Zhang et al. (2009) PLoS Genet 5: e1000438). Además, los niveles de metilación se distribuyeron bimodalmente entre secuencias altamente metiladas y no metiladas, lo que respalda aún más el patrón binario de la actividad de la metiltransferasa de ADN (Zhang et al. (2009) PLoS Genet 5: e1000438). El análisis de los tejidos murinos in vivo y las líneas celulares in vitro demostraron que solo alrededor del 0,3% de los promotores de alta densidad de CpG (HCP, definidos como que tenían >7% de secuencia de CpG dentro de una región de 300 pares de bases) estaban metilados, mientras que las áreas de baja densidad de CpG (LCP, definido como que tiene <5% de secuencia de CpG dentro de una región de 300 pares de bases) tendió a ser metilado con frecuencia en un patrón dinámico específico de tejido (Meissner et al. (2008) Nature 454: 766-70). Los PS incluyen promotores de genes ubicuos de limpieza y genes de desarrollo altamente regulados. Entre los sitios HCP metilados a >50% había varios marcadores establecidos como Wnt 2, NDRG2, SFRP2 y BMP3 (Meissner et al. (2008) Nature 454: 766-70).

35 **[0012]** Genes metilados se han detectado en la sangre y las heces de los pacientes con cáncer colorrectal y se han propuesto como marcadores de cribado de candidatos (Ahlquist DA, et al, Gastroenterology 2002; 122: Suppl. A40; Chen WD, et al, J Natl Cancer Inst 2005; 97: 1124-32; Zou HZ, Clin Cancer Res 2002; 8: 188-91; Itzkowitz SH, et al., Clin Gastroenterol Hepatol 2007; 5: 111-7; Kann L, et al., Clin Chem 2006; 52: 2299-302; Petko Z, et al., Clin Cancer Res 2005; 11: 1203-9; Muller HM, et al., Lancet 2004; 363: 1283-5; Leung WK, et al., Clin Chem 2004; 50: 2179-82; Ebert MP, et al., Gastroenterology 2006; 131: 1418-30; Grady WM, et al., Cancer Res 2001; 61: 900-2).

40 **[0013]** Zou y sus colegas, por ejemplo, genes evaluados frecuentemente metilados en la neoplasia colorrectal para identificar los las más discriminantes (Zou, et al, 2007 Cancer Epidemiol Biomarkers Prev. 16 (12): 2686-2696). Se seleccionaron cuatro genes específicamente metilados en el cáncer colorrectal (proteína morfogenética ósea 3 (BMP3), EYA2, homeobox-4 (ALX4) y vimentina sin arista) de 41 genes candidatos y evaluados en 74 cánceres, 62 adenomas y 70 epitelios normales. El estado de metilación se analizó cualitativa y cuantitativamente y se confirmó mediante secuenciación genómica de bisulfito. El efecto de la metilación en la expresión génica se evaluó en cinco líneas celulares de cáncer de colon. Las mutaciones K-ras y BRAF se detectaron mediante secuenciación. La metilación de BMP3, EYA2, ALX4 o vimentina se detectó respectivamente en 66%, 66%, 68% y 72% de los cánceres; 74%, 48%, 89% y 84% de adenomas; y 7%, 5%, 11% y 11% de epitelios normales (P <0,01, cáncer o adenoma versus normal). Se concluyó que los genes BMP3, EYA2, ALX4 y vimentina están metilados en la mayoría de las neoplasias colorrectales, pero rara vez en los epitelios normales.

55 **[0014]** El cribado del cáncer está en necesidad de un panel de marcadores o un marcador para el cáncer colorrectal que es ampliamente informativo y exhibe una alta especificidad para el cáncer colorrectal en el nivel de tejido cuando es interrogado en muestras tomadas de un sujeto (por ejemplo, una muestra de heces; una muestra de tejido colorrectal).

60 **[0015]** Por consiguiente, en el presente documento se proporciona tecnología para marcadores de detección de cáncer colorrectal que proporcionan una alta relación señal/ruido y un bajo nivel de fondo cuando se detectan de muestras tomadas de un sujeto (por ejemplo, una muestra de heces; una muestra de tejido colorrectal; muestra de suero; sangre o producto sanguíneo).

65 **[0016]** En particular, la presente invención se refiere a un kit que comprende un colector de muestra de heces para la obtención de una muestra de heces de un sujeto; reactivos para aislar un ácido nucleico de la muestra; un reactivo de bisulfito; y oligonucleótidos que se unen específicamente a una región genética que tiene las coordenadas del cromosoma 3 143119999-143120158.

[0017] En los experimentos realizados durante el curso del desarrollo de formas de realización de la presente descripción, los marcadores se identificaron en unos estudios de casos y de control mediante la comparación del estado de metilación de marcadores de ADN a partir de tejido colorrectal de sujetos con neoplasia colorrectal, adenoma, y/o pólipos de sésiles serrados (SSP) al estado de metilación de los mismos marcadores de ADN de sujetos de control (p. ej., tejido normal como colon normal) (véanse, Ejemplos 1-2, Tablas 1-5).

[0018] Por ejemplo, los marcadores y/o paneles de marcadores (por ejemplo, una región cromosómica que tiene una anotación seleccionada de FLI1, OPLAH, DTX1, MATK, SFMBT2 región 2, KCNK12, región Vav3 1, SFMBT2 región 3, PPP2R5C, CHST2 región 2, PKIA, PDGFD, ELOVL2, CHST2 región 1, SFMBT2 región 1, QKI, VAV3 región 2 y SLC8A3) se identificaron en un estudio de casos y controles comparando el estado de metilación de los marcadores de ADN (p. ej., del tejido colorrectal de sujetos con neoplasia colorrectal y/o adenoma) al estado de metilación de los mismos marcadores de ADN de sujetos de control (p. ej., tejido normal como el colon normal) (ver Ejemplo 1 y Tabla 1).

[0019] Los marcadores y/o paneles de marcadores (por ejemplo, una región cromosómica que tiene una anotación seleccionada de BMP3, NDRG4, PDGFG, CHST2 y SFMBT2) se identificaron en estudios de casos y controles comparando el estado de metilación de los marcadores de ADN del tejido colorrectal de sujetos con enfermedad inflamatoria intestinal y cáncer colorrectal al estado de metilación de los mismos marcadores de ADN de sujetos de control (p. ej., tejido normal como colon normal) (véase el Ejemplo 1 y la Tabla 2).

[0020] Además, 185, 244, y 111 marcadores de metilación de ADN específicos para los cánceres colorrectales, adenomas grandes, y pólipos de sésiles serrados, respectivamente, se identificaron (véase el Ejemplo 2 y en las Tablas 3, 4 y 5). Junto con los casos de cáncer colorrectal, se secuenciaron casos de adenoma grande y pólipos de sésiles serrados, mucosa colónica normal y ADN normal de glóbulos blancos.

[0021] Los experimentos adicionales realizados durante el curso del desarrollo de formas de realización para la presente descripción demostraron NDRG4, BMP3, OPLAH, FLI1, PDGFD, CHST 7889, SFMBT2_895, SFMBT2_896, SFMBT2_897, CHST2_7890, Vav3, y DTX1 como marcadores eficaces para la detección de cáncer colorrectal dentro de muestras de heces (ver, Ejemplo 3 y Tablas 6 y 7).

[0022] Como se describe en el presente documento, la tecnología proporciona un número de marcadores y subconjuntos de ADN metilado de los mismos (por ejemplo, series de 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 o más marcadores) con alta discriminación para la neoplasia colorrectal (p. ej., cáncer colorrectal, adenoma, SSP). Los experimentos aplicaron un filtro de selección a marcadores candidatos para identificar marcadores que proporcionan una alta relación señal/ruido y un bajo nivel de fondo para proporcionar una alta especificidad, por ejemplo, al analizar medios distantes (por ejemplo, heces, sangre, orina, tejido metastásico, etc.) para fines de detección o diagnóstico de cáncer colorrectal. Como tal, la tecnología proporciona marcadores específicos y combinaciones de marcadores para fines de detección o diagnóstico de cáncer colorrectal.

[0023] En algunas realizaciones, la tecnología está relacionada con la evaluación de la presencia y estado de metilación de uno o más de los marcadores identificados en la presente en una muestra biológica. Estos marcadores comprenden una o más regiones metiladas diferencialmente (DMR) como se discute en el presente documento, por ejemplo, como se proporciona en las Tablas 1-6. El estado de metilación se evalúa en realizaciones de la tecnología. Como tal, la tecnología proporcionada aquí no está restringida en el método por el cual se mide el estado de metilación de un gen. Por ejemplo, en algunas realizaciones, el estado de metilación se mide mediante un método de exploración del genoma. Por ejemplo, un método implica escaneo genómico de restricción de referencia (Kawai et al. (1994) Mol. Cell. Biol. 14: 7421-7427) y otro ejemplo involucra PCR cebada arbitrariamente sensible a la metilación (Gonzalzo et al. (1997) Cancer Res 57: 594-599). En algunas realizaciones, los cambios en los patrones de metilación en sitios CpG específicos se controlan mediante digestión de ADN genómico con enzimas de restricción sensibles a la metilación, seguido de análisis Southern de las regiones de interés (método Southern de digestión). En algunas realizaciones, el análisis de los cambios en los patrones de metilación implica un proceso basado en PCR que implica la digestión del ADN genómico con enzimas de restricción sensibles a la metilación antes de la amplificación por PCR (Singer-Sam et al. (1990) Nucl. Acids Res. 18: 687). Además, se han informado otras técnicas que utilizan el tratamiento con bisulfito de ADN como punto de partida para el análisis de metilación. Estos incluyen PCR específica de metilación (MSP) (Herman et al. (1992) Proc. Natl. Acad. Sci. EE.UU. 93: 9821-9826) y la digestión con enzimas de restricción de productos de PCR amplificados a partir de ADN convertido con bisulfito (Sadri y Hornsby (1996) Nucl. Acids Res. 24: 5058-5059; y Xiong y Laird (1997) Nucl. Acids Res. 25: 2532-2534). Se han desarrollado técnicas de PCR para la detección de mutaciones genéticas (Kuppuswamy et al. (1991) Proc. Natl. Acad. Sci. EE.UU. 88: 1143-1147) y la cuantificación de la expresión alélica específica (Szabo y Mann (1995) Genes Dev. 9: 3097-3108 y Singer-Sam et al. (1992) PCR Methods Appl. 1: 160-163). Dichas técnicas utilizan cebadores internos, que se unen a una plantilla generada por PCR y terminan inmediatamente en 5' del nucleótido único a analizar. Los métodos que usan un "ensayo cuantitativo de Ms-SNuPE" como se describe en la Patente de EE.UU. 7.037,650 se usan en algunas realizaciones.

[0024] Tras la evaluación de un estado de metilación, el estado de metilación se expresa a menudo como la fracción

o porcentaje de hebras individuales de ADN que está metilado en un sitio en particular (por ejemplo, en un solo nucleótido, en una región o locus particular, en una secuencia de interés más larga, por ejemplo, hasta un ~100 pb, 200 pb, 500 pb, 1000 pb subsecuencia de ADN o más) con respecto a la población total de ADN en la muestra que comprende ese sitio particular. Tradicionalmente, la cantidad de ácido nucleico no metilado se determina por PCR usando calibradores. Luego, una cantidad conocida de ADN se trata con bisulfito y la secuencia específica de metilación resultante se determina usando una PCR en tiempo real u otra amplificación exponencial, por ejemplo, un ensayo QuARTS (por ejemplo, según lo dispuesto por la Patente de EE.UU. N° 8,361,720; y Solicitud de Patente de EE.UU. N°s 2012/0122088 y 2012/0122106).

[0025] Por ejemplo, en algunas realizaciones métodos comprenden la generación de una curva estándar para la diana no metilada mediante el uso de estándares externos. La curva estándar se construye a partir de al menos dos puntos y relaciona el valor de Ct en tiempo real para el ADN no metilado con estándares cuantitativos conocidos. Luego, se construye una segunda curva estándar para la diana metilada a partir de al menos dos puntos y estándares externos. Esta segunda curva estándar relaciona el Ct para ADN metilado con estándares cuantitativos conocidos. A continuación, los valores de Ct de la muestra de prueba se determinan para las poblaciones metiladas y no metiladas y los equivalentes genómicos de ADN se calculan a partir de las curvas estándar producidas por los dos primeros pasos. El porcentaje de metilación en el sitio de interés se calcula a partir de la cantidad de ADN metilados en relación con la cantidad total de ADN en la población, por ejemplo, $(\text{número de ADN metilados}) / (\text{número de ADN metilados} + \text{número de ADN no metilados}) \times 100$.

[0026] Según otro aspecto de la presente descripción, neoplasia de una muestra biológica se indica cuando una relación de metilación de uno o más marcadores de metilación de ADN relativo a un nivel de número de copias de ADN tratado con bisulfito de un gen de referencia es diferente, en donde el uno o más marcadores de metilación de ADN comprenden una base en una región metilada diferencialmente (DMR) como se proporciona aquí. La relación de metilación incluye la relación del nivel de metilación del marcador de metilación del ADN y el nivel de una región en un gen de referencia determinado por los mismos medios utilizados para la determinación del nivel de metilación del biomarcador. Usualmente, la relación de metilación está representada por la relación del nivel de metilación del marcador de metilación de ADN y el nivel de una región en un gen de referencia determinado por los mismos medios utilizados para la determinación del nivel de metilación del marcador de metilación de ADN.

[0027] En algunas realizaciones, la relación de la metilación es la relación entre el nivel de metilación de un marcador de la metilación del ADN y el nivel de una región de un gen de referencia, ambos de los cuales se midió cuantitativamente usando polimerasa en tiempo real de reacción en cadena (PCR). Por ejemplo, el nivel de metilación de un marcador de metilación de ADN de una muestra de un sujeto puede medirse cuantitativamente usando un par de cebadores y una sonda de oligonucleótidos, donde un cebador, ambos cebadores, la sonda de oligonucleótidos o ambos cebadores y la sonda de oligonucleótidos son capaces de distinguir entre ácido nucleico metilado y no metilado, por ejemplo, después de que el ácido nucleico sea modificado por un agente modificador, por ejemplo, bisulfito que convierte la citosina no metilada en un ácido nucleico convertido.

[0028] La región de un gen de referencia puede ser cualquier región de un gen que tiene uno o más sitios o regiones que están desprovistos de sitios de metilación, por ejemplo, desprovistas de dinucleótidos CpG. Por ejemplo, la región de un gen de referencia puede ser una región que tiene dos sitios de unión de cebadores para la amplificación, tales como PCR que carecen de dinucleótidos CpG o una región que tiene al menos un sitio de unión de sonda oligonucleotídica específica para PCR en tiempo real que está desprovista de dinucleótidos CpG. La región de un gen de referencia puede ser una región del gen MYOD. La región de un gen de referencia puede ser una región del gen ACTB. La región de un gen de referencia puede ser una región que no está frecuentemente sujeta a alteraciones en el número de copias, como la amplificación o eliminación de genes.

[0029] En general, el nivel de una región de un gen de referencia se mide cuantitativamente usando PCR en tiempo real con cebadores y sondas específicos que se unen específicamente a sitios después de la conversión de bisulfito pero sin discriminar directa o indirectamente el estado de metilación de los sitios.

[0030] En ciertas realizaciones, se proporcionan métodos para detectar neoplasia en un sujeto. Tales métodos comprenden, por ejemplo, obtener una muestra que comprende ADN de un sujeto; tratar el ADN obtenido con un reactivo que modifica selectivamente los residuos de citosina no metilados en el ADN obtenido para producir residuos modificados pero que no modifica los residuos de citosina metilados; determinar el nivel de metilación de uno o más marcadores de metilación de ADN en el ADN que ha sufrido el tratamiento del paso b), en donde uno o más marcadores de metilación de ADN comprenden una base en una región metilada diferencialmente (DMR) como se proporciona aquí; comparar el nivel de metilación determinado de uno o más marcadores de metilación de ADN con referencias de nivel de metilación para uno o más marcadores de metilación de ADN para sujetos que no tienen neoplasia; e identificar al sujeto con neoplasia cuando el estado de metilación de o más de los marcadores de metilación del ADN es diferente al estado de metilación del marcador analizado en un sujeto que no tiene una neoplasia.

[0031] En algunas realizaciones, una determinación de la metilación elevada en uno o más de los marcadores de

metilación del ADN comprende una determinación de la metilación alterada dentro de una región seleccionada del grupo que consiste de una isla CpG y una orilla isla CpG.

5 **[0032]** En algunas realizaciones, una determinación de la metilación elevada dentro de dicha isla CpG o CpG en tierra comprende metilación elevada dentro de una región de codificación o una región reguladora del marcador de la metilación del ADN.

10 **[0033]** En algunas realizaciones, la determinación del nivel de metilación de uno o más marcadores de metilación del ADN en el ADN que hayan sido sometidos al tratamiento del paso b) comprende la determinación de la puntuación de la metilación y/o la frecuencia de metilación del uno o más marcadores de metilación del ADN.

[0034] En algunas realizaciones, el tratamiento del paso b) se lleva a cabo a través de modificación con bisulfito del ADN obtenido.

15 **[0035]** En algunas realizaciones, la determinación del nivel de metilación de uno o más marcadores de metilación del ADN en el ADN que hayan sido sometidas al tratamiento del paso b) se consigue mediante una técnica seleccionada del grupo que consiste en PCR de metilación específica, PCR cuantitativa de metilación específica, análisis de enzimas de restricción de ADN sensibles a la metilación, secuenciación cuantitativa de bisulfito y PCR de secuenciación genómica de bisulfito.

20 **[0036]** En algunas realizaciones, el neoplasma es un cáncer colorrectal, un gran adenoma colorrectal y/o un pólipo de sésiles serrado.

25 **[0037]** En ciertas realizaciones, se proporcionan métodos para detectar neoplasia en un sujeto. Dichas realizaciones comprenden, por ejemplo, determinar una relación de metilación de una muestra de un sujeto, en donde la relación de metilación es el nivel de metilación de una región tratada con bisulfito de uno o más marcadores de metilación de ADN en relación con un nivel de copia de ADN tratada con bisulfito número de un gen de referencia, en donde el uno o más marcadores de metilación del ADN comprenden una base en una región metilada diferencialmente (DMR) como se proporciona aquí, en donde el gen de referencia es MYOD o ACTB, identificando al sujeto con neoplasia cuando la relación de metilación de uno o más de los marcadores de metilación del ADN es diferente de la relación de metilación del marcador respectivo analizado en un sujeto que no tiene una neoplasia.

35 **[0038]** En algunas realizaciones, el nivel de metilación se determina mediante el uso de reacción en cadena de la polimerasa en tiempo real (PCR). En algunas realizaciones, el nivel de metilación se determina usando la reacción en cadena de la polimerasa (PCR) en tiempo real, en donde al menos un cebador usado en la PCR es capaz de distinguir entre ácido nucleico no metilado y metilado. En algunas realizaciones, el nivel de metilación se determina usando la reacción en cadena de la polimerasa (PCR) en tiempo real, en la que ambos cebadores utilizados en la PCR son capaces de distinguir entre ácido nucleico no metilado y metilado. En algunas realizaciones, el nivel de metilación se determina usando la reacción en cadena de la polimerasa (PCR) en tiempo real, en la que una sonda utilizada en la PCR es capaz de distinguir entre ácido nucleico no metilado y metilado. En algunas realizaciones, el nivel de metilación se determina usando la reacción en cadena de la polimerasa (PCR) en tiempo real, en la que los cebadores y una sonda utilizados en la PCR son capaces de distinguir entre el ácido nucleico metilado y no metilado. En algunas realizaciones, el nivel de la región en el gen de referencia se determina usando la reacción en cadena de la polimerasa (PCR) en tiempo real. En algunas realizaciones, el nivel de la región en el gen de referencia se determina usando la reacción en cadena de la polimerasa (PCR) en tiempo real, en donde la región contiene un primer y segundo sitio de unión del cebador y un sitio de unión de la sonda y en donde el primer y segundo sitio de unión del cebador y el sitio de unión de la sonda carecen de dinucleótidos CpG. En algunas realizaciones, la región en el gen de referencia carece de dinucleótidos CpG.

45 **[0039]** También se proporcionan en este documento composiciones y kits para la práctica de los métodos. Por ejemplo, en algunas realizaciones, los reactivos (por ejemplo, cebadores, sondas) específicos para uno o más marcadores se proporcionan solos o en conjuntos (por ejemplo, conjuntos de pares de cebadores para amplificar una pluralidad de marcadores). También se pueden proporcionar reactivos adicionales para realizar un ensayo de detección (p. ej., enzimas, tampones, controles positivos y negativos para realizar QuARTS, PCR, secuenciación, bisulfito u otros ensayos). En algunas realizaciones, se proporcionan los kits que contienen uno o más reactivos necesarios, suficientes o útiles para realizar un método. También se proporcionan mezclas de reacciones que contienen los reactivos. Además se proporcionan conjuntos de reactivos de mezcla maestra que contienen una pluralidad de reactivos que pueden agregarse entre sí y/o a una muestra de prueba para completar una mezcla de reacción.

60 **[0040]** En algunas realizaciones, la tecnología descrita en el presente documento está asociada con una máquina programable diseñada para llevar a cabo una secuencia de operaciones aritméticas o lógicas como proporcionada por los métodos descritos en el presente documento. Por ejemplo, algunas realizaciones de la tecnología están asociadas con (por ejemplo, implementadas en) software de computadora y/o hardware de computadora. En un aspecto, la tecnología se refiere a una computadora que comprende una forma de memoria, un elemento para realizar operaciones aritméticas y lógicas, y un elemento de procesamiento (por ejemplo, un microprocesador) para ejecutar una serie de instrucciones (por ejemplo, un método como se proporciona aquí) para leer, manipular y almacenar datos.

En algunas realizaciones, un microprocesador es parte de un sistema para determinar un estado de metilación (por ejemplo, de uno o más DMR, por ejemplo, como se proporciona en las Tablas 1-6); comparar estados de metilación (por ejemplo, de uno o más DMR, por ejemplo, como se proporciona en las Tablas 1-6); generar curvas estándar; determinar un valor de Ct; calcular una fracción, frecuencia o porcentaje de metilación (p. ej., de uno o más DMR, p. ej. como se proporciona en las Tablas 1-6); identificar una isla CpG; determinar una especificidad y/o sensibilidad de un ensayo o marcador; calcular una curva ROC y un AUC asociado; análisis de secuencia; todo como se describe en este documento o se conoce en la técnica.

[0041] En algunas realizaciones, un componente de software o hardware reciba los resultados de ensayos múltiples y determina un resultado de valor único que informar a un usuario que indica un riesgo de cáncer en base a los resultados de los múltiples ensayos (por ejemplo, la determinación del estado de metilación de DMR múltiple, por ejemplo, como se proporciona en las Tablas 1-6). Las realizaciones relacionadas calculan un factor de riesgo basado en una combinación matemática (p. ej., una combinación ponderada, una combinación lineal) de los resultados de múltiples ensayos, p. ej., determinando los estados de metilación de múltiples marcadores (como DMR múltiple, p. ej. como se proporciona en las Tablas 1-6). En algunas realizaciones, el estado de metilación de un DMR define una dimensión y puede tener valores en un espacio multidimensional y la coordenada definida por los estados de metilación de DMR múltiple es un resultado, por ejemplo, para informar a un usuario, por ejemplo, en relación con un riesgo de cáncer colorrectal.

[0042] Algunas realizaciones comprenden un medio de almacenamiento y componentes de memoria. Los componentes de la memoria (p. ej., memoria volátil y/o no volátil) encuentran uso en el almacenamiento de instrucciones (p. ej., una realización de un proceso como se proporciona aquí) y/o datos (p. ej., una pieza de trabajo como mediciones de metilación, secuencias y descripciones estadísticas asociadas con ella). Algunas realizaciones se refieren a sistemas que también comprenden una o más de una CPU, una tarjeta gráfica y una interfaz de usuario (por ejemplo, que comprende un dispositivo de salida como pantalla y un dispositivo de entrada como teclado).

[0043] Las máquinas programables asociadas con la tecnología comprenden tecnologías convencionales existentes y tecnologías en desarrollo o aún por desarrollar (por ejemplo, una computadora cuántica, una computadora química, una computadora de ADN, una computadora óptica, una computadora basada en espintrónica, etc.).

[0044] En algunas realizaciones, la tecnología comprende un cable (por ejemplo, cable metálico, de fibra óptica) o medio de transmisión inalámbrica para transmitir datos. Por ejemplo, algunas realizaciones se refieren a la transmisión de datos a través de una red (por ejemplo, una red de área local (LAN), una red de área amplia (WAN), una red ad-hoc, Internet, etc.). En algunas realizaciones, las máquinas programables están presentes en una red como pares y en algunas realizaciones las máquinas programables tienen una relación cliente/servidor.

[0045] En algunas realizaciones, los datos se almacenan en un medio de almacenamiento legible por ordenador tal como un disco duro, memoria flash, medios ópticos, un disquete, etc.

[0046] En algunas realizaciones, la tecnología proporcionada en el presente documento está asociado con una pluralidad de dispositivos programables que operan en concierto para realizar un método como se describe aquí. Por ejemplo, en algunas realizaciones, una pluralidad de computadoras (por ejemplo, conectadas por una red) pueden trabajar en paralelo para recopilar y procesar datos, por ejemplo, en una implementación de computación en clúster o computación en cuadrícula o alguna otra arquitectura de computadora distribuida que se basa en computadoras completas (con CPU incorporadas, almacenamiento, fuentes de alimentación, interfaces de red, etc.) conectadas a una red (privada, pública o Internet) por una interfaz de red convencional, como Ethernet, fibra óptica o mediante una tecnología de red inalámbrica.

[0047] Por ejemplo, algunas realizaciones proporcionan un equipo que incluye un medio legible por ordenador. La realización incluye una memoria de acceso aleatorio (RAM) acoplada a un procesador. El procesador ejecuta las instrucciones del programa ejecutable por computadora almacenadas en la memoria. Dichos procesadores pueden incluir un microprocesador, un ASIC, una máquina de estado u otro procesador, y pueden ser cualquiera de varios procesadores de computadora, como los procesadores de Intel Corporation of Santa Clara, California and Motorola Corporation of Schaumburg, Illinois. Dichos procesadores incluyen, o pueden estar en comunicación con medios, por ejemplo, medios legibles por computadora, que almacenan instrucciones que, cuando son ejecutadas por el procesador, hacen que el procesador realice los pasos descritos en este documento.

[0048] Las realizaciones de medios legibles por computadora incluyen, pero no se limitan a, un dispositivo electrónico, óptico, magnético u otro dispositivo de almacenamiento o transmisión capaz de proporcionar un procesador con instrucciones legibles por computadora. Otros ejemplos de medios adecuados incluyen, entre otros, un disquete, CD-ROM, DVD, disco magnético, chip de memoria, ROM, RAM, un ASIC, un procesador configurado, todos los medios ópticos, todas las cintas magnéticas u otros medios magnéticos o cualquier otro medio desde el cual un procesador de computadora pueda leer las instrucciones. Además, varias otras formas de medios legibles por computadora pueden transmitir o llevar instrucciones a una computadora, incluido un enrutador, una red pública o privada u otro dispositivo o canal de transmisión, tanto por cable como inalámbrico. Las instrucciones pueden incluir un código de

cualquier lenguaje de programación adecuado, incluidos, por ejemplo, C, C++, C#, Visual Basic, Java, Python, Perl y JavaScript.

5 [0049] Las computadoras están conectadas en algunas realizaciones a una red. Las computadoras también pueden incluir varios dispositivos externos o internos, como un ratón, un CD-ROM, DVD, un teclado, una pantalla u otros dispositivos de entrada o salida. Ejemplos de computadoras son computadoras personales, asistentes digitales, asistentes digitales personales, teléfonos celulares, teléfonos móviles, teléfonos inteligentes, buscapersonas, tabletas digitales, computadoras portátiles, dispositivos de internet y otros dispositivos basados en procesadores. En general, las computadoras relacionadas con los aspectos de la tecnología aquí provistos pueden ser cualquier tipo de plataforma basada en procesador que opere en cualquier sistema operativo, como Microsoft Windows, Linux, UNIX, Mac OS X, etc., capaz de soportar uno o más programas que comprenden la tecnología provista aquí. Algunas realizaciones comprenden una computadora personal que ejecuta otros programas de aplicación (por ejemplo, aplicaciones). Las aplicaciones pueden estar contenidas en la memoria y pueden incluir, por ejemplo, una aplicación de procesamiento de texto, una aplicación de hoja de cálculo, una aplicación de correo electrónico, una aplicación de mensajería instantánea, una aplicación de presentación, una aplicación de navegador de Internet, una aplicación de calendario/organizador y cualquier otra aplicación capaz de ser ejecutada por un dispositivo cliente.

10 [0050] Todos los tales componentes, equipos y sistemas descritos en el presente documento como asociados con la tecnología pueden ser lógicos o virtuales.

15 [0051] Por consiguiente, en este documento se proporciona tecnología relacionada con un método de detección de una neoplasia colorrectal en una muestra (por ejemplo, muestra de heces, muestra de tejido colorrectal; muestra de sangre; muestra de producto sanguíneo) obtenida de un sujeto (por ejemplo, un sujeto humano), comprendiendo el método analizar un estado de metilación de un marcador en una muestra obtenida de un sujeto; e identificar al sujeto que tiene una neoplasia colorrectal cuando el estado de metilación del marcador es diferente al estado de metilación del marcador analizado en un sujeto que no tiene una neoplasia colorrectal, en donde el marcador comprende una base en una región metilada diferencialmente (DMR) seleccionada de un grupo que consiste en un DMR como se proporciona en las Tablas 1-6. La tecnología también abarca la determinación del estado o etapa de un cáncer colorrectal, por ejemplo, en algunas realizaciones, la neoplasia es precancerosa. Algunas realizaciones proporcionan métodos que comprenden analizar una pluralidad de marcadores, por ejemplo, que comprenden analizar de 2 a 11 marcadores.

20 [0052] La tecnología no está limitada en el estado de metilación evaluado. En algunas realizaciones, evaluar el estado de metilación del marcador en la muestra comprende determinar el estado de metilación de una base. En algunas realizaciones, analizar el estado de metilación del marcador en la muestra comprende determinar el grado de metilación en una pluralidad de bases. Además, en algunas realizaciones, el estado de metilación del marcador comprende un aumento de la metilación del marcador con respecto a un estado de metilación normal del marcador. En algunas realizaciones, el estado de metilación del marcador comprende una metilación disminuida del marcador con respecto a un estado de metilación normal del marcador. En algunas realizaciones, el estado de metilación del marcador comprende un patrón diferente de metilación del marcador con respecto a un estado de metilación normal del marcador.

25 [0053] Además, en algunas realizaciones, el marcador es una región de 100 bases o menos, el marcador es una región de 500 bases o menos, el marcador es una región de 1000 bases o menos, el marcador es una región de 5000 o menos bases, o, en algunas realizaciones, el marcador es una base. En algunas realizaciones, el marcador está en un promotor de alta densidad de CpG.

30 [0054] La tecnología no está limitada por el tipo de muestra. Por ejemplo, en algunas realizaciones, la muestra es una muestra de heces, una muestra de tejido, una muestra de tejido colorrectal, una muestra de sangre (por ejemplo, plasma, suero, sangre completa), una excreción o una muestra de orina.

35 [0055] Además, la tecnología no está limitada en el método utilizado para determinar el estado de metilación. En algunas realizaciones, el ensayo comprende usar reacción en cadena de polimerasa específica de metilación, secuenciación de ácido nucleico, espectrometría de masas, nucleasa específica de metilación, separación basada en masa o captura de dianas. En algunas realizaciones, el ensayo comprende el uso de un oligonucleótido específico de metilación. En algunas realizaciones, la tecnología usa masivamente secuenciación paralela (por ejemplo, secuenciación de próxima generación) para determinar el estado de metilación, por ejemplo, secuenciación por síntesis, secuenciación en tiempo real (por ejemplo, de una sola molécula), secuenciación de emulsión de cuenta, secuenciación de nanoporos, etc.

40 [0056] La tecnología proporciona reactivos para la detección de un DMR, por ejemplo, en algunas realizaciones se proporcionan un conjunto de oligonucleótidos que comprenden las secuencias proporcionadas por la SEQ ID NOs: 1-110. En algunas realizaciones, se proporciona un oligonucleótido que comprende una secuencia complementaria a una región cromosómica que tiene una base en una DMR, por ejemplo, un oligonucleótido sensible al estado de metilación de una DMR.

[0057] La tecnología proporciona diversos paneles de marcadores, por ejemplo, en algunas realizaciones, el marcador comprende una región cromosómica que tiene una anotación que es FLI1, OPLAH, DTX1, MATK, SFMBT2 región 2, KCNK12, Vav3 región 1, SFMBT2 región 3, PPP2R5C, CHST2 región 2, PKIA, PDGFD, ELOVL2, CHST2 región 1, SFMBT2 región 1, QKI, VAV3 región 2 y SLC8A3, y eso comprende el marcador (ver Tabla 1). En algunas realizaciones, el marcador comprende una región cromosómica que tiene una anotación que es BMP3, NDRG4, PDGFG, CHST2 y SFMBT2, y que comprende el marcador (ver, Tabla 2). En algunas realizaciones, el marcador comprende una o más de las regiones cromosómicas proporcionadas en la Tabla 3 (para cáncer colorrectal), Tabla 4 (para adenoma) y Tabla 5 (para SSP).

[0058] Además, las realizaciones proporcionan un método de análisis de una DMR a partir de las Tablas 1-6. Algunas realizaciones proporcionan determinar el estado de metilación de un marcador, en donde una región cromosómica que tiene una anotación es FLI1, OPLAH, DTX1, MATK, SFMBT2 región 2, KCNK12, VAV3 región 1, SFMBT2 región 3, PPP2R5C, CHST2 región 2, PKIA, PDGFD, ELOVL2, CHST2 región 1, SFMBT2 región 1, QKI, VAV3 región 2 y SLC8A3, y/o una región cromosómica que tiene una anotación que es BMP3, NDRG4, PDGFG, CHST2 y SFMBT2 comprende el marcador. Algunas realizaciones proporcionan determinar el estado de metilación de un marcador, en donde una región cromosómica como se proporciona en las Tablas 3, 4 y/o 5 comprende el marcador.

[0059] Realizaciones de kit se describen en el presente documento, por ejemplo, un kit que comprende un reactivo de bisulfito; y un ácido nucleico de control que comprende una secuencia de una DMR seleccionada de cualquiera de las regiones cromosómicas proporcionadas en las Tablas 1-6 y que tiene un estado de metilación asociado con un sujeto que no tiene cáncer. En algunas realizaciones, los kits comprenden un reactivo de bisulfito y un oligonucleótido como se describe en el presente documento. En algunas realizaciones, los kits comprenden un reactivo de bisulfito; y un ácido nucleico de control que comprende una secuencia de una DMR seleccionada de cualquiera de las regiones cromosómicas proporcionadas en las Tablas 1-6 y que tiene un estado de metilación asociado con un sujeto que tiene cáncer colorrectal, adenoma y/o SSP. Algunas realizaciones del kit comprenden un colector de muestras para obtener una muestra de un sujeto (por ejemplo, una muestra de heces); reactivos para aislar un ácido nucleico de la muestra; un reactivo de bisulfito; y un oligonucleótido como se describe en el presente documento.

[0060] La tecnología está relacionada con formas de realización de composiciones (por ejemplo, mezclas de reacción). En algunas realizaciones, se proporciona una composición que comprende un ácido nucleico que comprende un DMR y un reactivo de bisulfito. Algunas realizaciones proporcionan una composición que comprende un ácido nucleico que comprende una DMR y un oligonucleótido como se describe en el presente documento. Algunas realizaciones proporcionan una composición que comprende un ácido nucleico que comprende una DMR y una enzima de restricción sensible a la metilación. Algunas realizaciones proporcionan una composición que comprende un ácido nucleico que comprende una DMR y una polimerasa.

[0061] Realizaciones de método adicionales relacionadas se proporcionan para la detección de una neoplasia colorrectal en una muestra obtenida de un sujeto, por ejemplo, un método que comprende la determinación de un estado de metilación de un marcador en la muestra que comprende una base en un DMR seleccionado de cualquiera de las regiones cromosómicas proporcionadas en las Tablas 1-6; comparar el estado de metilación del marcador de la muestra del sujeto con un estado de metilación del marcador de una muestra de control normal de un sujeto que no tiene cáncer; y determinar un intervalo de confianza y/o un valor p de la diferencia en el estado de metilación de la muestra del sujeto y la muestra de control normal. En algunas realizaciones, el intervalo de confianza es 90%, 95%, 97,5%, 98%, 99%, 99,5%, 99,9% o 99,99% y el valor p es 0,1, 0,05, 0,025, 0,02, 0,01, 0,005, 0,001, o 0,0001. Algunas realizaciones de métodos proporcionan pasos de hacer reaccionar un ácido nucleico que comprende una DMR con un reactivo de bisulfito para producir un ácido nucleico reaccionado con bisulfito; secuenciar el ácido nucleico reaccionado con bisulfito para proporcionar una secuencia de nucleótidos del ácido nucleico reaccionado con bisulfito; comparar la secuencia de nucleótidos del ácido nucleico reaccionado con bisulfito con una secuencia de nucleótidos de un ácido nucleico que comprende la DMR de un sujeto que no tiene cáncer para identificar diferencias en las dos secuencias; e identificar al sujeto que tiene una neoplasia colorrectal cuando hay una diferencia presente.

[0062] La tecnología proporciona sistemas para la detección de una neoplasia colorrectal en una muestra obtenida de un sujeto. Las realizaciones ejemplares de sistemas incluyen, por ejemplo, un sistema para la detección de una neoplasia colorrectal en una muestra obtenida de un sujeto, el sistema comprende un componente de análisis configurado para determinar el estado de metilación de una muestra, un componente de software configurado para comparar el estado de metilación de la muestra con una muestra de control o un estado de metilación de muestra de referencia registrado en una base de datos, y un componente de alerta configurado para alertar a un usuario de un estado de metilación asociado con cáncer. Una alerta se determina en algunas realizaciones mediante un componente de software que recibe los resultados de múltiples ensayos (p. ej., determinando los estados de metilación de múltiples marcadores, p. ej., DMR, p. ej., como se proporciona en las Tablas 1-6) y calculando un valor o resultado para informe basado en los múltiples resultados. Algunas realizaciones proporcionan una base de datos de parámetros ponderados asociados con cada DMR proporcionada en este documento para su uso en el cálculo de un valor o resultado y/o una alerta para informar a un usuario (por ejemplo, un médico, enfermera, médico, etc.). En algunas realizaciones, se informan todos los resultados de múltiples ensayos y en algunas realizaciones se usan uno o más resultados para

proporcionar una puntuación, valor o resultado basado en una combinación de uno o más resultados de múltiples ensayos que es indicativo de un riesgo de cáncer colorrectal en un sujeto.

[0063] En algunas realizaciones de sistemas, una muestra comprende un ácido nucleico que comprende una DMR. En algunas realizaciones el sistema comprende además un componente para aislar un ácido nucleico, un componente para recoger una muestra tal como un componente para recoger una muestra de heces. En algunas realizaciones, el sistema comprende secuencias de ácido nucleico que comprenden una DMR. En algunas realizaciones, la base de datos comprende secuencias de ácido nucleico de sujetos que no tienen cáncer. También se proporcionan ácidos nucleicos, por ejemplo, un conjunto de ácidos nucleicos, cada ácido nucleico tiene una secuencia que comprende una DMR. En algunas realizaciones, el conjunto de ácidos nucleicos en donde cada ácido nucleico tiene una secuencia de un sujeto que no tiene cáncer. Las realizaciones del sistema relacionadas comprenden un conjunto de ácidos nucleicos como se describe y una base de datos de secuencias de ácido nucleico asociadas con el conjunto de ácidos nucleicos. Algunas realizaciones comprenden además un reactivo de bisulfito. Y, algunas realizaciones comprenden además un secuenciador de ácido nucleico.

[0064] La tecnología está relacionada con formas de realización de composiciones (por ejemplo, mezclas de reacción). En algunas realizaciones, se proporciona una composición que comprende un ácido nucleico que comprende un DMR (por ejemplo, un DMR como se proporciona en las Tablas 1-6) y un reactivo de bisulfito. Algunas realizaciones proporcionan una composición que comprende un ácido nucleico que comprende una DMR y un oligonucleótido como se describe en el presente documento. Algunas realizaciones proporcionan una composición que comprende un ácido nucleico que comprende una DMR y una enzima de restricción sensible a la metilación. Algunas realizaciones proporcionan una composición que comprende un ácido nucleico que comprende una DMR y una polimerasa.

[0065] En ciertas realizaciones, las preocupaciones actuales de divulgación métodos de cribado para un neoplasma colorrectal en una muestra obtenida de un sujeto que tiene la enfermedad inflamatoria intestinal, comprendiendo el método 1) ensayar una metilación estado de un marcador en una muestra obtenida de un sujeto; y 2) identificar al sujeto que tiene una neoplasia cuando el estado de metilación del marcador es diferente al estado de metilación del marcador analizado en un sujeto que no tiene una neoplasia, en donde el marcador comprende una base en una región metilada diferencialmente (DMR) como se proporciona en la Tabla 2. En algunas realizaciones, la neoplasia es cáncer colorrectal y/o displasia plana.

[0066] Las realizaciones adicionales serán evidentes para los expertos en la técnica pertinente basándose en las enseñanzas contenidas en este documento.

DESCRIPCIÓN DETALLADA

[0067] Se describe en el presente documento tecnología relacionada con la detección de neoplasia colorrectal y particularmente, pero no exclusivamente, con métodos, composiciones y usos relacionados para detectar cáncer colorrectal premaligno y maligno. Como la tecnología se describe en este documento, los encabezados de sección utilizados son solo para fines organizativos y no deben interpretarse como limitantes del tema de ninguna manera.

[0068] En esta descripción detallada de las diversas formas de realización, para fines de explicación, numerosos detalles específicos se exponen para proporcionar una comprensión completa de las realizaciones descritas. Sin embargo, un experto en la materia apreciará que estas diversas realizaciones se pueden practicar con o sin estos detalles específicos. En otros casos, las estructuras y dispositivos se muestran en forma de diagrama de bloques. Además, un experto en la materia puede apreciar fácilmente que las secuencias específicas en las que se presentan y realizan los métodos son ilustrativas y se contempla que las secuencias pueden variarse.

[0069] A menos que se defina lo contrario, todos los términos técnicos y científicos usados en este documento tienen el mismo significado que el comúnmente entendido por un experto ordinario en la técnica a la que las diversas realizaciones descritas en este documento pertenece. Cuando las definiciones de los términos en las referencias parecen diferir de las definiciones proporcionadas en las presentes enseñanzas, prevalecerá la definición proporcionada en las presentes enseñanzas.

Definiciones

[0070] Para facilitar la comprensión de la tecnología actual, un número de términos y frases se definen a continuación. Se establecen definiciones adicionales a lo largo de la descripción detallada.

[0071] A lo largo de la memoria descriptiva y reivindicaciones, los siguientes términos tienen los significados asociados explícitamente en el presente documento, a menos que el contexto dicte claramente lo contrario. La frase "en una realización" como se usa en el presente documento no se refiere necesariamente a la misma realización, aunque puede hacerlo. Además, la frase "en otra realización" como se usa en el presente documento no se refiere necesariamente a una realización diferente, aunque puede hacerlo.

[0072] Además, tal como se utiliza aquí, el término "o" es un "o" inclusive operador y es equivalente a la expresión

"y/o" a menos que el contexto claramente dicte otra cosa. El término "basado en" no es exclusivo y permite basarse en factores adicionales no descritos, a menos que el contexto indique claramente lo contrario. Además, a lo largo de la especificación, el significado de "un", "una", "el" y "la" incluye referencias plurales. El significado de "en" incluye "en" y "sobre".

5 **[0073]** Como se usa en este documento, un "ácido nucleico" o "molécula de ácido nucleico" se refiere generalmente a cualquier ácido ribonucleico o ácido desoxirribonucleico, que puede ser modificado o no modificado de ADN o ARN. Los "ácidos nucleicos" incluyen, sin limitación, ácidos nucleicos monocatenarios y bicatenarios. Como se usa en el presente documento, el término "ácido nucleico" también incluye ADN como se describió anteriormente que contiene una o más bases modificadas. Por lo tanto, el ADN con un esqueleto modificado por estabilidad o por otras razones es un "ácido nucleico". El término "ácido nucleico" tal como se usa en el presente documento abarca formas de ácidos nucleicos modificadas química, enzimática o metabólicamente, así como las formas químicas de ADN características de virus y células, que incluyen, por ejemplo, células simples y complejas.

15 **[0074]** Los términos "oligonucleótido" o "polinucleótido" o "nucleótido" o "ácido nucleico" se refieren a una molécula que tiene dos o más desoxirribonucleótidos o ribonucleótidos, preferiblemente más de tres, y usualmente más de diez. El tamaño exacto dependerá de muchos factores, que a su vez dependen de la función o uso final del oligonucleótido. El oligonucleótido puede generarse de cualquier manera, incluyendo síntesis química, replicación de ADN, transcripción inversa o una combinación de los mismos. Los desoxirribonucleótidos típicos para el ADN son timina, adenina, citosina y guanina. Los ribonucleótidos típicos para ARN son uracilo, adenina, citosina y guanina.

[0075] Como se usa en este documento, los términos "locus" o "región" de un ácido nucleico se refieren a una subregión de un ácido nucleico, por ejemplo, un gen en un cromosoma, un solo nucleótido, una isla CpG, etc.

25 **[0076]** Los términos "complementario" y "complementariedad" se refieren a nucleótidos (por ejemplo, 1 nucleótido) o polinucleótidos (por ejemplo, una secuencia de nucleótidos) relacionados por las reglas de emparejamiento de bases. Por ejemplo, la secuencia 5'-AGT-3' es complementaria a la secuencia 3'-TCA-5'. La complementariedad puede ser "parcial", en la que solo algunas de las bases de los ácidos nucleicos coinciden de acuerdo con las reglas de emparejamiento de bases. O, puede haber complementariedad "completa" o "total" entre los ácidos nucleicos. El grado de complementariedad entre las cadenas de ácido nucleico afecta la eficiencia y la fuerza de la hibridación entre las cadenas de ácido nucleico. Esto es de particular importancia en las reacciones de amplificación y en los métodos de detección que dependen de la unión entre los ácidos nucleicos.

35 **[0077]** El término "gen" se refiere a un ácido nucleico (por ejemplo, ADN o ARN) secuencia que comprende las secuencias de codificación necesarias para la producción de un ARN, o de un polipéptido o su precursor. Un polipéptido funcional puede codificarse por una secuencia de codificación de longitud completa o por cualquier parte de la secuencia de codificación siempre que se retenga la actividad deseada o las propiedades funcionales (por ejemplo, actividad enzimática, unión de ligando, transducción de señal, etc.) del polipéptido. El término "porción" cuando se usa en referencia a un gen se refiere a fragmentos de ese gen. El tamaño de los fragmentos puede variar desde unos pocos nucleótidos hasta la secuencia génica completa menos un nucleótido. Por lo tanto, "un nucleótido que comprende al menos una porción de un gen" puede comprender fragmentos del gen o del gen completo.

45 **[0078]** El término "gen" también abarca las regiones codificantes de un gen estructural e incluye secuencias localizadas adyacentes a la región de codificación tanto en el 5' y 3', por ejemplo, para una distancia de aproximadamente 1 kb en cada extremo, tal que el gen corresponde a la longitud del ARNm de longitud completa (por ejemplo, que comprende secuencias de codificación, reguladoras, estructurales y de otro tipo). Las secuencias que están ubicadas en 5' de la región de codificación y que están presentes en el ARNm se denominan secuencias 5' no traducidas o no traducidas. Las secuencias que se encuentran 3' o aguas abajo de la región de codificación y que están presentes en el ARNm se denominan secuencias 3' no traducidas o 3' no traducidas. El término "gen" abarca tanto el ADNc como las formas genómicas de un gen. En algunos organismos (p. ej., eucariotas), una forma genómica o clon de un gen contiene la región codificante interrumpida con secuencias no codificantes denominadas "intrones" o "regiones intermedias" o "secuencias intermedias". Los intrones son segmentos de un gen que se transcribe en ARN nuclear (ARNhn); Los intrones pueden contener elementos reguladores tales como potenciadores. Los intrones se eliminan o "empalman" de la transcripción nuclear o primaria; por lo tanto, los intrones están ausentes en la transcripción del ARN mensajero (ARNm). El ARNm funciona durante la traducción para especificar la secuencia u orden de los aminoácidos en un polipéptido naciente.

60 **[0079]** Además de contener intrones, las formas genómicas de un gen pueden incluir también secuencias localizadas en los extremos 5' y 3' de las secuencias que están presentes en el transcrito de ARN. Estas secuencias se denominan secuencias o regiones "flanqueantes" (estas secuencias flanqueantes están ubicadas 5' o 3' a las secuencias no traducidas presentes en el transcrito de ARNm). La región flanqueante 5' puede contener secuencias reguladoras tales como promotores y potenciadores que controlan o influyen en la transcripción del gen. La región flanqueante 3' puede contener secuencias que dirigen la terminación de la transcripción, la escisión postranscripcional y la poliadenilación.

65 **[0080]** El término "de tipo silvestre" cuando se hace en referencia a un gen se refiere a un gen que tiene las

características de un gen aislado de una fuente de origen natural. El término "tipo salvaje" cuando se hace en referencia a un producto genético se refiere a un producto genético que tiene las características de un producto genético aislado de una fuente natural. El término "ocurrencia natural" tal como se aplica a un objeto se refiere al hecho de que un objeto se puede encontrar en la naturaleza. Por ejemplo, una secuencia de polipéptidos o polinucleótidos que está presente en un organismo (incluidos los virus) que puede aislarse de una fuente en la naturaleza y que no ha sido modificada intencionalmente por la mano de una persona en el laboratorio es natural. Un gen de tipo salvaje es a menudo ese gen o alelo que se observa con mayor frecuencia en una población y, por lo tanto, se designa arbitrariamente como la forma "normal" o "de tipo salvaje" del gen. En contraste, el término "modificado" o "mutante" cuando se hace en referencia a un gen o a un producto génico se refiere, respectivamente, a un gen o a un producto génico que muestra modificaciones en la secuencia y/o propiedades funcionales (por ejemplo, alterado características) en comparación con el gen de tipo salvaje o el producto génico. Se observa que los mutantes naturales pueden aislarse; estos se identifican por el hecho de que tienen características alteradas en comparación con el gen de tipo salvaje o el producto génico.

[0081] El término "alelo" se refiere a una variación de un gen; las variaciones incluyen, pero no se limitan a variantes y mutantes, loci polimórficos y loci polimórficos de un solo nucleótido, cambios de marco y mutaciones de empalme. Un alelo puede ocurrir naturalmente en una población o puede surgir durante la vida de cualquier individuo en particular de la población.

[0082] Por lo tanto, los términos "variante" y "mutante" cuando se usa en referencia a una secuencia de nucleótidos se refieren a un ácido nucleico de secuencia que difiere en uno o más nucleótidos de otro, por lo general relacionados, secuencia de ácido nucleótido. Una "variación" es una diferencia entre dos secuencias de nucleótidos diferentes; típicamente, una secuencia es una secuencia de referencia.

[0083] La "amplificación" es un caso especial de replicación de ácido nucleico que implica especificidad de plantilla. Debe contrastarse con la replicación de plantilla no específica (por ejemplo, la replicación que depende de la plantilla pero no depende de una plantilla específica). La especificidad de plantilla se distingue aquí de la fidelidad de replicación (por ejemplo, síntesis de la secuencia de polinucleótidos adecuada) y (ribo- o desoxirribo)especificidad de nucleótidos. La especificidad de la plantilla se describe con frecuencia en términos de especificidad de "diana". Las secuencias diana son "dianas" en el sentido de que se busca que se separen de otros ácidos nucleicos. Las técnicas de amplificación se han diseñado principalmente para esta clasificación.

[0084] La amplificación de ácidos nucleicos generalmente se refiere a la producción de copias múltiples de un polinucleótido, o una porción del polinucleótido, típicamente comenzando desde una pequeña cantidad del polinucleótido (por ejemplo, una molécula de polinucleótido único, 10 a 100 copias de una molécula de polinucleótido, que puede o no ser exactamente la misma), donde los productos de amplificación o amplicones son generalmente detectables. La amplificación de polinucleótidos abarca una variedad de procesos químicos y enzimáticos. La generación de múltiples copias de ADN a partir de una o unas pocas copias de una molécula de ADN diana o plantilla durante una reacción en cadena de la polimerasa (PCR) o una reacción en cadena de la ligasa (LCR; véase, por ejemplo, la Patente de EE.UU. N° 5,494,810) son formas de amplificación. Los tipos adicionales de amplificación incluyen, entre otros, PCR específica de alelo (Patente de EE.UU. N° 5,639,611), PCR de ensamblaje (Patente de EE.UU. N° 5,965,408), amplificación dependiente de helicasa (Patente de EE.UU. N° 7,662,594), PCR de arranque en caliente (Patentes de EE.UU. N°s 5,773,258 y 5,338,671), PCR específica de intersecuencia, PCR inversa (Triglia, et al. Al. (1988) *Nucleic Acids Res.*, 16: 8186), PCR mediada por ligadura (Guilfoyle, R. et al., *Nucleic Acids Research*, 25: 1854-1858 (1997); Patente de EE.UU. N° 5,508,169), PCR específica de metilación (Herman, et al., (1996) *PNAS* 93 (13) 9821-9826), minicizador PCR, ligadura múltiple dependiente de la amplificación de la sonda (Schouten, et al., (2002) *Nucleic Acids Research* 30 (12): e57), PCR multiplex (Chamberlain, et al., (1988) *Nucleic Acids Research* 16 (23) 11141-11156; Ballabio, et al., (1990) *Human Genetics* 84 (6) 571-573; Hayden, et al., (2008) *BMC Genetics* 9:80), PCR anidada, PCR de extensión y solapamiento (Higuchi, et al., (1988) *Nucleic Acids Research* 16 (15) 7351-7367), PCR en tiempo real (Higuchi, et al., (1992) *Biotechnology* 10: 413-417; Higuchi, et al., (1993) *Biotechnology* 11: 1026-1030), PCR de transcripción inversa (Bustin, SA (2000) *J. Molecular Endocrinology* 25: 169-193), PCR en fase sólida, PCR entrelazada asimétrica térmica y PCR de contacto. (Don, et al., *Nucleic Acids Research* (1991) 19 (14) 4008; Roux, K. (1994) *Biotechniques* 16 (5) 812-814; Hecker, et al., (1996) *Biotechniques* 20 (3) 478-485). La amplificación de polinucleótidos también se puede lograr mediante PCR digital (Kalinina, et al., *Nucleic Acids Research*. 25: 1999-2004, (1997); Vogelstein y Kinzler, *Proc Natl Acad Sci EE.UU.* 96: 9236-41, (1999); Publicación de Patente Internacional N° WO05023091A2; Publicación de Solicitud de Patente de EE.UU. No. 20070202525).

[0085] Como se usa en este documento, el término "ensayo de detección de ácido nucleico" se refiere a cualquier método para determinar la composición de nucleótidos de un ácido nucleico de interés. El ensayo de detección de ácido nucleico incluye, pero no se limita a, métodos de secuenciación de ADN, métodos de hibridación de sonda, ensayos de escisión específicos de estructura (por ejemplo, el ensayo INVADER, Hologic, Inc.) y se describen, por ejemplo, en las Patentes de EE.UU. N°s 5,846,717, 5,985,557, 5,994,069, 6,001,567, 6,090,543 y 6,872,816; Lyamichev y col., *Nat. Biotech.*, 17: 292 (1999), Hall et al., *PNAS*, EE. UU., 97: 8272 (2000) y US 2009/0253142; reacción en cadena de la polimerasa; métodos de hibridación ramificada (p. ej., Chiron, Patentes de EE.UU. N°s 5,849,481, 5,710,264, 5,124,246 y 5,624,802); replicación de círculo rodante (p. ej., Patentes de EE.UU. N°s 6,210,884, 6,183,960 y 6,235,502); NASBA (e.g., la Patente de EE.UU. N° 5,409,818); tecnología de baliza molecular (p. ej.,

Patente de EE.UU. N° 6,150,097); tecnología de sonda de ciclismo (p. ej., Patentes de Estados Unidos N°s 5,403,711, 5,011,769 y 5,660,988); reacción en cadena de ligasa (por ejemplo, Barnay Proc. Natl. Acad. Sci EE.UU. 88, 189-93 (1991)); Ensayo QuARTS (p. ej., Según lo provisto por la Patente de EE.UU. N° 8,361,720; y la Solicitud de Patente de EE.UU. y 2012/0122106); y métodos de hibridación en sándwich (p. ej., Patente de EE.UU. N° 5,288,609).

[0086] El término "cebador" se refiere a un oligonucleótido, tanto si se producen de forma natural como en un digesto de restricción purificado o producido sintéticamente, que es capaz de actuar como un punto de iniciación de la síntesis cuando se coloca bajo condiciones en las que la síntesis de un producto de extensión del cebador que es complementario a una cadena de ácido nucleico se induce (por ejemplo, en presencia de nucleótidos y un agente inductor tal como una polimerasa de ADN y a una temperatura y pH adecuados). El cebador es preferiblemente monocatenario para la máxima eficiencia en la amplificación, pero alternativamente puede ser bicatenario. Si es de doble hebra, la imprimación se trata primero para separar sus hebras antes de usarse para preparar productos de extensión. Preferiblemente, el cebador es un oligodesoxirribonucleótido. El cebador debe ser lo suficientemente largo para cebar la síntesis de productos de extensión en presencia del agente inductor. Las longitudes exactas de los cebadores dependerán de muchos factores, incluida la temperatura, la fuente del cebador y el uso del método.

[0087] El término "sonda" se refiere a un oligonucleótido (por ejemplo, una secuencia de nucleótidos), sea de origen natural como en un digesto de restricción purificado o producido sintéticamente, de forma recombinante, o mediante amplificación por PCR, que es capaz de hibridar con otro oligonucleótido de interesar. Una sonda puede ser monocatenaria o bicatenaria. Las sondas son útiles en la detección, identificación y aislamiento de secuencias de genes particulares (por ejemplo, una "sonda de captura"). Se contempla que cualquier sonda utilizada en la presente invención puede, en algunas realizaciones, marcarse con cualquier "molécula informadora", de modo que sea detectable en cualquier sistema de detección, que incluye, pero no se limita a enzima (por ejemplo, ELISA, así como ensayos histoquímicos basados en enzimas), sistemas fluorescentes, radiactivos y luminiscentes. No se pretende que la presente divulgación se limite a ningún sistema o etiqueta de detección particular.

[0088] Como se usa en este documento, "metilación" se refiere a la metilación de citosina en las posiciones C5 o N4 de la citosina, la posición N6 de adenina, u otros tipos de metilación ácida nucleica. El ADN amplificado in vitro generalmente no está metilado porque los métodos típicos de amplificación de ADN in vitro no retienen el patrón de metilación de la plantilla de amplificación. Sin embargo, "ADN no metilado" o "ADN metilado" también pueden referirse a ADN amplificado cuya plantilla original estaba no metilada o metilada, respectivamente.

[0089] Por consiguiente, como se usa en el presente documento, un "nucleótido metilado" o una "base de nucleótido metilada" se refiere a la presencia de un resto metilo en una base de nucleótidos, donde el resto metilo no está presente en una base nucleotídica típica reconocida. Por ejemplo, la citosina no contiene un resto metilo en su anillo de pirimidina, pero la 5-metilcitosina contiene un resto metilo en la posición 5 de su anillo de pirimidina. Por lo tanto, la citosina no es un nucleótido metilado y la 5-metilcitosina es un nucleótido metilado. En otro ejemplo, la timina contiene un resto metilo en la posición 5 de su anillo de pirimidina; sin embargo, para los fines de este documento, la timina no se considera un nucleótido metilado cuando está presente en el ADN, ya que la timina es una base típica de nucleótidos del ADN.

[0090] Tal como se utiliza aquí, una "molécula de ácido nucleico metilado" se refiere a una molécula de ácido nucleico que contiene uno o más nucleótidos metilados.

[0091] Como se usa en este documento, un "estado de metilación", "perfil de metilación", y "estado de metilación" de una molécula de ácido nucleico se refiere a la presencia de la ausencia de una o más bases de nucleótidos metilados en la molécula de ácido nucleico. Por ejemplo, una molécula de ácido nucleico que contiene una citosina metilada se considera metilada (por ejemplo, el estado de metilación de la molécula de ácido nucleico está metilado). Una molécula de ácido nucleico que no contiene ningún nucleótido metilado se considera no metilada.

[0092] El estado de metilación de una secuencia de ácido nucleico particular (por ejemplo, un marcador de genes o región de ADN como se describe en el presente documento) puede indicar el estado de metilación de cada base en la secuencia o puede indicar el estado de metilación de un subconjunto de las bases (por ejemplo, de una o más citosinas) dentro de la secuencia, o puede indicar información sobre la densidad de metilación regional dentro de la secuencia con o sin proporcionar información precisa de las ubicaciones dentro de la secuencia en que ocurre la metilación.

[0093] El estado de metilación de un locus de nucleótidos en una molécula de ácido nucleico se refiere a la presencia o ausencia de un nucleótido metilado en un locus particular, en la molécula de ácido nucleico. Por ejemplo, el estado de metilación de una citosina en el 7° nucleótido en una molécula de ácido nucleico se metila cuando el nucleótido presente en el 7° nucleótido en la molécula de ácido nucleico es 5-metilcitosina. De manera similar, el estado de metilación de una citosina en el 7° nucleótido en una molécula de ácido nucleico no está metilado cuando el nucleótido presente en el 7° nucleótido en la molécula de ácido nucleico es citosina (y no 5-metilcitosina).

[0094] El estado de metilación puede opcionalmente estar representado o indicado por un "valor de metilación" (por ejemplo, representando una frecuencia de metilación, la fracción, la relación, por ciento, etc.). Un valor de metilación

puede ser generado, por ejemplo, mediante la cuantificación de la cantidad de ácido nucleico intacto presente después de la digestión de restricción con una enzima de restricción dependiente de la metilación o mediante la comparación de perfiles de amplificación después de la reacción con bisulfito o mediante la comparación de secuencias de ácidos nucleicos tratados con bisulfito y sin tratar. En consecuencia, un valor, por ejemplo, un valor de metilación, representa el estado de metilación y, por lo tanto, puede usarse como un indicador cuantitativo del estado de metilación en múltiples copias de un locus. Esto es de uso particular cuando es deseable comparar el estado de metilación de una secuencia en una muestra con un umbral o valor de referencia.

[0095] Como se usa en este documento, "frecuencia de metilación" o "porcentaje de metilación (%)" se refieren al número de casos en los que una molécula o locus está metilado en relación con el número de instancias en que la molécula o locus es no metilado.

[0096] Como tal, el estado de metilación describe el estado de metilación de un ácido nucleico (por ejemplo, una secuencia genómica). Además, el estado de metilación se refiere a las características de un segmento de ácido nucleico en un locus genómico particular relevante para la metilación. Dichas características incluyen, pero no se limitan a, si alguno de los residuos de citosina (C) dentro de esta secuencia de ADN están metilados, la ubicación de los residuos C metilados, la frecuencia o el porcentaje de C metilado en cualquier región particular de un ácido nucleico y diferencias alélicas en la metilación debido, por ejemplo, a la diferencia en el origen de los alelos. Los términos "estado de metilación", "perfil de metilación" y "estado de metilación" también se refieren a la concentración relativa, concentración absoluta o patrón de C metilado o C no metilado en cualquier región particular de un ácido nucleico en una muestra biológica. Por ejemplo, si el (los) residuo(s) de citosina (C) dentro de una secuencia de ácido nucleico están metilados, pueden denominarse "hipermetilados" o "metilación incrementada", mientras que si el (los) residuo(s) de citosina (C) está(n) dentro de una secuencia ADN no están metilados, pueden denominarse "hipometilados" o que tienen "metilación disminuida". Del mismo modo, si el (los) residuo(s) de citosina (C) dentro de una secuencia de ácido nucleico se metilan en comparación con otra secuencia de ácido nucleico (por ejemplo, de una región diferente o de un individuo diferente, etc.), esa secuencia se considera hipermetilada o ha aumentado metilación en comparación con la otra secuencia de ácido nucleico. Alternativamente, si los residuos de citosina (C) dentro de una secuencia de ADN no están metilados en comparación con otra secuencia de ácido nucleico (por ejemplo, de una región diferente o de un individuo diferente, etc.), esa secuencia se considera hipometilada o ha disminuido metilación en comparación con la otra secuencia de ácido nucleico. Además, el término "patrón de metilación" como se usa en el presente documento se refiere a los sitios colectivos de nucleótidos metilados y no metilados sobre una región de un ácido nucleico. Dos ácidos nucleicos pueden tener la misma frecuencia de metilación o un porcentaje de metilación similar o similar, pero tienen patrones de metilación diferentes cuando el número de nucleótidos metilados y no metilados es igual o similar en toda la región, pero las ubicaciones de los nucleótidos metilados y no metilados son diferentes. Se dice que las secuencias están "metiladas diferencialmente" o que tienen una "diferencia en la metilación" o que tienen un "estado de metilación diferente" cuando difieren en la extensión (por ejemplo, se ha aumentado o disminuido la metilación en relación con la otra), la frecuencia o el patrón de metilación. El término "metilación diferencial" se refiere a una diferencia en el nivel o patrón de metilación de ácido nucleico en una muestra positiva de cáncer en comparación con el nivel o patrón de metilación de ácido nucleico en una muestra de cáncer negativo. También puede referirse a la diferencia en niveles o patrones entre pacientes que tienen recurrencia de cáncer después de la cirugía versus pacientes que no tienen recurrencia. La metilación diferencial y los niveles o patrones específicos de metilación del ADN son biomarcadores pronósticos y predictivos, por ejemplo, una vez que se han definido las características predictivas o de corte correctas.

[0097] La frecuencia de estado de metilación puede ser usada para describir una población de individuos o de una muestra de un solo individuo. Por ejemplo, un locus de nucleótidos que tiene una frecuencia de estado de metilación del 50% se metila en el 50% de los casos y no se metila en el 50% de los casos. Dicha frecuencia puede usarse, por ejemplo, para describir el grado en que un locus de nucleótidos o una región de ácido nucleico se metila en una población de individuos o en una colección de ácidos nucleicos. Por lo tanto, cuando la metilación en una primera población o grupo de moléculas de ácido nucleico es diferente de la metilación en una segunda población o grupo de moléculas de ácido nucleico, la frecuencia del estado de metilación de la primera población o grupo será diferente de la frecuencia del estado de metilación de la segunda población o piscina. Tal frecuencia también se puede usar, por ejemplo, para describir el grado en que un locus de nucleótidos o una región de ácido nucleico se metila en un solo individuo. Por ejemplo, dicha frecuencia puede usarse para describir el grado en que un grupo de células de una muestra de tejido se metila o no se metila en un locus de nucleótidos o región de ácido nucleico.

[0098] Como se usa en el presente documento un "locus de nucleótidos" se refiere a la ubicación de un nucleótido en una molécula de ácido nucleico. Un locus de nucleótidos de un nucleótido metilado se refiere a la ubicación de un nucleótido metilado en una molécula de ácido nucleico.

[0099] Típicamente, la metilación de ADN humano se produce en una secuencia de dinucleótido incluyendo una guanina adyacente y citosina donde se encuentra la citosina 5' de la guanina (también denominadas secuencias de dinucleótidos CpG). La mayoría de las citosinas dentro de los dinucleótidos CpG están metiladas en el genoma humano, sin embargo, algunas permanecen sin metilar en regiones genómicas específicas ricas en dinucleótidos CpG, conocidas como islas CpG (véase, por ejemplo, Antequera et al. (1990) Cell 62: 503-514).

5 **[0100]** Como se usa en el presente documento, una "isla CpG" se refiere a una región rica en G:C de ADN genómico que contiene un mayor número de dinucleótidos CpG con respecto al ADN genómico total. Una isla CpG puede tener al menos 100, 200 o más pares de bases de longitud, donde el contenido G:C de la región es al menos 50% y la proporción de frecuencia CpG observada sobre la frecuencia esperada es 0,6; en algunos casos, una isla CpG puede tener al menos 500 pares de bases de longitud, donde el contenido G:C de la región es al menos 55%) y la proporción de frecuencia CpG observada sobre la frecuencia esperada es 0,65. La frecuencia de CpG observada sobre la frecuencia esperada se puede calcular de acuerdo con el método proporcionado en Gardiner-Garden et al (1987) J. Mol. Biol. 196: 261-281. Por ejemplo, la frecuencia de CpG observada sobre la frecuencia esperada se puede calcular de acuerdo con la fórmula $R = (A \times B) / (C \times D)$, donde R es la relación de la frecuencia de CpG observada sobre la frecuencia esperada, A es el número de CpG dinucleótidos en una secuencia analizada, B es el número total de nucleótidos en la secuencia analizada, C es el número total de nucleótidos C en la secuencia analizada y D es el número total de nucleótidos G en la secuencia analizada. El estado de metilación se determina típicamente en islas CpG, por ejemplo, en regiones promotoras. Sin embargo, se apreciará que otras secuencias en el genoma humano son propensas a la metilación del ADN como CpA y CpT (véase Ramsahoye (2000) Proc. Natl. Acad. Sci. EE.UU. 97: 5237-5242; Salmon y Kaye (1970) Biochim Biophys. Acta. 204: 340-351; Grafstrom (1985) Nucleic Acids Res. 13: 2827-2842; Nyce (1986) Nucleic Acids Res. 14: 4353-4367; Woodcock (1987) Biochem. Biophys. Res. Commun. 145: 888-894).

20 **[0101]** Como se usa en el presente documento, un reactivo que modifica un nucleótido de la molécula de ácido nucleico en función del estado de metilación de la molécula de ácido nucleico, o un reactivo específico de metilación, se refiere a un compuesto o composición u otro agente que puede cambiar la secuencia de nucleótidos de una molécula de ácido nucleico de una manera que refleja el estado de metilación de la molécula de ácido nucleico. Los métodos para tratar una molécula de ácido nucleico con dicho reactivo pueden incluir poner en contacto la molécula de ácido nucleico con el reactivo, junto con pasos adicionales, si se desea, para lograr el cambio deseado de la secuencia de nucleótidos. Tal cambio en la secuencia de nucleótidos de la molécula de ácido nucleico puede dar como resultado una molécula de ácido nucleico en la que cada nucleótido metilado se modifica a un nucleótido diferente. Tal cambio en la secuencia de nucleótidos de ácido nucleico puede dar como resultado una molécula de ácido nucleico en la que cada nucleótido no metilado se modifica a un nucleótido diferente. Tal cambio en la secuencia de nucleótidos de ácido nucleico puede dar como resultado una molécula de ácido nucleico en la que cada uno de un nucleótido seleccionado que no está metilado (por ejemplo, cada citosina no metilada) se modifica a un nucleótido diferente. El uso de dicho reactivo para cambiar la secuencia de nucleótidos de ácido nucleico puede dar como resultado una molécula de ácido nucleico en la que cada nucleótido que es un nucleótido metilado (por ejemplo, cada citosina metilada) se modifica a un nucleótido diferente. Como se usa en este documento, el uso de un reactivo que modifica un nucleótido seleccionado se refiere a un reactivo que modifica un nucleótido de los cuatro nucleótidos que ocurren típicamente en una molécula de ácido nucleico (C, G, T y A para ADN y C, G, U, y A para ARN), de modo que el reactivo modifique un nucleótido sin modificar los otros tres nucleótidos. En una realización ejemplar, dicho reactivo modifica un nucleótido seleccionado no metilado para producir un nucleótido diferente. En otra realización ejemplar, dicho reactivo puede desaminar nucleótidos de citosina no metilados. Un reactivo ejemplar es el bisulfito.

40 **[0102]** Como se usa en este documento, el término "reactivo de bisulfito" se refiere a un reactivo que comprende en algunas realizaciones bisulfito, disulfito, sulfito de hidrógeno, o combinaciones de los mismos para distinguir entre citidinas metiladas y no metiladas, por ejemplo, en secuencias de dinucleótidos CpG.

45 **[0103]** El término "ensayo de metilación" se refiere a cualquier ensayo para determinar el estado de metilación de una o más secuencias de dinucleótidos CpG dentro de una secuencia de un ácido nucleico.

50 **[0104]** El término "MS AP-PCR" (reacción en cadena de la polimerasa preparada arbitrariamente sensible a la metilación) se refiere a la tecnología reconocida en la técnica que permite un escaneo global del genoma usando cebadores ricos en CG para centrarse en las regiones más probables para contener dinucleótidos CpG, y descritos por Gonzalzo et al. (1997) Cancer Research 57: 594-599.

55 **[0105]** El término "MethylLight™" se refiere a la técnica de PCR en tiempo real basada en fluorescencia reconocida en la técnica descrita por Eads et al. (1999) Cancer Res. 59: 2302-2306.

60 **[0106]** El término "HeavyMethyl™" se refiere a un ensayo en donde las sondas de bloqueo específicas de metilación (también denominadas en este documento como bloqueadores) que cubren las posiciones de CpG entre, o cubiertas por, los cebadores de amplificación permiten la amplificación selectiva específica de metilación de una muestra de ácido nucleico.

[0107] El término ensayo "HeavyMethyl™ MethylLight™" se refiere a un ensayo HeavyMethyl™ MethylLight™, que es una variación del ensayo MethylLight™, en donde el ensayo de MethylLight™ se combina con la metilación específica de bloqueo sondas que cubren las posiciones de CpG entre los cebadores de amplificación.

65 **[0108]** El término "Ms-SNuPE" (Methylation-sensitive Single Nucleotide Primer Extension) se refiere al ensayo reconocido en la técnica descrito por Gonzalzo y Jones (1997) Nucleic Acids Res. 25: 2529-2531.

[0109] El término "MSP" (PCR específica de metilación) se refiere al ensayo de metilación reconocido en la técnica descrito por Herman et al. (1996) Proc. Natl. Acad. Sci. EE.UU. 93: 9821-9826, y por la Patente de EE.UU. N° 5,786,146.

[0110] El término "COBRA" (Combined Bisulfite Restriction Analysis) se refiere al ensayo de metilación reconocido en la técnica descrito por Xiong & Laird (1997) Nucleic Acids Res. 25: 2532-2534.

[0111] El término "MCA" (Methylated CpG Island Amplification) se refiere al ensayo de metilación descrito por Toyota et al. (1999) Cancer Res. 59: 2307-12, y en el documento WO 00/26401A1.

[0112] Como se usa en el presente documento, un "nucleótido seleccionado" se refiere a un nucleótido de los cuatro nucleótidos que se producen típicamente en una molécula de ácido nucleico (C, G, T y A para ADN y C, G, U y A para ARN) y puede incluir derivados metilados de los nucleótidos que ocurren típicamente (por ejemplo, cuando C es el nucleótido seleccionado, tanto el C metilado como el no metilado se incluyen dentro del significado de un nucleótido seleccionado), mientras que un nucleótido seleccionado metilado se refiere específicamente a un nucleótido metilado que ocurre típicamente y un nucleótido seleccionado no metilado se refiere específicamente a un nucleótido no metilado que se produce típicamente.

[0113] Los términos "enzima de restricción específica de metilación" o "enzima de restricción sensible a metilación" se refieren a una enzima que digiere selectivamente un ácido nucleico dependiente del estado de metilación de su sitio de reconocimiento. En el caso de una enzima de restricción que corta específicamente si el sitio de reconocimiento no está metilado o está hemimetilado, el corte no tendrá lugar o tendrá una eficacia significativamente reducida si el sitio de reconocimiento está metilado. En el caso de una enzima de restricción que corta específicamente si el sitio de reconocimiento está metilado, el corte no tendrá lugar o tendrá una eficacia significativamente reducida si el sitio de reconocimiento no está metilado. Se prefieren las enzimas de restricción específicas de metilación, cuya secuencia de reconocimiento contiene un dinucleótido CG (por ejemplo, una secuencia de reconocimiento como CGCG o CCCGGG). Más preferidas para algunas realizaciones son las enzimas de restricción que no se cortan si la citosina en este dinucleótido se metila en el átomo de carbono C5.

[0114] Como se usa en el presente documento, un "nucleótido diferente" se refiere a un nucleótido que es químicamente diferente de un nucleótido seleccionado, típicamente tal que el nucleótido diferente tiene propiedades de emparejamiento de bases Watson-Crick que difieren del nucleótido seleccionado, con lo que el nucleótido que ocurre típicamente que es complementario al nucleótido seleccionado no es el mismo que el nucleótido que ocurre típicamente que es complementario a los diferentes nucleótidos. Por ejemplo, cuando C es el nucleótido seleccionado, U o T puede ser el nucleótido diferente, que se ejemplifica por la complementariedad de C a G y la complementariedad de U o T a A. Como se usa en el presente documento, un nucleótido que es complementario al nucleótido seleccionado o que es complementario a los diferentes nucleótidos se refiere a un nucleótido que se empareja en bases, en condiciones de alta rigurosidad, con el nucleótido seleccionado o un nucleótido diferente con mayor afinidad que el emparejamiento de bases del nucleótido complementario con tres de los cuatro nucleótidos que ocurren típicamente. Un ejemplo de complementariedad es el emparejamiento de bases de Watson-Crick en ADN (por ejemplo, AT y CG) y ARN (por ejemplo, AU y CG). Así, por ejemplo, los pares de bases G, en condiciones de alta rigurosidad, con mayor afinidad a C que los pares de bases G a G, A o T y, por lo tanto, cuando C es el nucleótido seleccionado, G es un nucleótido complementario al nucleótido seleccionado.

[0115] Como se usa en el presente documento, la "sensibilidad" de un marcador dado se refiere al porcentaje de muestras que informan un valor de metilación de ADN por encima de un valor umbral que distingue entre muestras neoplásicas y no neoplásicas. En algunas realizaciones, un positivo se define como una neoplasia confirmada por histología que informa un valor de metilación del ADN por encima de un valor umbral (por ejemplo, el rango asociado con la enfermedad), y un falso negativo se define como una neoplasia confirmada por histología que informa un valor de metilación de ADN por debajo del valor umbral (p. ej., el rango asociado con ninguna enfermedad). El valor de la sensibilidad, por lo tanto, refleja la probabilidad de que una medición de metilación del ADN para un marcador dado obtenido de una muestra enferma conocida esté dentro del rango de mediciones asociadas a la enfermedad. Como se define aquí, la relevancia clínica del valor de sensibilidad calculado representa una estimación de la probabilidad de que un marcador determinado detecte la presencia de una condición clínica cuando se aplica a un sujeto con esa condición.

[0116] Como se usa en este documento, la "especificidad" de un marcador dado se refiere al porcentaje de muestras no neoplásicas que informan un valor de metilación del ADN por debajo de un valor umbral que distingue entre muestras neoplásicas y no neoplásicas. En algunas realizaciones, un negativo se define como una muestra no neoplásica confirmada por histología que informa un valor de metilación del ADN por debajo del valor umbral (por ejemplo, el rango asociado con ninguna enfermedad) y un falso positivo se define como una muestra no neoplásica confirmada histológicamente que informa un valor de metilación del ADN por encima del valor umbral (por ejemplo, el rango asociado con la enfermedad). El valor de especificidad, por lo tanto, refleja la probabilidad de que una medición de metilación de ADN para un marcador dado obtenido de una muestra no neoplásica conocida esté en el rango de mediciones no asociadas a la enfermedad. Como se define aquí, la relevancia clínica del valor de especificidad

calculado representa una estimación de la probabilidad de que un marcador determinado detecte la ausencia de una condición clínica cuando se aplica a un paciente sin esa condición.

5 **[0117]** El término "AUC" como se usa en el presente documento es una abreviatura para el "área bajo una curva". En particular, se refiere al área bajo una curva de características operativas del receptor (ROC). La curva ROC es un gráfico de la tasa positiva verdadera contra la tasa de falsos positivos para los diferentes puntos de corte posibles de una prueba de diagnóstico. Muestra la compensación entre sensibilidad y especificidad según el punto de corte seleccionado (cualquier aumento en la sensibilidad irá acompañado de una disminución en la especificidad). El área bajo una curva ROC (AUC) es una medida para la precisión de una prueba de diagnóstico (cuanto mayor sea el área, mejor; el óptimo es 1; una prueba aleatoria tendría una curva ROC en la diagonal con un área de 0,5; para referencia: JP Egan. (1975) Signal Detection Theory and ROC Analysis, Academic Press, Nueva York).

15 **[0118]** Como se usa en el presente documento, el término "neoplasia" se refiere a "una masa anormal de tejido, cuyo crecimiento excede y está descoordinado con el de los tejidos normales" Véase, por ejemplo, Willis RA, "The Spread of Tumors in the Human Body", Londres, Butterworth & Co, 1952.

[0119] Como se usa en este documento, el término "adenoma" se refiere a un tumor benigno de origen glandular. Aunque estos crecimientos son benignos, con el tiempo pueden progresar hasta convertirse en malignos.

20 **[0120]** El término "pre-canceroso" o "pre-neoplásico" y sus equivalentes se refieren a cualquier trastorno proliferativo celular que esté experimentando transformación maligna.

25 **[0121]** Un "sitio" de una neoplasia, adenoma, cáncer, etc. es el tejido, órgano, tipo de célula, área anatómica, parte del cuerpo, etc. en el cuerpo de un sujeto donde se ubica la neoplasia, adenoma, cáncer, etc.

30 **[0122]** Como se usa en el presente documento, una aplicación de prueba de "diagnóstico" incluye la detección o identificación de un estado o condición de enfermedad de un sujeto, determinando la probabilidad de que un sujeto contraiga una enfermedad o condición dada, determinando la probabilidad de que un sujeto con una enfermedad o afección responderá a la terapia, determinando el pronóstico de un sujeto con una enfermedad o afección (o su probable progresión o regresión) y determinando el efecto de un tratamiento en un sujeto con una enfermedad o afección. Por ejemplo, se puede usar un diagnóstico para detectar la presencia o probabilidad de que un sujeto contraiga una neoplasia o la probabilidad de que dicho sujeto responda favorablemente a un compuesto (por ejemplo, un medicamento, por ejemplo, un medicamento) u otro tratamiento.

35 **[0123]** El término "marcador", como se usa en el presente documento, se refiere a una sustancia (por ejemplo, un ácido nucleico o una región de un ácido nucleico) que puede diagnosticar un cáncer al distinguir las células cancerosas de las células normales, por ejemplo, basándose en su estado de metilación.

40 **[0124]** El término "aislado" cuando se usa en relación con un ácido nucleico, como en "un oligonucleótido aislado" se refiere a una secuencia de ácido nucleico que se identifica y se separa de al menos un ácido nucleico contaminante con el que normalmente se asocia en su fuente natural. El ácido nucleico aislado está presente en una forma o entorno diferente al que se encuentra en la naturaleza. Por el contrario, los ácidos nucleicos no aislados, como el ADN y el ARN, se encuentran en el estado en que existen en la naturaleza. Los ejemplos de ácidos nucleicos no aislados incluyen: una secuencia de ADN dada (por ejemplo, un gen) que se encuentra en el cromosoma de la célula huésped en proximidad de genes vecinos; las secuencias de ARN, como una secuencia específica de ARNm que codifica una proteína específica, se encuentran en la célula como una mezcla con numerosos otros ARNm que codifican una multitud de proteínas. Sin embargo, el ácido nucleico aislado que codifica una proteína particular incluye, a modo de ejemplo, dicho ácido nucleico en las células que normalmente expresan la proteína, donde el ácido nucleico está en una ubicación cromosómica diferente de la de las células naturales, o está flanqueado por una secuencia de ácido nucleico diferente que la encontrada en la naturaleza. El ácido nucleico u oligonucleótido aislado puede estar presente en forma monocatenaria o bicatenaria. Cuando se va a utilizar un ácido nucleico u oligonucleótido aislado para expresar una proteína, el oligonucleótido contendrá como mínimo la cadena sentido o codificante (es decir, el oligonucleótido puede ser monocatenario), pero puede contener hebras tanto sentido como antisentido (es decir, el oligonucleótido puede ser bicatenario). Un ácido nucleico aislado puede, después del aislamiento de su entorno natural o típico, ser combinado con otros ácidos nucleicos o moléculas. Por ejemplo, un ácido nucleico aislado puede estar presente en una célula huésped en la que se ha colocado, por ejemplo, para expresión heteróloga.

55 **[0125]** El término "purificado" se refiere a moléculas, secuencias de ácido nucleico o de aminoácidos que se eliminan de su entorno natural, se aíslan o se separan. Por lo tanto, una "secuencia de ácido nucleico aislada" puede ser una secuencia de ácido nucleico purificada. Las moléculas "sustancialmente purificadas" son al menos 60% libres, preferiblemente al menos 75% libres, y más preferiblemente al menos 90% libres de otros componentes con los que están naturalmente asociados. Como se usa en este documento, los términos "purificado" o "purificar" también se refieren a la eliminación de contaminantes de una muestra. La eliminación de proteínas contaminantes da como resultado un aumento en el porcentaje de polipéptido o ácido nucleico de interés en la muestra. En otro ejemplo, los polipéptidos recombinantes se expresan en células huésped vegetales, bacterianas, de levadura o de mamífero y los

polipéptidos se purifican mediante la eliminación de las proteínas de la célula huésped; el porcentaje de polipéptidos recombinantes se incrementa así en la muestra.

5 **[0126]** El término "composición que comprende" una secuencia de polinucleótidos dada o polipéptido se refiere ampliamente a cualquier composición que contiene la secuencia de polinucleótidos dada o polipéptido. La composición puede comprender una solución acuosa que contiene sales (por ejemplo, NaCl), detergentes (por ejemplo, SDS) y otros componentes (por ejemplo, solución de Denhardt, leche en polvo, ADN de esperma de salmón, etc.).

10 **[0127]** El término "muestra" se usa en su sentido más amplio. En cierto sentido, puede referirse a una célula o tejido animal. En otro sentido, está destinado a incluir una muestra o cultivo obtenido de cualquier fuente, así como muestras biológicas y ambientales. Las muestras biológicas se pueden obtener de plantas o animales (incluidos los humanos) y abarcan fluidos, sólidos, tejidos y gases. Las muestras ambientales incluyen material ambiental como materia superficial, suelo, agua y muestras industriales.

15 **[0128]** Como se usa en el presente documento, una "muestra remota" como se usa en algunos contextos se refiere a una muestra recolectada indirectamente de un sitio que no es la fuente de células, tejidos u órganos de la muestra. Por ejemplo, cuando el material de muestra que se origina en el colon o el recto se evalúa en una muestra de heces (por ejemplo, no de una muestra tomada directamente de tejido colorrectal), la muestra es una muestra remota.

20 **[0129]** Como se usa en el presente documento, los términos "paciente" o "sujeto" se refieren a organismos sujetos a diversas pruebas proporcionadas por la tecnología. El término "sujeto" incluye animales, preferiblemente mamíferos, incluidos humanos. En una realización preferida, el sujeto es un primate. En una realización aún más preferida, el sujeto es un humano.

25 **[0130]** Como se usa en el presente documento, el término "kit" se refiere a cualquier sistema de entrega para entregar materiales. En el contexto de los ensayos de reacción, dichos sistemas de entrega incluyen sistemas que permiten el almacenamiento, transporte o entrega de reactivos de reacción (por ejemplo, oligonucleótidos, enzimas, etc. en los contenedores apropiados) y/o materiales de soporte (por ejemplo, tampones, instrucciones escritas para realizar el ensayo, etc.) de un lugar a otro. Por ejemplo, los kits incluyen uno o más recintos (p. ej., cajas) que contienen los reactivos de reacción y/o materiales de soporte relevantes. Como se usa en el presente documento, el término "kit fragmentado" se refiere a sistemas de suministro que comprenden dos o más recipientes separados que contienen cada uno una porción del total de componentes del kit. Los contenedores pueden ser entregados al destinatario previsto juntos o por separado. Por ejemplo, un primer recipiente puede contener una enzima para usar en un ensayo, mientras que un segundo recipiente contiene oligonucleótidos. El término "kit fragmentado" pretende abarcar kits que contienen reactivos específicos de analitos (ASR) regulados bajo la sección 520 (e) de la Ley Federal de Alimentos, Medicamentos y Cosméticos, pero no se limitan a los mismos. De hecho, cualquier sistema de entrega que comprenda dos o más contenedores separados, cada uno de los cuales contiene una parte del total de componentes del kit, se incluye en el término "kit fragmentado". En contraste, un "kit combinado" se refiere a un sistema de suministro que contiene todos los componentes de un ensayo de reacción en un único recipiente (por ejemplo, en una sola caja que contiene cada uno de los componentes deseados). El término "kit" incluye kits fragmentados y combinados.

Realizaciones de la tecnología

45 **[0131]** En conjunto, los cánceres gastrointestinales representan más mortalidad por cáncer que cualquier otro sistema de órganos. El cáncer colorrectal (CCR) es el segundo cáncer más mortal, en el país, con más de 600.000 muertes anuales. Si bien los cánceres colorrectales se examinan en los Estados Unidos, el cumplimiento es deficiente dado el costo, la incomodidad y la invasividad de la colonoscopia y el pésimo desempeño del menú actual de análisis de sangre fecal no invasiva. Para disminuir la carga de CCR en los individuos y la sociedad, se necesitan nuevas estrategias de prueba que sean efectivas y amigables para el paciente. Una prueba molecular precisa y no invasiva que utiliza biomarcadores ampliamente informativos puede proporcionar un enfoque racional. Un ensayo basado en heces es uno de esos ejemplos. Los cánceres y precánceres colorrectales arrojan células y ADN a la corriente digestiva y finalmente se excretan en las heces. Se han utilizado ensayos altamente sensibles para detectar marcadores genéticos y epigenéticos en heces de pacientes con cáncer de CCR y pólipos precancerosos. En un estudio multicéntrico reciente, estos marcadores se incorporaron a un ensayo de heces que exhibía una sensibilidad para el CCR esencialmente equivalente a la de la colonoscopia. Los nuevos marcadores serían de mayor valor si demostraran ser más sensibles, más específicos o más predictivos del sitio de la lesión que los marcadores existentes. Además, los marcadores idealmente detectarían las lesiones precancerosas críticas (pólipos adenomatosos y pólipos serrados) además del CCR cuando se aplican a una aplicación de detección.

60 **[0132]** Los mecanismos genómicos subyacentes al cáncer colorrectal implican genes y anomalías cromosómicas, incluyendo mutaciones de una sola base, aneuploidía, deleciones, translocaciones, alteraciones del número de copias, y cambios de expresión. Todos estos eventos se están estudiando intensamente utilizando nuevas tecnologías genómicas, incluida la secuenciación masiva paralela. Sin embargo, las alteraciones genéticas están demostrando ser muy heterogéneas y, en algunos aspectos, aleatorias, en lugar de recurrentes. El gen APC, por ejemplo, es el gen más mutado en las lesiones de CCR (~90%), pero los sitios de mutación se extienden a lo largo de 15 exones

65

codificadores que requieren un análisis de todo el gen. Esto agrega complejidad al desarrollo de ensayos con los niveles de rendimiento requeridos.

[0133] La metilación epigenético de ADN en sitios de la isla de citosina-fosfato-guanina (CpG) por metiltransferasas de ADN se ha estudiado como una clase potencial de biomarcadores en los tejidos de la mayoría de los tipos de tumores. En un mecanismo biológicamente atractivo, se cree que los eventos de metilación adquiridos en regiones promotoras de genes supresores de tumores silencian la expresión, contribuyendo a la oncogénesis. La metilación del ADN puede ser una herramienta de diagnóstico más estable química y biológicamente que la expresión de ARN o proteína. Además, los marcadores de metilación aberrantes son más ampliamente informativos y sensibles que las mutaciones de ADN individuales y ofrecen una excelente especificidad.

[0134] Las aplicaciones clínicas de marcadores altamente discriminantes podrían tener un gran impacto. Por ejemplo, el análisis de dichos marcadores en medios distantes, como heces o sangre, se utiliza en pruebas de detección o diagnósticos precisos para la detección de neoplasia colorrectal.

[0135] En experimentos realizados durante el desarrollo de realizaciones para la presente divulgación, se identificaron marcadores en un estudio de casos y controles comparando el estado de metilación de los marcadores de ADN del tejido colorrectal de sujetos con neoplasia colorrectal, adenoma y/o pólipos de sésiles serrados (SSP) al estado de metilación de los mismos marcadores de ADN de sujetos de control (p. ej., tejido normal como colon normal) (véanse, Ejemplos 1-2, Tablas 1-5).

[0136] Por ejemplo, marcadores y/o paneles de marcadores (por ejemplo, una región cromosómica que tiene una anotación seleccionada de FLI1, OPLAH, DTX1, MATK, SFMBT2 región 2, KCNK12, VAV3 región 1, SFMBT2 región 3, PPP2R5C, CHST2 región 2, PKIA, PDGFD, ELOVL2, CHST2 región 1, SFMBT2 región 1, QKI, VAV3 región 2 y SLC8A3) se identificaron en un estudio de casos y controles comparando el estado de metilación de los marcadores de ADN (p. ej., del tejido colorrectal de sujetos con neoplasia colorrectal y/o adenoma al estado de metilación de los mismos marcadores de ADN de los sujetos de control (por ejemplo, tejido normal como el de colon normal) (véase, ejemplo 1 y las Tablas 1A y 1B).

[0137] Los marcadores y/o paneles de marcadores (por ejemplo, una región cromosómica que tiene una anotación seleccionada de BMP3, NDRG4, PDGFG, CHST2 y SFMBT2) se identificaron en estudios de casos y controles al comparar el estado de metilación de los marcadores de ADN del tejido colorrectal de sujetos con enfermedad inflamatoria intestinal y cáncer colorrectal con el estado de metilación de los mismos marcadores de ADN de sujetos de control (p. ej., tejido normal tal como colon normal) (ver, Ejemplo 1 y Tabla 2).

[0138] Además, 185, 244, y 111 de marcadores de metilación ADN específicos para los cánceres colorrectales, adenomas grandes, y pólipos de sésiles serrados, respectivamente, se identificaron (véase el Ejemplo 2 y en las Tablas 3, 4 y 5). Junto con los casos de cáncer colorrectal, se secuenciaron casos de adenoma grande y pólipos de sésiles serrados, mucosa colónica normal y ADN normal de glóbulos blancos.

[0139] Experimentos adicionales realizados durante el desarrollo de realizaciones para la presente divulgación demostraron que NDRG4, BMP3, OPLAH, FLI1, PDGFD, CHST_7889, SFMBT2_895, SFMBT2_896, SFMBT2_897, CHST2_7890, VAV3 y DTX1 como marcadores efectivos para detectar cáncer colorrectal con muestras de heces (ver, Ejemplo 3 y Tablas 6 y 7).

[0140] En consecuencia, se divulga en el presente documento una tecnología para marcadores de detección de cáncer colorrectal que proporcionan una alta relación señal/ruido y un bajo nivel de fondo cuando se detecta a partir de muestras tomadas de un sujeto (por ejemplo, muestra de heces; una muestra de tejido colorrectal). Los marcadores se identificaron en un estudio de casos y controles comparando el estado de metilación de los marcadores de ADN del tejido colorrectal de sujetos con neoplasia colorrectal y/o adenoma con el estado de metilación de los mismos marcadores de ADN de los sujetos control (p. ej., tejido normal como el colon normal) (véanse los Ejemplos 1-3 y las Tablas 1-6).

[0141] Aunque la descripción en el presente documento se refiere a ciertas realizaciones ilustradas, debe entenderse que estas realizaciones se presentan a modo de ejemplo y no a modo de limitación.

[0142] En aspectos particulares, la presente tecnología se refiere a composiciones y métodos para identificar, determinar y/o clasificar un cáncer colorrectal. En algunas realizaciones, los métodos comprenden determinar el estado de metilación de al menos un marcador de metilación en una muestra biológica aislada de un sujeto (por ejemplo, una muestra de heces o una muestra de tejido colorrectal), en donde un cambio en el estado de metilación del marcador es indicativo de la presencia, clase o sitio de un cáncer colorrectal. Las realizaciones particulares se refieren a marcadores que comprenden una región metilada diferencialmente (DMR, por ejemplo, una DMR como se proporciona en las Tablas 1-6) que se utilizan para el diagnóstico o la detección de trastornos proliferativos celulares neoplásicos (por ejemplo, cáncer colorrectal), incluida la detección temprana durante estadios pre-cancerosos de la enfermedad. Además, los marcadores se usan para la diferenciación de trastornos proliferativos celulares neoplásicos

de benignos. En aspectos particulares, la presente tecnología proporciona un método en donde un trastorno proliferativo de células neoplásicas se distingue de un trastorno proliferativo de células benignas.

5 **[0143]** Los marcadores de la presente tecnología son particularmente eficaces para detectar o distinguir entre trastornos proliferativos colorrectales, proporcionando de este modo medios mejorados para la detección temprana, clasificación y tratamiento de cáncer colorrectal.

10 **[0144]** Además de las realizaciones en las que se analiza el análisis de metilación de al menos un marcador, una región de un marcador o una base de un marcador que comprende un DMR (por ejemplo, un DMR como se proporciona en las Tablas 1-6) proporcionado aquí, la tecnología también proporciona paneles de marcadores que comprenden al menos un marcador, región de un marcador o base de un marcador que comprende un DMR con utilidad para la detección de cánceres colorrectales. Además de las realizaciones en las que se analiza el análisis de metilación de al menos un marcador, una región de un marcador o una base de un marcador que comprende un DMR (por ejemplo, un DMR como se proporciona en las Tablas 1-6) que se proporciona en el presente documento, la tecnología también proporciona paneles de marcadores que comprenden cualquier tipo o clase de fabricantes (por ejemplo, marcador completo, región de un marcador, base de un marcador, etc.) que tienen utilidad para la detección de cánceres colorrectales (por ejemplo, un marcador de expresión, cantidad de ADN, péptido, hemoglobina, etc.).

20 **[0145]** Algunas realizaciones de la tecnología se basan en el análisis del estado de metilación de CpG de al menos un marcador, región de un marcador o base de un marcador que comprende un DMR.

25 **[0146]** En algunas realizaciones, la presente tecnología proporciona el uso de una técnica de bisulfito, en combinación con uno o más ensayos de metilación para determinar el estado de metilación de secuencias de dinucleótidos CpG dentro de al menos un marcador que comprende un DMR (por ejemplo, un DMR como se proporciona en las Tablas 1-6). Los dinucleótidos CpG genómicos pueden estar metilados o no metilados (alternativamente conocidos como metilados hacia arriba y hacia abajo, respectivamente). Sin embargo, los métodos de la presente divulgación son adecuados para el análisis de muestras biológicas de naturaleza heterogénea, por ejemplo, una baja concentración de células tumorales o materiales biológicos a partir de ellas, dentro de un fondo de una muestra remota (por ejemplo, sangre, efluente de órganos o heces). En consecuencia, al analizar el estado de metilación de una posición de CpG dentro de una muestra de este tipo, se puede usar un ensayo cuantitativo para determinar el nivel (por ejemplo, porcentaje, fracción, relación, proporción o grado) de metilación en una posición de CpG particular.

35 **[0147]** La determinación del estado de metilación de secuencias de dinucleótidos CpG en marcadores que comprenden una DMR tiene utilidad tanto en el diagnóstico como en la caracterización de cánceres colorrectales.

Combinaciones de marcadores

40 **[0148]** En algunas realizaciones, la tecnología se relaciona con la evaluación del estado de metilación de combinaciones de marcadores que comprenden un DMR de las Tablas 1-6 (por ejemplo, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 27, 29, 30) o más marcadores que comprenden un DMR. En algunas realizaciones, evaluar el estado de metilación de más de un marcador aumenta la especificidad y/o sensibilidad de un cribado o diagnóstico para identificar una neoplasia colorrectal en un sujeto. Además de las realizaciones en las que se analiza el análisis de metilación de al menos un marcador, una región de un marcador o una base de un marcador que comprende un DMR (por ejemplo, un DMR como se proporciona en las Tablas 1-6) que se proporciona en el presente documento, también se analiza la tecnología proporciona paneles de marcadores que comprenden cualquier tipo o clase de fabricantes (por ejemplo, marcador completo, región de un marcador, base de un marcador, etc.) que tienen utilidad para la detección de cánceres colorrectales (por ejemplo, un marcador de expresión, cantidad de ADN, péptido, hemoglobina, etc.).

50 **[0149]** Se predicen varios cánceres mediante diversas combinaciones de marcadores, por ejemplo, como se identifica mediante técnicas estadísticas relacionadas con la especificidad y la sensibilidad de la predicción. La tecnología proporciona métodos para identificar combinaciones predictivas y combinaciones predictivas validadas para algunos tipos de cáncer.

Métodos para analizar el estado de metilación

60 **[0150]** Un método para analizar un ácido nucleico para la presencia de 5-metilcitosina se basa en el método de bisulfito descrito por Frommer, et al. para la detección de 5-metilcitosinas en el ADN (Frommer et al. (1992) Proc. Natl. Acad. Sci. EE.UU. 89: 1827-31) o variaciones de las mismas. El método de bisulfito para mapear las 5-metilcitosinas se basa en la observación de que la citosina, pero no la 5-metilcitosina, reacciona con el ion sulfito de hidrógeno (también conocido como bisulfito). En algunas realizaciones, la reacción se realiza de acuerdo con los siguientes pasos: primero, la citosina reacciona con sulfito de hidrógeno para formar una citosina sulfonada. A continuación, la desaminación espontánea del intermedio de reacción sulfonado da como resultado un uracilo sulfonado. Finalmente, el uracilo sulfonado se desulfona en condiciones alcalinas para formar uracilo. La detección es posible porque el uracilo forma pares de bases con adenina (comportándose así como timina), mientras que los pares de bases de 5-metilcitosina con guanina (comportándose así como citosina). Esto hace posible la discriminación de las citosinas metiladas de las

citiosinas no metiladas mediante, por ejemplo, la secuenciación genómica de bisulfito (Grigg G y Clark S, *Bioessays* (1994) 16: 431-36; Grigg G, *DNA Seq.* (1996) 6: 189 -98) o PCR específica de metilación (MSP) como se describe, por ejemplo, en la Patente de EE.UU. Nº 5,786,146.

5 **[0151]** Algunas tecnologías convencionales están relacionadas con métodos que comprenden encerrar el ADN para analizar en una matriz de agarosa, evitando así la difusión y renaturalización del ADN (el bisulfito solo reacciona con el ADN monocatenario), y reemplazando los pasos de precipitación y purificación con una diálisis rápida (Olek A, et al. (1996) "A modified and improved method for bisulfite based cytosine methylation analysis" *Nucleic Acids Res.* 24: 5064-6). Por lo tanto, es posible analizar el estado de metilación de las células individuales, ilustrando la utilidad y la
10 sensibilidad del método. Rein, T., y cols. proporcionan una visión general de los métodos convencionales para detectar 5-metilcitosina. (1998) *Nucleic Acids Res.* 26: 2255.

[0152] La técnica de bisulfito típicamente implica amplificar fragmentos cortos y específicos de un ácido nucleico conocido después de un tratamiento con bisulfito, y luego analizar el producto por secuenciación (Olek y Walter (1997) *Nat. Genet.* 17: 275-6) o una reacción de extensión del cebador (Gonzalzo & Jones (1997) *Nucleic Acids Res.* 25: 2529-31; WO 95/00669; Patente de EE.UU. Nº 6,251,594) para analizar posiciones individuales de citosina. Algunos métodos utilizan la digestión enzimática (Xiong y Laird (1997) *Nucleic Acids Res.* 25: 2532-4). La detección por hibridación también se ha descrito en la técnica (Olek et al., WO 99/28498). Además, se ha descrito el uso de la técnica de bisulfito para la detección de metilación con respecto a genes individuales (Grigg y Clark (1994) *Bioessays* 16: 431-6; Zeschnigk et al. (1997) *Hum Mol Genet.* 6: 387-95; Feil et al. (1994) *Nucleic Acids Res.* 22: 695; Martin et al. (1995) *Gene* 157: 261-4; WO 9746705; WO 9515373).

[0153] Se conocen varios procedimientos de ensayo de metilación en la técnica y se pueden usar junto con el tratamiento con bisulfito de acuerdo con la tecnología actual. Estos ensayos permiten la determinación del estado de metilación de uno o una pluralidad de dinucleótidos CpG (por ejemplo, islas CpG) dentro de una secuencia de ácido nucleico. Dichos ensayos implican, entre otras técnicas, la secuenciación de ácido nucleico tratado con bisulfito, PCR (para amplificación específica de secuencia), análisis de transferencia Southern y uso de enzimas de restricción sensibles a la metilación.

30 **[0154]** Por ejemplo, la secuenciación genómica se ha simplificado para el análisis de patrones de metilación y distribuciones de 5-metilcitosina mediante el tratamiento con bisulfito (Frommer et al. (1992) *Proc. Natl. Acad. Sci. EE.UU.* 89: 1827-1831). Además, la digestión con enzimas de restricción de productos de PCR amplificados a partir de ADN convertido con bisulfito encuentra uso en la evaluación del estado de metilación, por ejemplo, según lo descrito por Sadri y Hornsby (1997) *Nucl. Acidos Res.* 24: 5058-5059 o como se incorpora en el método conocido como COBRA (Combined Bisulfite Restriction Analysis) (Xiong y Laird (1997) *Nucleic Acids Res.* 25: 2532-2534).

[0155] El análisis COBRA™ es un ensayo de metilación cuantitativa útil para determinar los niveles de metilación del ADN en loci específico en pequeñas cantidades de ADN genómico (Xiong y Laird, *Nucleic Acids Res.* 25: 2532-2534, 1997). Brevemente, la digestión con enzimas de restricción se usa para revelar diferencias de secuencia dependientes de la metilación en productos de PCR de ADN tratado con bisulfito de sodio. Las diferencias de secuencia dependientes de la metilación se introducen primero en el ADN genómico mediante tratamiento con bisulfito estándar de acuerdo con el procedimiento descrito por Frommer et al. (*Proc. Natl. Acad. Sci. EE.UU.* 89: 1827-1831, 1992). La amplificación por PCR del ADN convertido con bisulfito se realiza utilizando cebadores específicos para las islas CpG de interés, seguido de digestión con endonucleasa de restricción, electroforesis en gel y detección mediante sondas de hibridación marcadas específicas. Los niveles de metilación en la muestra de ADN original están representados por las cantidades relativas de producto de PCR digerido y no digerido de forma linealmente cuantitativa en un amplio espectro de niveles de metilación de ADN. Además, esta técnica se puede aplicar de manera confiable al ADN obtenido de muestras de tejido embebido en parafina microdisecionadas.

50 **[0156]** Los reactivos típicos (por ejemplo, como se podría encontrar en un kit basado en COBRA™ típica) para análisis COBRA™ pueden incluir, pero no se limitan a: cebadores de PCR para loci específicos (por ejemplo, genes específicos, marcadores, DMR, regiones de genes, regiones de marcadores, secuencia de ADN tratada con bisulfito, isla CpG, etc.); enzima de restricción y tampón apropiado; oligonucleótido de hibridación génica; control de oligonucleótido de hibridación; kit de etiquetado de quinasa para sonda de oligonucleótidos; y nucleótidos marcados.
55 Además, los reactivos de conversión de bisulfito pueden incluir: tampón de desnaturalización de ADN; tampón de sulfonación; reactivos o kits de recuperación de ADN (por ejemplo, precipitación, ultrafiltración, columna de afinidad); tampón de desulfonación; y componentes de recuperación de ADN.

[0157] Preferiblemente, ensayos como "MethyLight™" (una técnica de PCR en tiempo real basada en fluorescencia) (Eads et al., *Cancer Res.* 59: 2302-2306, 1999), Ms-SNuPE™ (Methylation-sensitive Single Nucleotide Primer Extension) (Gonzalzo & Jones, *Nucleic Acids Res.* 25: 2529-2531, 1997), PCR específica de metilación ("MSP"; Herman et al., *Proc. Natl. Acad. Sci. EE.UU.* 93: 9821- 9826, 1996; Patente de EE.UU. Nº 5,786,146), y la amplificación de isla metílica CpG ("MCA"; Toyota et al., *Cancer Res.* 59: 2307-12, 1999) se usan solos o en combinación con uno o más de estos métodos.

65 **[0158]** La técnica de ensayo "HeavyMethyl™" es un método cuantitativo para evaluar diferencias de metilación basado

en la amplificación específica para la metilación de ADN tratado con bisulfito. Las sondas de bloqueo específicas de metilación ("bloqueadores") que cubren las posiciones de CpG entre, o cubiertas por, los cebadores de amplificación permiten la amplificación selectiva específica de metilación de una muestra de ácido nucleico.

5 **[0159]** El término ensayo de "HeavyMethyl™ MethyLight™" se refiere a un ensayo HeavyMethyl™ MethyLight™, que es una variación del ensayo MethyLight™, en donde el ensayo de MethyLight™ se combina con sondas de bloqueo específicas a metilación que cubren las posiciones de CpG entre los cebadores de amplificación. El ensayo HeavyMethyl™ también se puede usar en combinación con cebadores de amplificación específicos de metilación.

10 **[0160]** Los reactivos típicos (por ejemplo, como se podría encontrar en un kit típico basado en MethyLight™) para el análisis HeavyMethyl™ puede incluir, pero no se limitan a: cebadores de PCR para loci específicos (por ejemplo, genes específicos, marcadores, DMR, regiones de genes, regiones de marcadores, secuencia de ADN tratada con bisulfito, isla CpG, o secuencia de ADN tratada con bisulfito o isla CpG, etc.); oligonucleótidos de bloqueo; tampones de PCR optimizados y desoxinucleótidos; y polimerasa Taq.

15 **[0161]** La MSP (PCR específica a metilación) permite evaluar el estado de metilación de prácticamente cualquier grupo de sitios CpG dentro de una isla CpG, independientemente del uso de enzimas de restricción sensibles a la metilación (Herman et al. Proc. Natl. Acad. Sci EE.UU. 93: 9821-9826, 1996; Patente de EE.UU. N° 5,786,146). Brevemente, el ADN es modificado por bisulfito de sodio, que convierte las citosinas no metiladas, pero no metiladas, en uracilo, y los productos se amplifican posteriormente con cebadores específicos para ADN metilado versus no metilado. MSP requiere solo pequeñas cantidades de ADN, es sensible a los alelos metilados al 0,1% de un locus de isla CpG dado, y puede realizarse en ADN extraído de muestras incluidas en parafina. Los reactivos típicos (p. ej., como se puede encontrar en un kit típico basado en MSP) para el análisis de MSP pueden incluir, entre otros: cebadores de PCR metilados y no metilados para loci específicos (p. ej., genes específicos, marcadores, DMR, regiones de genes), regiones de marcadores, secuencia de ADN tratada con bisulfito, isla CpG, etc.); tampones de PCR optimizados y desoxinucleótidos, y sondas específicas.

20 **[0162]** El ensayo de MethyLight™ es un ensayo de metilación de alto rendimiento cuantitativo que utiliza PCR en tiempo real basado en la fluorescencia (por ejemplo, TaqMan®) que no requiere manipulaciones adicionales después del paso de PCR (Eads et al, Cancer Res.. 59: 2302-2306, 1999). Brevemente, el proceso MethyLight™ comienza con una muestra mixta de ADN genómico que se convierte, en una reacción de bisulfito de sodio, en un conjunto mixto de diferencias de secuencia dependientes de la metilación de acuerdo con los procedimientos estándar (el proceso de bisulfito convierte los residuos de citosina no metilados en uracilo). La PCR basada en fluorescencia se realiza luego en una reacción "sesgada", por ejemplo, con cebadores de PCR que se superponen a los dinucleótidos CpG conocidos. La discriminación de secuencia ocurre tanto al nivel del proceso de amplificación como al nivel del proceso de detección de fluorescencia.

30 **[0163]** El ensayo MethyLight™ se usa como una prueba cuantitativa para los patrones de metilación en un ácido nucleico, por ejemplo, una muestra de ADN genómico, en donde la discriminación de secuencia ocurre al nivel de hibridación de la sonda. En una versión cuantitativa, la reacción de PCR proporciona una amplificación específica de metilación en presencia de una sonda fluorescente que se superpone a un sitio de metilación putativo particular. Un control imparcial de la cantidad de ADN de entrada es proporcionado por una reacción en la que ni los cebadores ni la sonda se superponen a ningún dinucleótido CpG. Alternativamente, se logra una prueba cualitativa para la metilación genómica al sondear el conjunto de PCR sesgado con oligonucleótidos de control que no cubren los sitios de metilación conocidos (por ejemplo, una versión basada en fluorescencia de las técnicas HeavyMethyl™ y MSP) o con oligonucleótidos que cubren posibles sitios de metilación.

35 **[0164]** El proceso MethyLight™ se usa con cualquier sonda adecuada (por ejemplo, una sonda "TaqMan®", una sonda Lightcycler®, etc.) Por ejemplo, en algunas aplicaciones, el ADN genómico bicatenario se trata con bisulfito de sodio y se somete a uno de los dos conjuntos de reacciones de PCR usando sondas TaqMan®, por ejemplo, con cebadores MSP y/o oligonucleótidos bloqueadores de HeavyMethyl y una sonda TaqMan®. La sonda TaqMan® tiene doble etiqueta con moléculas fluorescentes de "indicador" y "desactivador" y está diseñada para ser específica para una región de contenido de GC relativamente alta de modo que se derrita a una temperatura aproximadamente 10°C más alta en el ciclo de PCR que la delantera o cebadores inversos. Esto permite que la sonda TaqMan® permanezca completamente hibridada durante el paso de recocido/extensión de PCR. A medida que la polimerasa Taq sintetiza enzimáticamente una nueva cadena durante la PCR, eventualmente alcanzará la sonda TaqMan® recocida. La actividad de endonucleasa Taq polimerasa 5' a 3' desplazará la sonda TaqMan® al digerirla para liberar la molécula informadora fluorescente para la detección cuantitativa de su señal ahora no apagada utilizando un sistema de detección fluorescente en tiempo real.

40 **[0165]** Los reactivos típicos (por ejemplo, como se podría encontrar en un kit típico basado en MethyLight™) para el análisis MethyLight™ pueden incluir, pero no se limitan a: cebadores de PCR para loci específicos (por ejemplo, genes específicos, marcadores, DMR, regiones de genes, regiones de marcadores, secuencia de ADN tratada con bisulfito, isla CpG, etc.); sondas TaqMan® o Lightcycler®; tampones de PCR optimizados y desoxinucleótidos; y polimerasa Taq.

60

65

[0166] El ensayo QM™ (metilación cuantitativa) es una prueba cuantitativa alternativa para los patrones de metilación en muestras de ADN genómico, en donde la discriminación de secuencia se produce al nivel de hibridación de la sonda. En esta versión cuantitativa, la reacción de PCR proporciona una amplificación imparcial en presencia de una sonda fluorescente que se superpone a un sitio de metilación putativo particular. Un control imparcial de la cantidad de ADN de entrada es proporcionado por una reacción en la que ni los cebadores ni la sonda se superponen a ningún dinucleótido CpG. Alternativamente, se logra una prueba cualitativa para la metilación genómica sondeando el conjunto de PCR sesgado con oligonucleótidos de control que no cubren los sitios de metilación conocidos (una versión basada en fluorescencia de las técnicas HeavyMethyl™ y MSP) o con oligonucleótidos que cubren los sitios de metilación potenciales.

[0167] El proceso QM™ puede usarse con cualquier sonda adecuada, por ejemplo, sondas "TaqMan®", sondas Lightcycler®, en el proceso de amplificación. Por ejemplo, el ADN genómico bicatenario se trata con bisulfito de sodio y se somete a cebadores no sesgados y a la sonda TaqMan®. La sonda TaqMan® tiene doble etiqueta con moléculas fluorescentes de "indicador" y "desactivador", y está diseñada para ser específica para una región de contenido de GC relativamente alta de modo que se derrita a una temperatura aproximadamente 10°C más alta en el ciclo de PCR que los cebadores hacia adelante o hacia atrás. Esto permite que la sonda TaqMan® permanezca completamente hibridada durante el paso de recocido/extensión de PCR. A medida que la polimerasa Taq sintetiza enzimáticamente una nueva cadena durante la PCR, eventualmente alcanzará la sonda TaqMan® recocida. La actividad de endonucleasa Taq polimerasa 5' a 3' desplazará la sonda TaqMan® al digerirla para liberar la molécula informadora fluorescente para la detección cuantitativa de su señal ahora no apagada utilizando un sistema de detección fluorescente en tiempo real. Los reactivos típicos (p. ej., como se puede encontrar en un kit típico basado en QM™) para el análisis QM™ pueden incluir, entre otros: cebadores de PCR para loci específicos (p. ej., genes específicos, marcadores, DMR, regiones de genes, regiones de marcadores, secuencia de ADN tratada con bisulfito, isla CpG, etc.); sondas TaqMan® o Lightcycler®; tampones de PCR optimizados y desoxinucleótidos; y polimerasa Taq.

[0168] La técnica Ms-SNuPE™ es un método cuantitativo para evaluar las diferencias de metilación en sitios CpG específicos basados en el tratamiento con bisulfito de ADN, seguido de extensión de cebador de un solo nucleótido (Gonzalzo & Jones, *Nucleic Acids Res.* 25: 2529-2531, 1997). Brevemente, el ADN genómico se hace reaccionar con bisulfito de sodio para convertir la citosina no metilada en uracilo mientras se deja la 5-metilcitosina sin cambios. La amplificación de la secuencia diana deseada se realiza luego usando cebadores de PCR específicos para ADN convertido con bisulfito, y el producto resultante se aísla y se usa como plantilla para el análisis de metilación en el sitio de interés CpG. Se pueden analizar pequeñas cantidades de ADN (p. ej., secciones de patología microdisecionadas) y se evita la utilización de enzimas de restricción para determinar el estado de metilación en los sitios CpG.

[0169] Los reactivos típicos (por ejemplo, como se podría encontrar en un kit típico basado en Ms-SNuPE™) para análisis Ms-SNuPE™ pueden incluir, pero no se limitan a: cebadores de PCR para loci específicos (por ejemplo, genes específicos, marcadores, DMR, regiones de genes, regiones de marcadores, secuencia de ADN tratada con bisulfito, isla CpG, etc.); tampones de PCR optimizados y desoxinucleótidos; kit de extracción de gel; cebadores de control positivo; cebadores Ms-SNuPE™ para loci específicos; tampón de reacción (para la reacción Ms-SNuPE); y nucleótidos marcados. Además, los reactivos de conversión de bisulfito pueden incluir: tampón de desnaturalización de ADN; tampón de sulfonación; reactivos o kit de recuperación de ADN (por ejemplo, precipitación, ultrafiltración, columna de afinidad); tampón de desulfonación; y componentes de recuperación de ADN.

[0170] La secuenciación de bisulfito de representación reducida (RRB) comienza con el tratamiento de bisulfito de ácido nucleico para convertir todas las citosinas no metiladas en uracilo, seguido de digestión con enzimas de restricción (por ejemplo, por una enzima que reconoce un sitio que incluye una secuencia CG tales como MspI) y completa secuenciación de fragmentos después del acoplamiento a un ligando adaptador. La elección de la enzima de restricción enriquece los fragmentos de las regiones densas de CpG, reduciendo el número de secuencias redundantes que pueden mapearse en múltiples posiciones de genes durante el análisis. Como tal, RRBS reduce la complejidad de la muestra de ácido nucleico seleccionando un subconjunto (por ejemplo, mediante selección de tamaño usando electroforesis en gel preparativa) de fragmentos de restricción para la secuenciación. A diferencia de la secuenciación de bisulfito de genoma completo, cada fragmento producido por la digestión con enzimas de restricción contiene información de metilación del ADN para al menos un dinucleótido CpG. Como tal, RRBS enriquece la muestra para promotores, islas CpG y otras características genómicas con una alta frecuencia de sitios de corte de enzimas de restricción en estas regiones y, por lo tanto, proporciona un ensayo para evaluar el estado de metilación de uno o más loci genómicos.

[0171] Un protocolo típico para RRBS comprende los pasos de digerir una muestra de ácido nucleico con una enzima de restricción como MspI, rellenar voladizos y colas A, adaptadores de ligadura, conversión de bisulfito y PCR. Ver, por ejemplo, et al. (2005) *Nat Methods* 7: 133-6; Meissner y col. (2005) *Nucleic Acids Res.* 33: 5868-77.

[0172] En algunas realizaciones, se usa una diana en tiempo real específica a alelo cuantitativo y ensayo de amplificación de señal (QuARTS) para evaluar el estado de metilación. Se producen tres reacciones secuencialmente en cada ensayo QuARTS, incluida la amplificación (reacción 1) y la escisión de la sonda diana (reacción 2) en la reacción primaria; y escisión FRET y generación de señal fluorescente (reacción 3) en la reacción secundaria. Cuando

el ácido nucleico diana se amplifica con cebadores específicos, una sonda de detección específica con una secuencia de colgajo se une libremente al amplicón. La presencia del oligonucleótido invasivo específico en el sitio de unión a la diana hace que la escisión libere la secuencia del colgajo cortando entre la sonda de detección y la secuencia del colgajo. La secuencia de la aleta es complementaria a una porción sin pinza de un casete FRET correspondiente. En consecuencia, la secuencia del colgajo funciona como un oligonucleótido invasivo en el casete FRET y produce una escisión entre el fluoróforo del casete FRET y un apagador, que produce una señal fluorescente. La reacción de escisión puede cortar múltiples sondas por diana y así liberar múltiples fluoróforos por colgajo, proporcionando una amplificación de señal exponencial. QuARTS puede detectar múltiples dianas en un solo pozo de reacción mediante el uso de casetes FRET con diferentes tintes. Ver, por ejemplo, en Zou et al. (2010) "Sensitive quantification of methylated markers with a novel methylation specific technology" Clin Chem 56: A199; Pat. N° 8,361,720 y la Patente de EE.UU. Sol. Ser. N°s 13/594,674, 12/946,745 y 12/946,752.

[0173] En algunas realizaciones, el ácido nucleico diana se aísla de una muestra mediante, por ejemplo, una captura directa de genes. Por ejemplo, en algunas realizaciones, el ácido nucleico diana se aísla de una muestra mediante, por ejemplo, la eliminación de agentes inhibidores del ensayo para producir una muestra clarificada (por ejemplo, con PVP, PVPP y/o el uso de un filtro giratorio), captura de un ácido nucleico diana (si está presente) de la muestra clarificada con un reactivo de captura para formar un complejo de captura, aislando el complejo de captura de la muestra clarificada, recuperando el ácido nucleico diana (si está presente) del complejo de captura en una solución de ácido nucleico, y, opcionalmente, repetir el aislamiento de diferentes dianas (véanse, por ejemplo, las Solicitudes de Patente de EE.UU. N°s de Serie 14/145,082, 14/145,087, 14/145,070, 14/145,056, 13/470,251, 13/470,018, 13/469,999 y 13/469,989).

[0174] En algunas realizaciones, los fragmentos del ADN tratado se amplifican usando conjuntos de oligonucleótidos cebadores (por ejemplo, véase la Tabla 2) y una enzima de amplificación. La amplificación de varios segmentos de ADN puede llevarse a cabo simultáneamente en uno y el mismo recipiente de reacción. Típicamente, la amplificación se lleva a cabo usando una reacción en cadena de la polimerasa (PCR). Los amplicones son típicamente de 100 a 2000 pares de bases de longitud.

[0175] En otra realización del método, el estado de metilación de las posiciones de CpG dentro o cerca de un marcador que comprende un DMR (por ejemplo, un DMR como se proporciona en las Tablas 1-6) puede detectarse mediante el uso de oligonucleótidos cebadores específicos de metilación. Esta técnica (MSP) se ha descrito en la Patente de EE.UU. N° 6,265,171 de Herman. El uso de cebadores específicos del estado de metilación para la amplificación de ADN tratado con bisulfito permite la diferenciación entre ácidos nucleicos metilados y no metilados. Los pares de cebadores MSP contienen al menos un cebador que se hibrida con un dinucleótido CpG tratado con bisulfito. Por lo tanto, la secuencia de dichos cebadores comprende al menos un dinucleótido CpG. Los cebadores de MSP específicos para ADN no metilado contienen una "T" en la posición de la posición C en el CpG.

[0176] Los fragmentos obtenidos por medio de la amplificación pueden llevar un marcador detectable directa o indirectamente. En algunas realizaciones, los marcadores son marcadores fluorescentes, radionucleidos o fragmentos de moléculas desmontables que tienen una masa típica que puede detectarse en un espectrómetro de masas. Cuando dichas etiquetas son etiquetas de masa, algunas realizaciones proporcionan que los amplicones marcados tengan una sola carga neta positiva o negativa, lo que permite una mejor detectabilidad en el espectrómetro de masas. La detección puede llevarse a cabo y visualizarse mediante, por ejemplo, espectrometría de masas por ionización/desorción asistida por matriz (MALDI) o utilizando espectrometría de masas por pulverización electrónica (ESI).

[0177] En algunas realizaciones, los métodos para aislar ADN comprenden el aislamiento de ácidos nucleicos como se describe en la Patente de EE.UU. Sol. Ser. N° 13/470,251 ("Isolation of Nucleic Acids").

50 Métodos

[0178] En algunas realizaciones de la tecnología, se describen métodos que comprenden los siguientes pasos:

- 1) poner en contacto un ácido nucleico (p. ej., ADN genómico, p. ej., aislado de fluidos corporales como una muestra de heces o tejido colorrectal) obtenido del sujeto con al menos un reactivo o serie de reactivos que distingue entre dinucleótidos CpG metilados y no metilados dentro de al menos un marcador DMR (por ejemplo, un DMR como se proporciona en las Tablas 1-6) y
- 2) detectar una neoplasia colorrectal o trastorno proliferativo (por ejemplo, cáncer colorrectal, adenoma grande, SSP) (por ejemplo, con una sensibilidad mayor o igual al 80% y una especificidad mayor o igual al 80%).

[0179] En algunas realizaciones, se describen métodos que comprenden los siguientes pasos:

- 1) poner en contacto un ácido nucleico (por ejemplo, ADN genómico, por ejemplo, aislado de fluidos corporales, como una muestra de heces o tejido colorrectal) obtenido del sujeto con al menos un reactivo o serie de reactivos que distingue entre dinucleótidos CpG metilados y no metilados dentro de al menos un

marcador seleccionado de una región cromosómica que tiene una anotación seleccionada del grupo que consiste en BMP3, NDRG4, FLI1, OPLAH, DTX1, MATK, SFMBT2 región 2, KCNK12, VAV3 región 1, SFMBT2 región 3, PPP2R5C, CHST2 región 2, PKIA, PDGFD, ELOVL2, CHST2 región 1, SFMBT2 región 1, QKI, VAV3 región 2 y SLC8A3, y

2) detectar el cáncer colorrectal (por ejemplo, con una sensibilidad mayor o igual al 80% y una especificidad mayor o igual al 80%).

[0180] Preferiblemente, la sensibilidad es de aproximadamente 70% a aproximadamente 100%, o de aproximadamente 80% a aproximadamente 90%, o de aproximadamente 80% a aproximadamente 85%. En algunas realizaciones, la especificidad es de aproximadamente 70% a aproximadamente 100%, o de aproximadamente 80% a aproximadamente 90%, o de aproximadamente 80% a aproximadamente 85%. En algunas realizaciones, la especificidad es de aproximadamente 90% a 100%, 91% a 99%, 93% a 97%, 94% a 96%, 95% a 99%, 96% a 99,5%, 97% a 99,9%, etc.)

[0181] El ADN genómico se puede aislar por cualquier medio, incluyendo el uso de kits disponibles comercialmente. Brevemente, en donde el ADN de interés está encapsulado por una membrana celular, la muestra biológica debe ser interrumpida y lisada por medios enzimáticos, químicos o mecánicos. La solución de ADN se puede eliminar de proteínas y otros contaminantes, por ejemplo, por digestión con proteinasa K. El ADN genómico se recupera de la solución. Esto puede llevarse a cabo mediante una variedad de métodos que incluyen la salazón, la extracción orgánica o la unión del ADN a un soporte en fase sólida. La elección del método se verá afectada por varios factores, incluidos el tiempo, los gastos y la cantidad requerida de ADN. Todos los tipos de muestra que comprenden materia neoplásica o materia preneoplásica son adecuados para usar en el presente método, por ejemplo, líneas celulares, portaobjetos histológicos, biopsias, tejido embebido en parafina, fluidos corporales, heces, efluentes del colon, orina, plasma sanguíneo, suero sanguíneo., sangre completa, células sanguíneas aisladas, células aisladas de la sangre y combinaciones de las mismas.

[0182] La tecnología no está limitada en los métodos utilizados para preparar las muestras y proporcionar un ácido nucleico para la prueba. Por ejemplo, en algunas realizaciones, se aísla un ADN de una muestra de heces o de sangre o de una muestra de plasma usando captura directa de genes, por ejemplo, como se detalla en las Solicitudes de Patente de EE.UU. Nos de Serie 14/145,082, 14/145,087, 14/145,070, 14/145,056, 13/470,251, 13/470,018, 13/469,999 y 13/469,989.

[0183] La muestra de ADN genómico se trata luego con al menos un reactivo, o una serie de reactivos, que distingue entre dinucleótidos CpG metilados y no metilados dentro de al menos un marcador que comprende una DMR (por ejemplo, una DMR como se proporciona en las Tablas 1-6).

[0184] En algunas realizaciones, el reactivo convierte bases de citosina que no están metiladas en la posición 5' en uracilo, timina u otra base que es diferente a la citosina en términos de comportamiento de hibridación. Sin embargo, en algunas realizaciones, el reactivo puede ser una enzima de restricción sensible a la metilación.

[0185] En algunas realizaciones, la muestra de ADN genómico se trata de tal manera que las bases de citosina que no están metiladas en la posición 5' se convierten en uracilo, timina u otra base que es diferente a la citosina en términos de comportamiento de hibridación. En algunas realizaciones, este tratamiento se lleva a cabo con bisulfato (sulfito de hidrógeno, disulfito) seguido de hidrólisis alcalina.

[0186] A continuación se analiza el ácido nucleico tratado para determinar el estado de metilación de las secuencias de genes diana (por lo menos un gen, secuencia genómica o nucleótido de un marcador que comprende un DMR, por ejemplo, al menos un DMR como proporcionado en las Tablas 1- 6). En algunas realizaciones, el método de análisis es QuARTS y/o MSP como se describe aquí.

[0187] Metilación aberrante, más específicamente la hipermetilación de un marcador que comprende un DMR (por ejemplo, un DMR como proporcionado en las Tablas 1-6) está asociada con un cáncer colorrectal.

[0188] La tecnología se relaciona con el análisis de cualquier muestra asociada con un cáncer colorrectal. Por ejemplo, en algunas realizaciones, la muestra comprende un tejido y/o fluido biológico obtenido de un paciente. En algunas realizaciones, la muestra comprende una secreción. En algunas realizaciones, la muestra comprende sangre, suero, plasma, secreciones gástricas, tejido colorrectal, tejido tumoral colorrectal, una muestra de biopsia colorrectal, jugo pancreático, una muestra de biopsia gastrointestinal, células microdisecionadas de una biopsia gastrointestinal, células gastrointestinales caídas en la luz gastrointestinal, y/o células gastrointestinales recuperadas de las heces. En algunas realizaciones, el sujeto es humano. Estas muestras pueden originarse en el tracto gastrointestinal superior, el tracto gastrointestinal inferior, o comprender células, tejidos y/o secreciones tanto del tracto gastrointestinal superior como del tracto gastrointestinal inferior. La muestra puede incluir células, secreciones o tejidos del hígado, conductos biliares, páncreas, estómago, colon, recto, esófago, intestino delgado, apéndice, duodeno, pólipos, vesícula biliar, ano y/o peritoneo. En algunas realizaciones, la muestra comprende líquido celular, ascitis, orina, heces, líquido pancreático, líquido obtenido durante la endoscopia, sangre, moco o saliva. En algunas realizaciones, la muestra es una muestra de heces.

- 5 [0189] Dichas muestras se pueden obtener por cualquier variedad de técnicas. Por ejemplo, las muestras de orina y heces son fácilmente obtenibles, mientras que las muestras de sangre, ascitis, suero, líquido colorrectal o pancreático se pueden obtener por vía parenteral utilizando una aguja y una jeringa, por ejemplo. Se pueden obtener muestras sin células o sustancialmente sin células sometiendo la muestra a varias técnicas que incluyen, pero no se limitan a, centrifugación y filtración. Aunque generalmente se prefiere que no se usen técnicas invasivas para obtener la muestra, aún puede ser preferible obtener muestras tales como homogeneizados de tejidos, secciones de tejidos y muestras de biopsia.
- 10 [0190] La tecnología se puede utilizar para tratar a un paciente (por ejemplo, un paciente con cáncer colorrectal, con cáncer colorrectal en etapa temprana, o que puede desarrollar cáncer colorrectal), comprendiendo el método determinar el estado de metilación de una o más DMR como se proporciona en este documento y administrar un tratamiento al paciente basado en los resultados de determinar el estado de metilación. El tratamiento puede realizar una colonoscopia, administrar un compuesto farmacéutico, una vacuna, realizar una cirugía, tomar imágenes del paciente, realizar otra prueba. Preferiblemente, dicho uso es en un método de detección clínica, un método de evaluación del pronóstico, un método para monitorear los resultados de la terapia, un método para identificar a los pacientes con más probabilidades de responder a un tratamiento terapéutico particular, un método para obtener imágenes de un paciente o sujeto, y un método para la detección y desarrollo de drogas.
- 15 [0191] En algunas realizaciones, el pronóstico clínico del cáncer incluye determinar la agresividad del cáncer y la probabilidad de recurrencia tumoral para planificar la terapia más eficaz. Si se puede hacer un pronóstico más preciso o incluso se puede evaluar un riesgo potencial de desarrollar cáncer, se puede elegir una terapia adecuada y, en algunos casos, una terapia menos severa para el paciente. La evaluación (por ejemplo, determinar el estado de metilación) de los biomarcadores de cáncer es útil para separar a los sujetos con buen pronóstico y/o bajo riesgo de desarrollar cáncer que no necesitarán terapia o terapia limitada de aquellos con más probabilidades de desarrollar cáncer o sufrir una recurrencia de cáncer que podría beneficiarse de tratamientos o monitoreo más intensivos.
- 20 [0192] En algunas realizaciones de la materia actualmente descrita, se realizan múltiples determinaciones de los biomarcadores a lo largo del tiempo para facilitar el diagnóstico y/o pronóstico. Se utiliza un cambio temporal en el biomarcador para predecir un resultado clínico, controlar la progresión del cáncer gastrointestinal y/o controlar la eficacia de las terapias apropiadas dirigidas contra el cáncer. En tal realización, por ejemplo, se podría esperar ver un cambio en el estado de metilación de uno o más biomarcadores (p. ej., DMR) divulgados aquí (y potencialmente uno o más biomarcadores adicionales, si se monitorean) en una muestra biológica durante tiempo durante el curso de una terapia efectiva.
- 25 [0193] Los métodos y composiciones de la presente descripción pueden emplearse para el tratamiento o diagnóstico de la enfermedad en una etapa temprana, por ejemplo, antes de que aparezcan los síntomas de la enfermedad.
- 30 [0194] En algunas realizaciones, un análisis estadístico asocia un indicador pronóstico con una predisposición a un resultado adverso. Por ejemplo, en algunas realizaciones, un estado de metilación diferente del de una muestra de control normal obtenida de un paciente que no tiene cáncer puede indicar que un sujeto tiene más probabilidades de sufrir un cáncer que los sujetos con un nivel que es más similar al estado de metilación en la muestra de control, según lo determinado por un nivel de significación estadística. Además, un cambio en el estado de metilación desde un nivel basal (p. ej., "normal") puede reflejar el pronóstico del sujeto, y el grado de cambio en el estado de metilación puede estar relacionado con la gravedad de los eventos adversos. La significación estadística a menudo se determina comparando dos o más poblaciones y determinando un intervalo de confianza y/o un valor p . Véase, por ejemplo, Dowdy y Wearden, *Statistics for Research*, John Wiley & Sons, Nueva York, 1983. Los intervalos de confianza ejemplares del presente tema son 90%, 95%, 97,5%, 98%, 99%, 99,5%, 99,9 % y 99,99%, mientras que los valores p ejemplares son 0,1, 0,05, 0,025, 0,02, 0,01, 0,005, 0,001 y 0,0001.
- 35 [0195] En otras realizaciones, se puede establecer un grado umbral de cambio en el estado de metilación de un biomarcador de pronóstico o diagnóstico descrito en este documento (por ejemplo, una DMR), y el grado de cambio en el estado de metilación del marcador biológico en una muestra biológica simplemente se compara con el grado umbral de cambio en el estado de metilación. Un cambio de umbral preferido en el estado de metilación para los biomarcadores proporcionados aquí es aproximadamente 5%, aproximadamente 10%, aproximadamente 15%, aproximadamente 20%, aproximadamente 25%, aproximadamente 30%, aproximadamente 50%, aproximadamente 75%, aproximadamente 100% y aproximadamente 150%. En otras realizaciones más, se puede establecer un "nomograma", mediante el cual el estado de metilación de un indicador de pronóstico o diagnóstico (biomarcador o combinación de biomarcadores) está directamente relacionado con una disposición asociada hacia un resultado dado.
- 40 El experto en la técnica está familiarizado con el uso de tales nomogramas para relacionar dos valores numéricos con el entendimiento de que la incertidumbre en esta medición es la misma que la incertidumbre en la concentración del marcador porque se hace referencia a mediciones de muestras individuales, no a promedios de población.
- 45 [0196] En algunas realizaciones, una muestra de control se analiza simultáneamente con la muestra biológica, de modo que los resultados obtenidos de la muestra biológica se pueden comparar con los resultados obtenidos de la muestra de control. Además, se contempla que se pueden proporcionar curvas estándar, con las cuales se pueden
- 50
- 55
- 60
- 65

comparar los resultados del ensayo para la muestra biológica. Dichas curvas estándar presentan estados de metilación de un biomarcador en función de las unidades de ensayo, por ejemplo, intensidad de señal fluorescente, si se usa un marcador fluorescente. Usando muestras tomadas de múltiples donantes, se pueden proporcionar curvas estándar para controlar los estados de metilación de uno o más biomarcadores en el tejido normal, así como para los niveles "en riesgo" de uno o más biomarcadores en el tejido tomado de donantes con metaplasia o de donantes con cáncer gastrointestinal. En ciertas realizaciones del método, se identifica a un sujeto que tiene metaplasia al identificar un estado de metilación aberrante de uno o más DMR proporcionados aquí en una muestra biológica obtenida del sujeto. En otras realizaciones del método, la detección de un estado de metilación aberrante de uno o más de tales biomarcadores en una muestra biológica obtenida del sujeto da como resultado que el sujeto sea identificado como que tiene cáncer.

[0197] El análisis de marcadores puede llevarse a cabo por separado o simultáneamente con marcadores adicionales dentro de una muestra de ensayo. Por ejemplo, se pueden combinar varios marcadores en una sola prueba para el procesamiento eficiente de un múltiplo de muestras y para proporcionar potencialmente una mayor precisión diagnóstica y/o pronóstica. Además, un experto en la materia reconocería el valor de probar múltiples muestras (por ejemplo, en puntos de tiempo sucesivos) del mismo sujeto. Dichas pruebas de muestras en serie pueden permitir la identificación de cambios en los estados de metilación del marcador con el tiempo. Los cambios en el estado de metilación, así como la ausencia de cambios en el estado de metilación, pueden proporcionar información útil sobre el estado de la enfermedad que incluye, entre otros, la identificación del tiempo aproximado desde el inicio del evento, la presencia y la cantidad de tejido recuperable, la idoneidad de las terapias farmacológicas, la efectividad de varias terapias y la identificación del resultado del sujeto, incluido el riesgo de eventos futuros.

[0198] El análisis de biomarcadores puede llevarse a cabo en una variedad de formatos físicos. Por ejemplo, el uso de placas de microtitulación o automatización puede usarse para facilitar el procesamiento de grandes cantidades de muestras de prueba. Alternativamente, se pueden utilizar formatos de muestra única para facilitar el tratamiento inmediato y el diagnóstico de manera oportuna, por ejemplo, en transporte ambulatorio o en salas de emergencias.

[0199] En algunas realizaciones, al sujeto se le diagnostica un cáncer colorrectal si, en comparación con un estado de metilación de control, hay una diferencia apreciable en el estado de metilación de al menos un biomarcador en la muestra. Por el contrario, cuando no se identifica ningún cambio en el estado de metilación en la muestra biológica, se puede identificar al sujeto como no con cáncer colorrectal, sin riesgo de cáncer o con bajo riesgo de cáncer. A este respecto, los sujetos que tienen el cáncer o el riesgo del mismo pueden diferenciarse de los sujetos que tienen un cáncer o un riesgo bajo o sustancialmente nulo. Aquellos sujetos que corren el riesgo de desarrollar un cáncer colorrectal pueden someterse a un programa de detección más intensivo y/o regular, incluida la vigilancia endoscópica. Por otro lado, aquellos sujetos que tienen un riesgo bajo o prácticamente nulo pueden evitar ser sometidos a una endoscopia, hasta que un examen de detección futuro, por ejemplo, un examen de detección realizado de acuerdo con la tecnología actual, indique que apareció en esos sujetos un riesgo de cáncer colorrectal.

[0200] Como se mencionó anteriormente, dependiendo de la realización del método de la presente tecnología, detectar un cambio en el estado de metilación de uno o más biomarcadores puede ser una determinación cualitativa o puede ser una determinación cuantitativa. Como tal, el paso de diagnosticar que un sujeto tiene o está en riesgo de desarrollar un cáncer gastrointestinal indica que se realizan ciertas mediciones de umbral, por ejemplo, el estado de metilación de uno o más biomarcadores en la muestra biológica varía de un estado de metilación de control predeterminado. En algunas realizaciones del método, el estado de metilación de control es cualquier estado de metilación detectable del biomarcador. En otras realizaciones del método donde una muestra de control se prueba simultáneamente con la muestra biológica, el estado de metilación predeterminado es el estado de metilación en la muestra de control. En otras realizaciones del método, el estado de metilación predeterminado se basa y/o identifica mediante una curva estándar. En otras realizaciones del método, el estado de metilación predeterminado es un estado o rango de estado específico. Como tal, el estado de metilación predeterminado puede elegirse, dentro de límites aceptables que serán evidentes para los expertos en la técnica, en base en parte a la forma de realización del método siendo practicado y la especificidad deseada, etc.

[0201] La materia actualmente divulgada incluye además un sistema para diagnosticar un cáncer gastrointestinal en un sujeto. El sistema puede proporcionarse, por ejemplo, como un kit comercial que puede usarse para detectar un riesgo de cáncer gastrointestinal o diagnosticar un cáncer gastrointestinal en un sujeto del que se ha recogido una muestra biológica. Un sistema ejemplar proporcionado de acuerdo con la presente tecnología incluye evaluar el estado de metilación de un DMR como se proporciona en las Tablas 1-6.

EJEMPLOS

Ejemplo 1.

[0202] Este ejemplo describe estudios de tejido de control de casos independientes que identificaron marcadores de metilación altamente discriminantes para la neoplasia colorrectal. Los ejemplos utilizaron RRBS para la fase de descubrimiento y la PCR específica de metilación (MSP) y la diana en tiempo real específica a alelos cuantitativos y la amplificación de señal (QuART) para la fase de validación.

[0203] Se identificaron muestras de tejido de registros de cáncer existentes. La población accesible incluía a aquellos que se sometieron a colectomía abierta o laparoscópica, o biopsia de colon con una muestra archivada. Todos los tejidos fueron revisados por un patólogo gastrointestinal experto para confirmar la clasificación correcta. Los tejidos de casos de neoplasia colorrectal incluyeron cánceres colorrectales en estadios I-IV (congelados frescos), adenomas avanzados >1 cm de tamaño (congelados frescos) y pólipos de sésiles serrados del lado derecho (FFPE). Hubo dos grupos de control estudiados. El primer grupo de control incluyó 18 tejidos epiteliales de colon de pacientes confirmados como libres de neoplasia de colon. El segundo grupo incluyó 18 muestras de pelaje normal de pacientes sin cáncer. Los casos y ambos controles se combinaron por sexo y edad. Además, el CRC y las cohortes de adenoma avanzado se distribuyeron uniformemente entre las lesiones del lado derecho e izquierdo. En un laboratorio central, se microdisecionaron tejidos de casos y controles y se extrajo el ADN usando una técnica de fenol-cloroformo, produciendo al menos 500 ng de ADN. La identificación de casos, el emparejamiento y la extracción de ADN fueron realizados por personal independiente para mantener el cegamiento del personal de laboratorio al estado del caso y el control.

[0204] El ADN genómico (300 ng) se fragmentó mediante digestión con 10 unidades de MspI, una enzima de restricción específica para la metilación que reconoce motivos que contienen CpG. Esto enriquece las muestras para el contenido de CpG y elimina áreas redundantes del genoma. Los fragmentos digeridos se repararon por el extremo y se colocaron en cola A con 5 unidades de fragmento Klenow (3'-5' exo-), y se ligaron durante la noche a adaptadores TruSeq metilados (Illumina, San Diego CA) que contienen una de cuatro secuencias de códigos de barras (para unir cada uno fragmento a su ID de muestra). La selección del tamaño de los fragmentos de 160-340 pb (insertos de 40-220 pb) se realizó usando cuentas/tampón Agencourt AMPure XP SPRI (Beckman Coulter, Brea CA). Los límites de tampón fueron de 0,7X a 1,1X volúmenes de muestra de cuentas/tampón. El volumen de elución final fue de 22 μ L (tampón EB - Qiagen, Germantown MD). Se usó qPCR para medir la eficiencia de la ligadura y la calidad del fragmento en una pequeña alícuota de muestra. Luego, las muestras se sometieron a conversión de bisulfito (dos veces) utilizando un protocolo EpiTect modificado (Qiagen). qPCR y PCR convencional (PfuTurbo Cx hotstart - Agilent, Santa Clara CA) seguido de la evaluación Bioanalyzer 2100 (Agilent) en alícuotas de muestra convertidas determinaron el número de ciclo de PCR óptimo antes de la amplificación de la biblioteca final. Condiciones para la PCR final: 50uL rxn: 5uL de 10X tampón, 1,25uL de 10mM cada dNTP's, 5uL primer cocktail (~5uM), 15uL plantilla (muestra), 1uL PfuTurbo Cx hotstart, 22,75 agua. 95C-5min; 98C-30seg; 16 ciclos de 98C-10seg, 65C-30seg, 72C-30seg; 72C-5min; 4C. Las muestras se combinaron (equimolar) en bibliotecas de 4 plex con base en el esquema de aleatorización y se probaron con el bioanalizador para la verificación del tamaño final, y con qPCR usando estándares phiX y cebadores específicos del adaptador.

[0205] Las muestras se cargaron en carriles de célula de flujo de acuerdo con una asignación aleatoria de carriles con carriles adicionales reservados para controles de ensayo internos. La secuenciación fue realizada por el núcleo de secuenciación de próxima generación en el Centro de Genoma Médico de la Clínica Mayo en Illumina HiSeq 2000. Las lecturas fueron unidireccionales durante 101 ciclos. Cada carril de célula de flujo generó 100-120 millones de lecturas, suficiente para una cobertura media de 30-50 veces la profundidad de secuencia (número de lectura por CpG) para secuencias alineadas. Se usó el software estándar de canalización Illumina para llamadas de base y generación de lectura de secuencia en formato fastq. Como anteriormente (véase, por ejemplo, Sun, et al., 2012 *Bioinformatics* 28 (16): 2180-1), se utilizó SAAP-RRBS, una tubería de análisis y anotación simplificada para la secuenciación de bisulfito de representación reducida, para la alineación de secuencias y la extracción de metilación.

[0206] Se realizaron dos estudios de validación basados en MSP en conjuntos de muestras expandidas para confirmar la precisión y la reproducibilidad de los candidatos observados diferencialmente metilados. El primero, un estudio de *validación interna* (MSP, por sus siglas en inglés), se realizó en muestras emparejadas y ciegas utilizando réplicas biológicas y técnicas de neoplasia colorrectal, colon normal y leucocitos normales. Este paso se realizó para garantizar que los sitios de metilación diferencial identificados por la filtración de datos RRBS, donde el % de metilación era la unidad de análisis, se reflejarían en MSP, donde la unidad de análisis es el número absoluto de copias genómicas de la secuencia diana corregido por la concentración de ADN de entrada para cada muestra. El segundo experimento de *validación externa* utilizó la tecnología QuART para evaluar a los principales candidatos en neoplasia colorrectal independiente, asignada al azar, cegada e independiente y muestras de colon normales.

[0207] Los cebadores para cada marcador se diseñaron para apuntar a las secuencias metiladas modificadas con bisulfito de cada gen diana (IDT, Coralville IA) y una región sin sitios de citosina-fosfato-guanina en el gen de β -actina, como referencia del tratamiento con bisulfito. y entrada de ADN. El diseño se realizó mediante el software Methprimer (Universidad de California, San Francisco, CA) o mediante métodos semi-manuales. Los ensayos se probaron y optimizaron ejecutando qPCR con tintes SYBR Green (Life Technologies, Grand Island NY) en diluciones de controles de ADN genómico universalmente metilados y no metilados.

[0208] Las reacciones de MSP se realizaron en ADN extraído de tejido como se describió previamente (véase, por ejemplo, Kisiel, et al., 2012 *Cancer* 118 (10): 2623-2631). Brevemente, el ADN se trató con bisulfito usando el kit de metilación de ADN EZ (Zymo Research, Orange, CA) y se eluyó en tampón. Se usó 1 ml de ADN tratado con bisulfito como plantilla para la cuantificación de la metilación con una PCR en tiempo real basada en fluorescencia, realizada con la mezcla maestra SYBR Green (Roche, Mannheim, Alemania). Las reacciones se realizaron en Roche 480

LightCyclers (Indianápolis, IN), donde se usó ADN metilado universal CpGenome tratado con bisulfito (Millipore, Billerica, MA) como control positivo, y diluido en serie para crear curvas estándar para todas las placas. Las secuencias de oligonucleótidos y las temperaturas de recocido están disponibles a pedido.

5 **[0209]** La comparación primaria de interés era la diferencia de metilación entre casos y controles de colon en cada CpG asignada. Las islas CpG se definen bioquímicamente por una relación CpG observada a esperada superior a 0,6 (31). Sin embargo, para este modelo, se crearon unidades de análisis de CpG "región metilada diferencialmente (DMR)" en función de la distancia entre las ubicaciones del sitio CpG para cada cromosoma. Como la distancia entre cualquier CpG dada excedió la ubicación anterior o siguiente en más de 100 bps, se creó un nuevo identificador de isla. Se excluyeron las islas con un solo CpG. El resultado secundario fue la misma comparación entre casos y controles de leucocitos. Los sitios de CpG individuales se consideraron para el análisis diferencial solo si la profundidad total de cobertura por grupo de enfermedad fue ≥ 200 lecturas (lo que equivale aproximadamente a un promedio de 10 lecturas por sujeto) y la varianza del % de metilación fue mayor que cero (sitios de CpG no informativos con 0 varianza fueron excluidos). Los criterios para la profundidad de lectura se basaron en el poder estadístico deseado para detectar una diferencia del 10% en la tasa de metilación entre dos grupos en los cuales el tamaño de la muestra de los individuos para cada grupo era 18.

20 **[0210]** La significación estadística se determinó por regresión logística en el % de metilación por DMR (usando los recuentos reales) con los grupos definidos como neoplasia colorrectal, colon normal y leucocitos normales. Para tener en cuenta las diferentes profundidades de lectura entre sujetos individuales, se utilizó un modelo de regresión logística sobredispersado, donde el parámetro de dispersión se estimó utilizando la estadística Chi-cuadrado de Pearson de los residuos del modelo ajustado. Para evaluar la metilación específica del filamento, las regiones directa e inversa se analizaron por separado. Los DMR se clasificaron de acuerdo con su nivel de significación y se consideraron como una región marcadora viable si la tasa de metilación en los controles era $\leq 2,5\%$ pero $\geq 10\%$ en los casos. Cada DMR significativa se consideró como un marcador candidato.

30 **[0211]** Para el estudio de validación interna, el resultado primario fue el área bajo la curva de características operativas del receptor (AUC) para cada marcador. Esto se calculó utilizando la regresión logística (JMP versión 9.0.1, SAS Institute, Cary NC) para modelar la fuerza del número de copias medianas corregidas de cada marcador con neoplasia colorrectal en comparación con el colon normal y los leucocitos normales. Los marcadores con los valores más altos de AUC y la relación más amplia de número de copias de marcadores medianos entre casos y controles fueron seleccionados para el estudio de validación externa. El resultado primario para el experimento de validación externa fue el AUC para cada marcador trazado frente a la intensidad de la señal de cada marcador, medido por el registro de la relación de % de metilación mediano corregido en los casos en comparación con los controles. Con dieciocho casos, hay $>80\%$ de potencia para detectar un área bajo la curva de 0,85 o superior a partir de la hipótesis nula de 0,5 a un nivel de significancia de dos lados 0,05.

40 **[0212]** Descubrimiento del marcador RRBS Extractos de ADN coincidentes, cegados, asignados aleatoriamente de 18 cánceres colorrectales, 18 adenomas avanzados (>1 cm), 18 pólipos de sésiles serrados, 18 tejidos epiteliales de colon normales y 18 controles leucocitarios derivados de la capa leucocitaria normal fueron secuenciados por RRBS. La mediana de edad fue de 65 años (rango intercuartil 60-70), y el 51% eran mujeres. Se generaron 4-5 millones de CpG por muestra a partir de los datos de secuenciación. Después de seleccionar solo los sitios CpG donde se cumplieron los criterios de cobertura y varianza del grupo, se consideró un rango aproximado de 2-3 millones de sitios CpG para su posterior análisis. Las muestras de SSA, que se derivaron de bloques de tejido FFPE, tenían un número menor de CpG (200.000 - 1,2 millones) debido a la menor calidad inherente. Las DMR 1068 (CRC), 1200 (adenoma avanzado) y 268 (SSA) cumplieron los criterios de significación para la metilación diferencial. Estos se agruparon en 185, 244 y 109 regiones candidatas con suficientes firmas de metilación para el diseño del cebador MSP. La longitud de los DMR oscilaba entre 30 y más de 1000 bases. Las firmas de metilación contenían 6 a 69 CpG contiguas. El 84% de los DMR anotan en las regiones reguladoras 5' (promotoras) de genes, la mayoría de las cuales se asocian con islas CpG más grandes. En el caso de las DMR de CCR, aproximadamente el 15% tiene asociaciones previas con neoplasia colorrectal. 50% han sido reportados en otros tipos de cáncer. Del 25% restante, aproximadamente la mitad se segrega en vías relevantes para el cáncer (factores de transcripción, señalización, reguladores del ciclo celular, transportadores de membrana, etc.) Varios sitios anotados contienen múltiples DMR.

55 **[0213]** Validación interna Basándose en el número de CpG vecinas en cada firma de metilación del gen candidato, se diseñaron cebadores para los 50 candidatos principales de los 185 marcadores CRC. La clasificación se realizó por referencia a las métricas de regresión logística y la relación de número de copias caso/control. 35 de estas pasaron el control de calidad interno y 15 fueron rechazadas. Luego se usó MSP para analizar los 35 candidatos en muestras de ADN de cohortes parcialmente independientes, emparejadas, cegadas, que incluyen 36 lesiones de CCR, 36 adenomas avanzados y 36 muestras de epitelio colónico normal. Además, 31 muestras de ADN extraídas de pacientes con enfermedad inflamatoria intestinal (EII) (18 con CRC, 2 con alto grado de displasia, y 13 controles normales.) β -actina amplificado en todas las muestras. De los 35 ensayos de MSP, 18 marcadores candidatos tenían un AUC $>0,88$, relaciones de número de copias caso/control >50 y metilación de fondo $<1\%$ (ver Tabla 1A y 1B). Estos fueron seleccionados para su inclusión en una validación externa independiente.

65

Tabla 1A: Marcadores más discriminatorios en tejidos independientes por MSP, incluidas las coordenadas genómicas de DMR

Marcador	AUC	S/N Mediana	Metilación de fondo (%)	Cromosoma	Coordenadas genómicas DMR
FLI1	0,996	>1000	0	11	128563956-128564209
OPLAH	0,981	67	0,25	8	145106349-145106456
DTX1	0,974	>1000	0	12	113494586-113494957
MATK	0,972	64	0,63	19	3785828-3786371
SFMBT2 <i>región 2</i>	0,971	90	0,71	10	7452746-7452779
KCNK12	0,963	>1000	0	2	47797187-47797452
VAV3 <i>región 1</i>	0,954	135	0,27	1	108507074-108507498
SFMBT2 <i>región 3</i>	0,952	>1000	0	10	7452885-7452956
PPP2R5C	0,949	>1000	0,1	14	102247525-102247929
CHST2 <i>región 2</i>	0,947	>1000	0	3	142838645-142839023
PKIA	0,945	>1000	0,01	8	79428485-79428684
PDGFD	0,944	>1000	0	11	104034769-104034920
ELOVL2	0,935	70	0,79	6	11044395-11044834
CHST2 <i>región 1</i>	0,931	390	0,07	3	142838025-142838494
SFMBT2 <i>región 1</i>	0,931	>1000	0	10	7452029-7452452
QKI	0,921	50	0,61	6	163834534-163834925
VAV3 <i>región 2</i>	0,892	>1000	0	1	108507609-108507674
SLC8A3	0,883	596	0,1	14	70655516-70655712

Tabla 1B: La mayoría de los marcadores discriminatorios en tejidos independientes por MSP, incluidos los cebadores de MSP delantero y los cebadores de MSP inverso

Marcador	Reenviar MSP Primer	Reverse MSP Primer
FLI1	GGGAGTGAGGGTAGGGGTTT (SEQ ID NO: 1)	CTCGAACCCCTTCGAATTAACCCG (SEQ ID NO: 2)
OPLAH	TGCGTAGGTGATAGGGAGGGTTAC (SEQ ID NO: 3)	ACAAAACACATCCTATTAACGCGAA (SEQ ID NO: 4)
DTX1	GAGTCGCGGTTTGGTTTT (SEQ ID NO: 5)	GACGCGACGACCGGAAAAAC (SEQ ID NO: 6)
MATK	TGCACACCCCGAGGCGGTCCCG (SEQ ID NO: 7)	CGCCCCAAAATAAAAAAACGAA (SEQ ID NO: 8)
SFMBT2 región 2	GCGTTTAGTTGGTCGGAGA (SEQ ID NO: 10)	CCTAACCAACGCACTCAACC (SEQ ID NO: 11)
KCNK12	CGTAGCGTGGCGTTTTAGCGC (SEQ ID NO: 12)	TCGAAAAACCCGACGAAACGAAAAACG (SEQ ID NO: 13)
VAV3 región 1	GCGTAAGGTCGAAATATTTGAGTCGA (SEQ ID NO: 14)	AAAATACTACCCACCAACCCGAA (SEQ ID NO: 15)
SFMBT2 región 3	GTCGTCGTTTCGAGAGGGTA (SEQ ID NO: 16)	CGAACAAAAACGAAACGAAACGAA (SEQ ID NO: 17)
PPP2R5C	TCGATTTTATTTTTGTTGTCGTTGTAGAT TCGC (SEQ ID NO: 18)	GAAAAAACTAAAAAACGACAAAAAACCC CGACG (SEQ ID NO: 19)
CHST2 región 2	GGAACGAGTAGTAGTCGGATAGTTCGTC (SEQ ID NO: 20)	CGCCCCGAAAAACGACCCCG (SEQ ID NO: 21)
PKIA	CCCGCCGAA TACTCGATCAACTCG (SEQ ID NO: 23)	
PDGFD	CCGAACGCGTATAAATACCGCACTT (SEQ ID NO: 25)	
ELOVL2	CGGTTTTATTATTATGATTCGTAGCGG (SEQ ID NO: 26)	CGACTACCCCTAAACAACGCATCGC (SEQ ID NO: 27)
CHST2 región 1	CGAGTTCGGTAGTTGTACGTAGA (SEQ ID NO: 28)	CGAAATACGAAACGCGAAATCTAAACT (SEQ ID NO: 29)
SFMBT2 región 1	GCGACGTAGTCGTCGTTGT (SEQ ID NO: 30)	CCAACGCGAAAAAACGCG (SEQ ID NO: 31)
QKI	GAGGCGGACGTCGCGGTAC (SEQ ID NO: 32)	CGCCACGACGCGAAATCTTAACCTAGC (SEQ ID NO: 33)
VAV3 región 2	GGATCGAGGGAGTAGGAGTCGC (SEQ ID NO: 34)	CGAAAAACCGAACCTAACGCGGACG (SEQ ID NO: 35)
SLC8A3	AGTTTTTTCGCGCGTTTTTTTTG (SEQ ID NO: 36)	GCCGAACTCCGCCCTTACACG (SEQ ID NO: 37)

[0214] También se realizaron análisis logísticos en las muestras de casos/controles de EII. De los 35 marcadores probados, se seleccionaron 3 marcadores (PDGFD, CHST2 7889, SFMBT2 896) para compararlos con los marcadores de metilación ColoGuard (Exact Sciences) BMP3 y NDRG4 existentes en una validación externa basada en heces.

5 **[0215]** Validación externa El ADN emparejado, cegado, asignado aleatoriamente de 40 CRC, 24 adenomas avanzados y 40 muestras epiteliales de colon normales se analizaron mediante QuART para 18 candidatos principales. La mediana de edad de este subconjunto fue de 60 (rango intercuartil 52-67). El equilibrio de género fue de 50:50. Lo mismo para porcentajes de lesión distal:proximal en los casos. β -actina amplificada en todas las muestras. Los 18 marcadores demostraron un rendimiento en línea con los resultados de la validación interna. En el análisis CRC, 9 marcadores exhibieron un rendimiento superior en términos de características AUC y relaciones de metilación entre casos y controles. Como se muestra en la tabla a continuación (Tabla 1C), estos 9 mostraron una excelente asociación con el cáncer colorrectal.

Tabla 1C: Marcadores más discriminatorios que muestran asociación con CCR

Marcador	ROC AUC	Des. Est.	Valor Z	Pr (>z)	Inferior. 95	Superior. 95	% metilación media - Caso/control
vav3_877 (VAV3 región 2)	0,8718	0,035 4	10,5	0	0,8024	0,941 2	5046
chst2_7889 (CHST2 región 1)	0,8462	0,037 4	9,25	0	0,7728	0,919 5	10127
sfmbt2_896 (sfmbt2 región 2)	0,9526	0,025 1	18,05	0	0,9034	1,001 7	139
sfmbt2_895 (sfmbt2 región 1)	0,9439	0,026	16,72	0	0,8919	0,996	580
pdgfd	0,8814	0,035 1	10,85	0	0,8125	0,950 3	1351
dtx1	0,9606	0,022 2	20,77	0	0,9171	1,004	883
fli1	0,9744	0,017 9	26,51	0	0,9393	1,009 4	1131
sfmbt2_897 (sfmbt2 región 3)	0,9311	0,029 3	14,73	0	0,8737	0,988 5	323
chst2_7890 (región 2 de chst2)	0,9231	0,029 3	14,46	0	0,8657	0,980 4	6970

40 **[0216]** Para el estudio de validación de heces de EII, se eligieron tres marcadores - (PDGFD, CHST2 7889, SFMBT2 896), y se corrieron en comparación con BMP3 y NDRG4.

45 **[0217]** Se descubrieron marcadores de ADN metilados y se pusieron a prueba para medir la detección de CRN asociado a IBD (IBDCRN: displasia de bajo grado [LGD], displasia de alto grado [HGD] y cáncer colorrectal [CRC]).

50 **[0218]** Los marcadores se identificaron y probaron en 3 pasos discretos: descubrimiento; validación biológica; y pilotaje clínico. Primero, un experimento de descubrimiento identificó marcadores por secuenciación de bisulfito de representación reducida (RRBS) en el ADN extraído de muestras de tejido congeladas archivadas de CRC esporádico y adenomas. En segundo lugar, los marcadores candidatos se validaron mediante un ensayo de PCR específico de metilación (MSP) en ADN extraído de tejidos de archivo de pacientes con EII-CRC y controles de EII sin CRN. En tercer lugar, las heces de archivo de casos independientes de IBD-CRN y controles de IBD se analizaron mediante una diana cuantitativa específica de alelos en tiempo real y amplificación de señal (QuARTS). Se excluyeron los pacientes sin biopsias de vigilancia o con trasplante previo de órganos sólidos. La regresión logística midió la sensibilidad y la especificidad. La influencia de la variable clínica se probó mediante la suma de rangos de Chi-cuadrado y Wilcoxon para datos categóricos y continuos, respectivamente.

60 **[0219]** 18 CRC esporádicos, 18 adenomas avanzados y 18 muestras de colon normales fueron secuenciadas por RRBS; los mejores 20 candidatos fueron probados por MSP en muestras independientes, incluyendo 18 IBD-CRC y 13 controles IBD. Se seleccionaron tres marcadores (PDGFD, CHST2, SFMBT2) para compararlos con BMP3 y NDRG4; todos fueron analizados por QuARTS en muestras de heces de 33 casos de IBD-CRN (8 CRC, 8 HGD, 8 LGD \geq 1cm, 10 LGD <1cm) y 50 controles de IBD. Se excluyeron cuatro controles por insuficiencia de β -actina. La mediana de la duración de la enfermedad de la EII fue de 23 años (rango intercuartil [RIQ] 9-35) en los casos y 13 (8-20) años en controles ($p = 0,0009$). No se observaron otras diferencias significativas al comparar la edad, el sexo, la gravedad de la inflamación, la extensión de la EII o la colangitis esclerosante primaria comórbida. Los niveles de PDGFD, CHST2, SFMBT2 fueron moderadamente influenciados por la duración de la enfermedad ($p = 0,04, 0,02,$

0,04), pero *BMP3* y *NDRG4* no. Otras variables no fueron significativas. Se informan tasas de detección al 90% de especificidad (Tabla 2A, 2B y 2C).

5 **[0220]** Para muestras de CRC (n = 1) y HGD (n = 2) que fueron negativas, se revisaron los bloques H&E. Una muestra de DAG fue reclasificada como cambio reactivo. El ADN se extrajo de los tejidos restantes de HGD y CRC y se analizó mediante QuARTS para cada marcador anterior. Estos fueron muy positivos.

10 **[0221]** Ensayos de QuARTS se repitieron en las correspondientes muestras de heces con un cambio mínimo en % de metilación; sin embargo, los números de copias sin procesar aumentaron para cada marcador, de modo que *BMP3*, *NDRG4*, *PDGFG*, *CHST27889* pudieron detectar 8/8 CRC y 6/7 HGD (14/15 CRC + HGD, sensibilidad del 93%) en un rango de especificidad de 89-93%.

15 **Tabla 2A: Detección de neoplasias colorrectales asociadas a EII por ADN de heces metiladas con una especificidad del 90%**

Marcador metilado	CRC	CRC + HGD (n = 16)	HGD	LGD ≥ 1cm	LGD <1 cm
	(n = 8)		(n = 8)	(n = 8)	(n = 10)
BMP3	7 (88%)	13 (81%)	6 (75%)	6 (75%)	6 (60%)
NDRG4	7 (88%)	13 (81%)	6 (75%)	5 (63%)	4 (40%)
PDGFG	7 (88%)	13 (81%)	6 (75%)	4 (50%)	4 (40%)
CHST2	7 (88%)	13 (81%)	6 (75%)	5 (63%)	4 (40%)
SFMBT2	7 (88%)	13 (81%)	6 (75%)	5 (63%)	3 (30%)

25 **Tabla 2B: Detección de neoplasias colorrectales asociadas a EII por ADN de heces metiladas con una especificidad del 90%, incluidas las coordenadas genómicas de DMR**

Marcador metilado	Cromosoma	Coordenadas genómicas DMR
BMP3	4	81952348-81952402
NDRG4	16	58497395-58497451
PDGFG	11	104034769-104034920
CHST2	3	142838025-142838494
SFMBT2	10	7452029-7452452

40

45

50

55

60

65

Tabla 2C: Detección de neoplasias colorrectales asociadas a la EII por ADN de heces metiladas al 90% de especificidad, incluidos cebadores de MSP directa, secuencia de sonda y cebadores de MSP inversos

Marcador metilado	Adelante Primer	Secuencia de la sonda	Imprimación inversa
BMP3	GTTTAATTTTCGGTTTCGTCGTC (SEQ ID NO: 38)	CGCCGAG GCGGTTTT TTGCG (SEQ ID NO: 39)	CGCTACGAAACTCCGA (SEQ ID NO: 40)
NDRG4	CGGTTTTTCGTTCTTTTTCG (SEQ ID NO: 41)	CCACGGA CGGTTCGT TTATCG (SEQ ID NO: 42)	CCGCCCTTCTACGCGACTA (SEQ ID NO: 43)
PDGFG	GCGAATAAATAAACGTTAATTG TTGTTGTTTC (SEQ ID NO: 44)	CCACGGA CGCGCACT TCCTTA (SEQ ID NO: 45)	
CHST2	CGAGTTCGGTAGTTGTACGTAG A (SEQ ID NO: 47)	CGCCGAG GTCGTCGA TACCG (SEQ ID NO: 48)	
SFMBT2	GCGACGTAGTCGTCGTTGT (SEQ ID NO: 50)	CCACGGA CGGAAAAC GCGAAA (SEQ ID NO: 51)	CCAACGCGAAAAAACGCG (SEQ ID NO: 52)

Ejemplo 2.

[0222] Este ejemplo describe la identificación de marcadores de ADN metilados para la detección de cáncer colorrectal y precáncer.

[0223] Aquí se divulga un conjunto de 185 marcadores de metilación del ADN para el cáncer colorrectal (Tablas 3A y 3B), 244 para adenomas grandes (≥ 1 cm) (Tablas 4A y 4B), y 111 para pólipos de sésiles serrados (SSP) (Tablas 5A y 5B): todos identificados a partir de los datos generados por el enriquecimiento de la isla CpG junto con la secuenciación masivamente paralela de un conjunto de muestras de tejido de control de casos. Los adenomas y los SSP son las lesiones precancerosas críticas para el CRC, y su detección en una aplicación de detección es importante. Los controles incluyeron epitelios colónicos normales y ADN derivado de glóbulos blancos normales. La técnica utilizó la secuenciación de bisulfito de representación reducida (RRBS). Los análisis terciarios y cuaternarios son únicos e integrales al proceso de selección de marcadores. El paso terciario consiste en analizar los datos en términos de límites de cobertura, excluyendo todos los sitios no informativos, contrastando el % de metilación entre subgrupos de diagnóstico mediante regresión logística, creando "islas" CpG in silico basadas en agrupaciones definidas de sitios de metilación contiguos, y análisis de características operativas receptoras. El paso cuaternario filtra los datos de % de metilación (tanto CpG individuales como los grupos in silico) para seleccionar marcadores que maximicen las relaciones señal/ruido, minimicen la metilación de fondo, expliquen la heterogeneidad del tumor y enfatizen el rendimiento de ROC. Además, este enfoque analítico asegura la identificación de CpG de punto de acceso en una longitud de ADN definida para un desarrollo fácil y un diseño óptimo de ensayos de marcadores posteriores: PCR específica de metilación, secuenciación profunda de fragmentos pequeños, etc.

[0224] Las muestras de CRC, adenoma y control produjeron aproximadamente 2-3 millones de CpG de alta calidad, que después del análisis y el filtrado dieron como resultado aproximadamente 1068 (CRC) y 1200 (adenoma) sitios altamente discriminados. Estos se agruparon en 185 regiones localizadas de metilación diferencial para CRC y 244 para adenoma, algunas extendiéndose para 30-40 bases y otras más de una kilobase. De una revisión superficial de la literatura, menos del 15% de las DMR parecen tener una asociación previa de cáncer colorrectal. Sin embargo, casi el 50% tiene alguna asociación previa con el cáncer, pero no tiene un carácter específicamente epigenético. El resto tiene asociaciones de cáncer muy débiles o nulas, aunque muchas están involucradas en vías funcionales que pueden ser relevantes para la tumorigénesis. 30 DMR (CRC) y 42 (adenoma) no tenían anotaciones ni referencias en ninguna parte de la literatura. Estos fueron nombrados MAX seguido de ubicación cromosómica y coordenadas. Todos los DMR tenían AUC de 0,85 o más (algunos demostraron una discriminación perfecta de los casos de los controles con un AUC de 1,0). Además de alcanzar el umbral de AUC de 0,85, los marcadores de cáncer de colon y adenoma que se identificaron tuvieron que exhibir una densidad de metilación >50 veces mayor en el tumor que en la mucosa gástrica o colónica normal y $<1,0$ % de metilación en la mucosa colónica normal. Con base en la experiencia previa con el descubrimiento y validación de marcadores de cáncer de páncreas, estas características de marcadores en el tejido predicen una alta discriminación en medios distantes como las heces o la sangre.

[0225] Las muestras de SSP se derivaron de bloques FFPE y fueron de menor calidad. Para estos, solo la mitad tuvo un número adecuado de lecturas (>100.000). Como tal, las limitaciones de filtrado se redujeron ligeramente en comparación con las utilizadas con las muestras congeladas. Después de la clasificación, hubo 268 sitios discriminados que se agruparon en 111 regiones localizadas.

[0226] Algunos de los marcadores descritos se comparten entre grupos de casos por pares, y un número menor entre los tres. Otros son exclusivos de un grupo específico. La mayoría son parte de islas CpG definidas (p. ej., contenido de GC superior al 55% y una relación CpG observada a esperada del 65% o más), y cuando hay una anotación (asociación con un gen conocido), la ubicación es generalmente en la región promotora.

Tabla 3A - DMR de cáncer colorrectal

Cromosoma	Coordenadas cromosómicas	Anotación de genes
2	207308703-207308890	ADAM23
7	45613877-45613977	ADCY1
22	24820148-24820373	ADORA2A
7	44143993-44144413	AEBP1
2	100721643-100721967	AFF3
1	49242089-49242514	AGBL4
6	151561236-151561473	AKAP12
7	134142981-134143723	AKR1B1
8	41754327-41754726	ANK1
17	27940469-27940612	ANKRD13B
5	10565042-10565191	ANKRD33B

ES 2 812 753 T3

(Continuación)

	Cromosoma	Coordenadas cromosómicas	Anotación de genes
	5	10564655-10564807	ANKRD33B
5	11	110582796-110583345	ARHGAP20
	11	110582170-110582657	ARHGAP20
	11	110581912-110582039	ARHGAP20
	12	103351885-103352327	ASCL1
10	13	25946118-25946206	ATP8A2
	8	104152806-104153145	BAALC
	9	96715112-96715603	BARX1
	8	65493937-65494105	BHLHE22
	6	7727566-7728088	BMP6
15	6	105584685-105585220	BVES
	17	43339242-43339498	C17orf46
	1	1475560-1475650	C1orf70
	1	1476065-1476127	C1orf70
	10	16562866-16563332	C1QL3
20	10	16563667-16563892	C1QL3
	10	16562465-16562672	C1QL3
	9	132382813-132382909	C9orf50
	18	70211543-70211719	CBLN2
25	3	128720801-128720885	CCDC48
	3	128719995-128720631	CCDC48
	2	101033758-101034005	CHST10
	12	104850745-104851001	CHST11
	3	142838025-142838494	CHST2
30	3	142838645-142839023	CHST2
	3	142839223-142839576	CHST2
	5	178016833-178017456	COL23A1
	7	30721941-30722028	CRHR2
35	9	124461296-124461420	DAB2IP
	12	64062131-64062443	DPY19L2
	12	113494586-113494957	DTX1
	10	64575060-64575283	EGR2
	2	31456804-31457263	EHD3
40	7	37487755-37488565	ELMO1
	7	37487539-37487623	ELMO1
	6	11044395-11044834	ELOVL2
	6	80656845-80657306	ELOVL4
	6	152129389-152129636	ESR1
45	6	133562485-133562878	EYA4
	15	48938056-48938252	FBN1
	11	128563956-128564209	FLI1
	11	128562780-128563522	FLI1
50	6	159590083-159590220	FNDC1
	9	101471421-101471519	GABBR2
	2	31360809-31360992	GALNT14
	2	31360542-31360640	GALNT14
	11	134146132-134146380	GLB1L3
55	7	42276418-42277414	GLI3
	12	52400569-52400726	AGARRAR
	12	52400919-52401166	AGARRAR
	16	28074472-28074761	GSG1L
	17	1959348-1959370	HIC1
60	7	50343838-50344453	IKZF1
	2	182321830-182321983	ITGA4
	1	226925082-226925651	ITPKB
	4	6201350-6201560	JAKMIP1
	2	47797187-47797452	KCNK12
65	2	149633039-149633137	KIF5C

ES 2 812 753 T3

(Continuación)

	Cromosoma	Coordenadas cromosómicas	Anotación de genes
	7	149411729-149411847	KRBA1
5	19	8274584-8274671	LASS4
	2	30455594-30455705	LBH
	5	38556357-38556743	LIFR
	19	2290273-2290393	LINGO3
10	19	2290645-2290738	LINGO3
	19	42905798-42906349	LIPE
	7	140772542-140772873	LOC100131199
	2	100938402-100939005	LONRF2
	7	127671918-127672318	LRRC4
15	19	41119795-41119907	LTBP4
	6	6546375-6546598	LY86-AS1
	19	3785828-3786371	MATK
	10	22541502-22541671	MAX.chr10,22541502-22541671
	10	22541884-22542001	MAX.chr10,22541884-22542001
20	10	22765155-22765223	MAX.chr10,22765155-22765223
	11	123301058-123301255	MAX.chr11,123301058-123301255
	11	123301366-123301506	MAX.chr11,123301366-123301506
	11	123301853-123301941	MAX.chr11,123301853-123301941
	11	57250528-57250611	MAX.chr11,57250528-57250611
25	12	8171360-8171769	MAX.chr12,8171360-8171769
	14	100437680-100437767	MAX.chr14,100437680-100437767
	19	22034447-22034696	MAX.chr19,22034447-22034696
	19	22034799-22034887	MAX.chr19,22034799-22034887
30	19	42444999-42445053	MAX.chr19,42444999-42445053
	2	144694517-144695025	MAX.chr2,144694517-144695025
	20	44936022-44936246	MAX.chr20,44936022-44936246
	3	13324501-13324623	MAX.chr3,13324501-13324623
	3	13324760-13324864	MAX.chr3,13324760-13324864
35	3	44039952-44040054	MAX.chr3,44039952-44040054
	7	142494755-142494915	MAX.chr7,142494755-142494915
	8	30769438-30769680	MAX.chr8,30769438-30769680
	9	99983730-99984118	MAX.chr9,99983730-99984118
40	6	41606074-41606126	MDF1
	3	150804938-150804971	MED12L
	5	88185490-88185589	MEF2C
	22	39853199-39853295	MGAT3
45	2	220416703-220417434	MIR3132
	6	132722283-132722484	MOXD1
	8	72755813-72756349	MSC
	16	58497251-58497370	NDRG4
	19	3361105-3361330	NFIC
	17	47573986-47574084	NGFR
50	7	108095348-108095805	NRCAM
	8	32406662-32406901	NRG1
	17	8925482-8925838	NTN1
	15	53082447-53083044	ONECUT1
55	8	145106742-145106921	OPLAH
	8	145106349-145106456	OPLAH
	5	76506245-76506578	PDE8B
	11	104034769-104034920	PDGFD
	22	45405722-45405819	PHF21B
60	8	79428485-79428684	PKIA
	1	150122783-150123157	PLEKHO1
	1	38510915-38511213	POU3F1
	17	56833684-56833978	PPM1E
	20	37434246-37434800	PPP1R16B
65	14	102248062-102248216	PPP2R5C

ES 2 812 753 T3

(Continuación)

	Cromosoma	Coordenadas cromosómicas	Anotación de genes
	14	102247525-102247929	PPP2R5C
5	16	23847825-23848168	PRKCB
	19	47778181-47778372	PRR24
	6	163834534-163834925	QKI
	18	9708397-9709392	RAB31
	20	4803201-4803703	RASSF2
10	2	161263880-161264733	RBMS1
	6	127440413-127441057	RSPO3
	17	26698693-26699117	SARM1
	8	97505964-97506676	SDC2
15	17	75369224-75369327	Septin9
	17	75368800-75369056	Septin9
	10	7452746-7452779	SFMBT2
	10	7452885-7452956	SFMBT2
	10	7452029-7452452	SFMBT2
20	10	7450242-7450831	SFMBT2
	10	7451097-7451185	SFMBT2
	14	70655516-70655712	SLC8A3
	5	101632152-101632237	SLCO4C1
25	7	128829103-128829184	SMO
	11	65601167-65601514	SNX32
	13	95363646-95363959	SOX21
	12	24715703-24715776	SOX5
	12	24715012-24715416	SOX5
30	12	24716178-24716294	SOX5
	17	70114081-70114176	SOX9
	7	75896637-75896925	SRRM3
	6	166581771-166582044	T
	12	65218900-65218994	TBC1D30
35	12	65218335-65218778	TBC1D30
	8	67874670-67875083	TCF24
	18	53255390-53255565	TCF4
	5	1294873-1295322	TERT
40	21	32930226-32930576	TIAM1
	21	32716063-32716545	TIAM1
	2	74741941-74742264	TLX2
	15	83776196-83776373	TM6SF1
	2	135476019-135476390	TMEM163
45	7	19156788-19156858	TWIST1
	1	108507609-108507674	VAV3
	1	108507074-108507498	VAV3
	5	82768837-82769031	VCAN
	2	175547056-175547390	WIPF1
50	8	10873760-10874271	XKR6
	8	10872819-10873619	XKR6
	10	31609049-31609227	ZEB1
	2	145274517-145274600	ZEB2
	2	145274704-145275062	ZEB2
55	19	58951402-58951530	ZNF132
	19	54024023-54024436	ZNF331
	19	53661526-53662618	ZNF347
	19	22018452-22018639	ZNF43
60	16	88496963-88497197	ZNF469
	19	37407197-37407365	ZNF568
	19	12267378-12267677	ZNF625
	19	12203466-12203641	ZNF788
65	2	185463105-185463763	ZNF804A

(Continuación)

Cromosoma	Coordenadas cromosómicas	Anotación de genes
19	53970869-53971374	ZNF813

Tabla 3B: DMR de cáncer colorrectal clasificadas por área bajo la curva ROC

Anotación de genes	Cromosoma	Coordenada cromosómica	Área bajo la curva ROC
SFMBT2	10	7452885-7452956	1.0000
LIFR	5	38556357-38556743	0,9969
OPLAH	8	145106742-145106921	0,9969
CRHR2	7	30721941-30722028	0,9966
AGBL4	1	49242089-49242514	0,9954
ZEB2	2	145274704-145275062	0,9938
ZNF788	19	12203466-12203641	0,9936
SFMBT2	10	7452746-7452779	0,9926
ESR1	6	152129389-152129636	0,9918
ANKRD33B	5	10565042-10565191	0,9907
CHST2	3	142838645-142839023	0,9907
FNDC1	6	159590083-159590220	0,9902
ZNF469	16	88496963-88497197	0,9899
AKR1B1	7	134142981-134143723	0,9892
OPLAH	8	145106349-145106456	0,9886
MSC	8	72755813-72756349	0,9876
KCNK12	2	47797187-47797452	0,9861
MAX.chr10,22541884-22542001	10	22541884-22542001	0,9861
ONECUT1	15	53082447-53083044	0,9861
RASSF2	20	4803201-4803703	0,9861
BHLHE22	8	65493937-65494105	0,9841
ARHGAP20	11	110582170-110582657	0,9837
EYA4	6	133562485-133562878	0,9830
LINGO3	19	2290273-2290393	0,9830
MATK	19	3785828-3786371	0,9830
RSPO3	6	127440413-127441057	0,9830
MGAT3	22	39853199-39853295	0,9828
AGARRAR	12	52400919-52401166	0,9814
ZEB2	2	145274517-145274600	0,9806
GSG1L	16	28074472-28074761	0,9804
ZNF625	19	12267378-12267677	0,9799
NDRG4	16	58497251-58497370	0,9771
PPP2R5C	14	102247525-102247929	0,9771
FLI1	11	128562780-128563522	0,9739
ZEB1	10	31609049-31609227	0,9739
C1QL3	10	16562465-16562672	0,9737
C1orf70	1	1476065-1476127	0,9733
ANKRD13B	17	27940469-27940612	0,9721
DAB2IP	9	124461296-124461420	0,9721
GALNT14	2	31360542-31360640	0,9721
ATP8A2	13	25946118-25946206	0,9716
CCDC48	3	128720801-128720885	0,9707
LONRF2	2	100938402-100939005	0,9706
LRRC4	7	127671918-127672318	0,9706
PDGFD	11	104034769-104034920	0,9706
SFMBT2	10	7452029-7452452	0,9706
SFMBT2	10	7450242-7450831	0,9706
CHST2	3	142839223-142839576	0,9690
MAX.chr7,142494755-142494915	7	142494755-142494915	0,9690

(Continuación)

	Anotación de genes	Cromosoma	Coordenada cromosómica	Área bajo la curva ROC
5	MAX.chr11,123301058-123301255	11	123301058-123301255	0,9673
	SOX9	17	70114081-70114176	0,9665
	MAX.chr20,44936022-44936246	20	44936022-44936246	0,9659
10	TWIST1	7	19156788-19156858	0,9652
	ELMO1	7	37487539-37487623	0,9644
	NFIC	19	3361105-3361330	0,9644
	PLEKHO1	1	150122783-150123157	0,9644
15	POU3F1	1	38510915-38511213	0,9644
	NGFR	17	47573986-47574084	0,9633
	LINGO3	19	2290645-2290738	0,9624
	AEBP1	7	44143993-44144413	0,9613
20	PPP1R16B	20	37434246-37434800	0,9598
	SFMBT2	10	7451097-7451185	0,9585
	SMO	7	128829103-128829184	0,9583
	FLI1	11	128563956-128564209	0,9575
	DTX1	12	113494586-113494957	0,9551
	TIAM1	21	32930226-32930576	0,9551
25	GABBR2	9	101471421-101471519	0,9542
	PRKCB	16	23847825-23848168	0,9539
	RAB31	18	9708397-9709392	0,9536
	VAV3	1	108507074-108507498	0,9536
30	LASS4	19	8274584-8274671	0,9533
	ANK1	8	41754327-41754726	0,9526
	ANKRD33B	5	10564655-10564807	0,9520
	SARM1	17	26698693-26699117	0,9520
	TM6SF1	15	83776196-83776373	0,9516
35	ZNF568	19	37407197-37407365	0,9505
	C1orf70	1	1475560-1475650	0,9497
	ITGA4	2	182321830-182321983	0,9495
	GLB1L3	11	134146132-134146380	0,9493
40	MAX.chr12,8171360-8171769	12	8171360-8171769	0,9489
	MAX.chr14,100437680-100437767	14	100437680-100437767	0,9481
	ZNF132	19	58951402-58951530	0,9479
45	BAALC	8	104152806-104153145	0,9474
	CHST10	2	101033758-101034005	0,9458
	IKZF1	7	50343838-50344453	0,9443
	MAX.chr2,144694517-144695025	2	144694517-144695025	0,9443
50	NRG1	8	32406662-32406901	0,9443
	AFF3	2	100721643-100721967	0,9412
	FBN1	15	48938056-48938252	0,9412
	C1QL3	10	16562866-16563332	0,9397
	Septin9	17	75368800-75369056	0,9396
55	MED12L	3	150804938-150804971	0,9387
	MDFI	6	41606074-41606126	0,9381
	MAX.chr11,123301366-123301506	11	123301366-123301506	0,9381
60	C9orf50	9	132382813-132382909	0,9375
	ITPKB	1	226925082-226925651	0,9365
	TIAM1	21	32716063-32716545	0,9350
	LTBP4	19	41119795-41119907	0,9342
	LOC100131199	7	140772542-140772873	0,9334
65	TCF24	8	67874670-67875083	0,9334

(Continuación)

	Anotación de genes	Cromosoma	Coordenada cromosómica	Área bajo la curva ROC
5	MAX.chr11,123301853-123301941	11	123301853-123301941	0,9321
	SDC2	8	97505964-97506676	0,9319
	TERT	5	1294873-1295322	0,9319
	ELOVL2	6	11044395-11044834	0,9303
10	GALNT14	2	31360809-31360992	0,9298
	Septin9	17	75369224-75369327	0,9289
	GLI3	7	42276418-42277414	0,9288
	SOX5	12	24715703-24715776	0,9286
15	BMP6	6	7727566-7728088	0,9272
	ELMO1	7	37487755-37488565	0,9257
	HIC1	17	1959348-1959370	0,9246
	VAV3	1	108507609-108507674	0,9229
	ELOVL4	6	80656845-80657306	0,9203
20	CCDC48	3	12871995-128720631	0,9195
	COL23A1	5	178016833-178017456	0,9195
	PPM1E	17	56833684-56833978	0,9195
	PKIA	8	79428485-79428684	0,9186
	ASCL1	12	103351885-103352327	0,9180
25	ZNF347	19	53661526-53662618	0,9173
	MAX.chr19,22034447-22034696	19	22034447-22034696	0,9164
	NRCAM	7	108095348-108095805	0,9164
30	C17orf46	17	43339242-43339498	0,9149
	ARHGAP20	11	110582796-110583345	0,9134
	SOX5	12	24716178-24716294	0,9134
	TLX2	2	74741941-74742264	0,9133
	ZNF43	19	22018452-22018639	0,9131
35	BVES	6	105584685-105585220	0,9118
	CBLN2	18	70211543-70211719	0,9102
	MAX.chr19,42444999-42445053	19	42444999-42445053	0,9080
40	BARX1	9	96715112-96715603	0,9071
	AGARRAR	12	52400569-52400726	0,9071
	PRR24	19	47778181-47778372	0,9040
	MAX.chr3,13324501-13324623	3	13324501-13324623	0,9034
45	MAX.chr11,57250528-57250611	11	57250528-57250611	0,9031
	C1QL3	10	16563667-16563892	0,9025
	MAX.chr3,13324760-13324864	3	13324760-13324864	0,9003
50	KRBA1	7	149411729-149411847	0,9002
	ADORA2A	22	24820148-24820373	0,8957
	TCF4	18	53255390-53255565	0,8955
	TBC1D30	12	65218900-65218994	0,8948
	XKR6	8	10872819-10873619	0,8947
55	SLC8A3	14	70655516-70655712	0,8922
	NTN1	17	8925482-8925838	0,8916
	MAX.chr10,22765155-22765223	10	22765155-22765223	0,8916
60	MAX.chr10,22541502-22541671	10	22541502-22541671	0,8884
	MEF2C	5	88185490-88185589	0,8870
	TBC1D30	12	65218335-65218778	0,8870
65	MAX.chr9,99983730-99984118	9	99983730-99984118	0,8854

(Continuación)

	Anotación de genes	Cromosoma	Coordenada cromosómica	Área bajo la curva ROC
5	MAX.chr19,22034799-22034887	19	22034799-22034887	0,8839
	SOX21	13	95363646-95363959	0,8839
	ZNF804A	2	185463105-185463763	0,8839
10	SLCO4C1	5	101632152-101632237	0,8805
	ADAM23	2	207308703-207308890	0,8799
	TMEM163	2	135476019-135476390	0,8777
	PDE8B	5	76506245-76506578	0,8769
	EHD3	2	31456804-31457263	0,8762
15	MAX.chr8,30769438-30769680	8	30769438-30769680	0,8762
	QKI	6	163834534-163834925	0,8762
	AKAP12	6	151561236-151561473	0,8754
	ADCY1	7	45613877-45613977	0,8731
20	LBH	2	30455594-30455705	0,8731
	SNX32	11	65601167-65601514	0,8731
	PPP2R5C	14	102248062-102248216	0,8721
	MAX.chr3,44039952-44040054	3	44039952-44040054	0,8715
25	RBMS1	2	161263880-161264733	0,8715
	ARHGAP20	11	110581912-110582039	0,8684
	LIPE	19	42905798-42906349	0,8684
	SRRM3	7	75896637-75896925	0,8684
30	CHST2	3	142838025-142838494	0,8669
	XKR6	8	10873760-10874271	0,8669
	WIPF1	2	175547056-175547390	0,8653
	VCAN	5	82768837-82769031	0,8628
	KIF5C	2	149633039-149633137	0,8612
35	MOXD1	6	132722283-132722484	0,8607
	MIR3132	2	220416703-220417434	0,8591
	T	6	166581771-166582044	0,8591
	DPY19L2	12	64062131-64062443	0,8584
	LY86-AS1	6	6546375-6546598	0,8560
40	ZNF331	19	54024023-54024436	0,8545
	PHF21B	22	45405722-45405819	0,8541
	EGR2	10	64575060-64575283	0,8529
	ZNF813	19	53970869-53971374	0,8522
45	CHST11	12	104850745-104851001	0,8514
	SOX5	12	24715012-24715416	0,8514
	JAKMIP1	4	6201350-6201560	0,8506

Tabla 4A: DMR de adenoma grande

	Cromosoma	Coordenadas cromosómicas	Anotación de genes
50	7	45613877-45613977	ADCY1
	2	70994754-70995045	AGREGAR2
	22	24820148-24820373	ADORA2A
55	2	100721643-100721967	AFF3
	1	49242089-49242514	AGBL4
	6	151561236-151561473	AKAP12
	6	151561598-151561873	AKAP12
60	7	134142981-134143723	AKR1B1
	8	41754327-41754726	ANK1
	5	10564406-10564551	ANKRD33B
	5	10564655-10564807	ANKRD33B
	5	10565042-10565191	ANKRD33B
65	11	110582796-110583345	ARHGAP20

ES 2 812 753 T3

(Continuación)

	Cromosoma	Coordenadas cromosómicas	Anotación de genes
5	12	103351885-103352327	ASCL1
	8	104152806-104153145	BAALC
	8	65494269-65494355	BHLHE22
	6	7727566-7728088	BMP6
	6	105584685-105585220	BVES
10	12	21680721-21680828	C12orf39
	12	48577334-48577557	C12orf68
	16	4588091-4588817	C16orf5
	1	1475560-1475650	C1orf70
	1	1476065-1476127	C1orf70
15	10	16562866-16563332	C1QL3
	10	16563667-16563892	C1QL3
	10	16562465-16562672	C1QL3
	20	3388089-3388291	C20orf194
20	6	74019826-74019955	C6orf147
	9	132382813-132382909	C9orf50
	7	44364925-44365359	CAMK2B
	5	110559508-110560719	CAMK4
	3	12838197-12838303	CAND2
25	18	70211543-70211719	CBLN2
	6	74405903-74406086	CD109
	12	133464655-133464819	CHFR
	11	45686306-45686534	CHST1
	2	101033758-101034005	CHST10
30	12	104851372-104851465	CHST11
	12	104850745-104851001	CHST11
	10	125852012-125852098	CHST15
	10	125852559-125852792	CHST15
35	10	125852905-125853007	CHST15
	10	125851544-125851700	CHST15
	3	142838025-142838494	CHST2
	3	142838645-142839023	CHST2
	3	142839223-142839576	CHST2
40	3	139654045-139654299	CLSTN2
	5	178016833-178017456	COL23A1
	9	124461296-124461420	DAB2IP
	3	186079767-186080092	DGKG
	1	65731412-65731782	DNAJC6
45	2	225906664-225906922	DOCK10
	12	64062131-64062443	DPY19L2
	12	113494586-113494957	DTX1
	2	233352345-233352431	ECEL1
50	10	64575060-64575283	EGR2
	2	31456804-31457263	EHD3
	7	37487755-37488565	ELMO1
	6	11044395-11044834	ELOVL2
	6	80656845-80657306	ELOVL4
55	6	152129389-152129636	ESR1
	6	133562229-133562380	EYA4
	6	133562485-133562878	EYA4
	7	23053043-23053438	FAM126A
	1	53098973-53099237	FAM159A
60	1	206137408-206137473	FAM72A
	1	120839339-120839381	FAM72B
	1	120838272-120838775	FAM72B
	1	120836675-120836768	FAM72B
65	1	27960931-27961018	FGR
	11	128563956-128564209	FLI1

ES 2 812 753 T3

(Continuación)

	Cromosoma	Coordenadas cromosómicas	Anotación de genes
5	11	128562780-128563522	FLI1
	14	62584068-62584109	FLJ43390
	13	28674199-28674862	FLT3
	9	101471421-101471519	GABBR2
	2	31360809-31360992	GALNT14
10	17	10102237-10102576	GAS7
	19	19006296-19006511	GDF1
	5	179780627-179781188	GFPT2
	5	137610023-137610333	GFRA3
	11	134146132-134146380	GLB1L3
15	7	42276418-42277414	GLI3
	1	54204505-54204712	GLIS1
	12	52400919-52401166	AGARRAR
	10	88125930-88126495	GRID1
	7	6570511-6570865	GRID2IP
20	16	28074472-28074761	GSG1L
	6	32632785-32632860	HLA-DQB1
	15	83621302-83621657	HOMER2
	7	50343838-50344453	IKZF1
25	2	182321830-182321983	ITGA4
	21	46351838-46352381	ITGB2
	1	226925082-226925651	ITPKB
	4	6202051-6202410	JAKMIP1
	4	6201350-6201560	JAKMIP1
30	2	47797187-47797452	KCNK12
	20	62103225-62103324	KCNQ2
	6	73972941-73973104	KHDC1
	2	149633039-149633137	KIF5C
35	7	149411729-149411847	KRBA1
	19	8274360-8274430	LASS4
	19	8274584-8274671	LASS4
	19	2290273-2290393	LINGO3
	19	42905798-42906349	LIPE
40	11	8284746-8284871	LMO1
	7	140773610-140773855	LOC100131199
	7	140772542-140772873	LOC100131199
	2	100938402-100939005	LONRF2
	7	127671918-127672318	LRRC4
45	19	41119795-41119907	LTBP4
	6	6546375-6546598	LY86-AS1
	11	63828346-63828436	MACROD1
	19	3785828-3786371	MATK
50	1	244012766-244012875	MAX.chr1,244012766-244012875
	1	244013190-244013393	MAX.chr1,244013190-244013393
	1	39269813-39270150	MAX.chr1,39269813-39270150
	10	22541502-22541671	MAX.chr10,22541502-22541671
	10	22541884-22542001	MAX.chr10,22541884-22542001
	10	22765155-22765223	MAX.chr10,22765155-22765223
55	11	120435350-120435981	MAX.chr11,120435350-120435981
	11	123301058-123301255	MAX.chr11,123301058-123301255
	11	123301366-123301506	MAX.chr11,123301366-123301506
	11	123301853-123301941	MAX.chr11,123301853-123301941
60	11	44749119-44749205	MAX.chr11,44749119-44749205
	11	8040551-8040677	MAX.chr11,8040551-8040677
	12	133484966-133485857	MAX.chr12,133484966-133485857
	14	100437680-100437767	MAX.chr14,100437680-100437767
	14	105400087-105400182	MAX.chr14,105400087-105400182
65	15	34806855-34807014	MAX.chr15,34806855-34807014

ES 2 812 753 T3

(Continuación)

	Cromosoma	Coordenadas cromosómicas	Anotación de genes
5	18	77558550-77558609	MAX.chr18,77558550-77558609
	19	20959229-20959691	MAX.chr19,20959229-20959691
	19	22034447-22034696	MAX.chr19,22034447-22034696
	19	22034799-22034887	MAX.chr19,22034799-22034887
	2	144694517-144695025	MAX.chr2,144694517-144695025
10	21	47063135-47064177	MAX.chr21,47063135-47064177
	3	115231555-115231576	MAX.chr3,115231555-115231576
	3	44039952-44040054	MAX.chr3,44039952-44040054
	8	30769438-30769680	MAX.chr8,30769438-30769680
	6	37664238-37664539	MDGA1
15	5	88185490-88185589	MEF2C
	22	39853199-39853295	MGAT3
	17	74864552-74864821	MGAT5B
	3	154797723-154797909	MME
20	6	132722283-132722484	MOXD1
	17	8533282-8534168	MYH10
	11	112832731-112832815	NCAM1
	16	58497979-58498250	NDRG4
	17	47573986-47574084	NGFR
25	8	41503949-41504137	NKX6-3
	5	142784971-142785160	NR3C1
	7	108095348-108095805	NRCAM
	17	8925482-8925838	NTN1
30	1	107683961-107684314	NTNG1
	1	107683064-107683372	NTNG1
	1	107684447-107684545	NTNG1
	3	13461109-13461191	NUP210
	11	79150971-79151076	ODZ4
35	15	53082447-53083044	ONECUT1
	8	145106742-145106921	OPLAH
	8	145106349-145106456	OPLAH
	3	142682282-142682813	PAQR9
	5	140855415-140856027	PCDHGA1
40	5	76506245-76506578	PDE8B
	11	104034769-104034920	PDGFD
	4	55099106-55099473	PDGFRA
	1	9711854-9711974	PIK3CD
	8	79428485-79428684	PKIA
45	1	150122783-150123157	PLEKHO1
	1	38510915-38511213	POU3F1
	17	56833684-56833978	PPM1E
	14	102248062-102248216	PPP2R5C
50	14	102247525-102247929	PPP2R5C
	5	122425730-122425886	PRDM6
	20	47444582-47444776	PREX1
	16	23847825-23848168	PRKCB
	16	23846951-23847056	PRKCB
55	8	30890580-30890912	PURG
	6	163834534-163834925	QKI
	6	163835376-163835472	QKI
	20	4803969-4804077	RASSF2
	20	4803201-4803703	RASSF2
60	18	56936593-56936656	RAX
	15	93631919-93632242	RGMA
	9	77111900-77112005	RORB
	6	127440413-127441057	RSPO3
65	17	26698693-26699117	SARM1
	8	97505964-97506676	SDC2

ES 2 812 753 T3

(Continuación)

	Cromosoma	Coordenadas cromosómicas	Anotación de genes
5	7	3339895-3340903	SDK1
	19	40005836-40005892	SELV
	17	75368800-75369056	Septin9
	22	26565137-26565417	SEZ6L
	10	7452746-7452779	SFMBT2
10	10	7452885-7452956	SFMBT2
	10	7452029-7452452	SFMBT2
	10	7450242-7450831	SFMBT2
	10	7451097-7451185	SFMBT2
15	6	118228394-118228979	SLC35F1
	14	70655516-70655712	SLC8A3
	14	70655268-70655368	SLC8A3
	11	65601167-65601514	SNX32
	13	95363646-95363959	SOX21
20	12	24715703-24715776	SOX5
	12	24715012-24715416	SOX5
	12	24716178-24716294	SOX5
	8	10587893-10588143	SOX7
	7	75896637-75896925	SRRM3
25	8	134583587-134583963	ST3GAL1
	12	65218900-65218994	TBC1D30
	17	45810562-45810819	TBX21
	8	67874670-67875083	TCF24
	18	53255390-53255565	TCF4
30	21	32931523-32931688	TIAM1
	21	32930226-32930576	TIAM1
	15	83776196-83776373	TM6SF1
	2	135476019-135476390	TMEM163
35	2	39893089-39893224	TMEM178
	2	12857915-12858230	TRIB2
	7	19156788-19156858	TWIST1
	1	213124472-213124778	VASH2
	1	108507609-108507674	VAV3
40	1	108507074-108507498	VAV3
	5	82768837-82769031	VCAN
	10	17271896-17271978	EMPUJE
	10	17270955-17271052	EMPUJE
45	13	27131683-27131757	WASF3
	2	175547056-175547390	WIPF1
	1	228195339-228195413	WNT3A
	8	10873760-10874271	XKR6
	8	10872819-10873619	XKR6
50	10	31608798-31608892	ZEB1
	10	31608394-31608690	ZEB1
	10	31609049-31609227	ZEB1
	19	58951402-58951530	ZNF132
	4	332064-332199	ZNF141
55	17	16472295-16472694	ZNF287
	19	54024023-54024436	ZNF331
	19	53661526-53662618	ZNF347
	19	22018746-22019004	ZNF43
	16	88496963-88497197	ZNF469
60	19	37064200-37064435	ZNF529
	19	37960066-37960505	ZNF569
	19	12267378-12267677	ZNF625
	19	20149796-20149923	ZNF682
65	2	185463105-185463763	ZNF804A

Tabla 4B: DMR de adenoma grande clasificados por área bajo la curva ROC

	Anotación de genes	Cromosoma	Coordenada cromosómica	Área bajo la curva ROC
5	AGREGAR2	2	70994754-70995045	1.0000
	AGBL4	1	49242089-49242514	1.0000
	AKAP12	6	151561598-151561873	1.0000
	ANKRD33B	5	10565042-10565191	1.0000
10	ASCL1	12	103351885-103352327	1.0000
	C1orf70	1	1475560-1475650	1.0000
	CHST11	12	104851372-104851465	1.0000
	CHST15	10	125851544-125851700	1.0000
	DTX1	12	113494586-113494957	1.0000
15	ECEL1	2	233352345-233352431	1.0000
	EYA4	6	133562485-133562878	1.0000
	FLI1	11	128562780-128563522	1.0000
	FLJ43390	14	62584068-62584109	1.0000
	FLT3	13	28674199-28674862	1.0000
20	AGARRAR	12	52400919-52401166	1.0000
	ITGA4	2	182321830-182321983	1.0000
	KCNQ2	20	62103225-62103324	1.0000
	LOC100131199	7	140772542-140772873	1.0000
25	LONRF2	2	100938402-100939005	1.0000
	MGAT3	22	39853199-39853295	1.0000
	OPLAH	8	145106742-145106921	1.0000
	OPLAH	8	145106349-145106456	1.0000
	PDE8B	5	76506245-76506578	1.0000
30	PDGFD	11	104034769-104034920	1.0000
	PKIA	8	79428485-79428684	1.0000
	POU3F1	1	38510915-38511213	1.0000
	QKI	6	163834534-163834925	1.0000
35	RASSF2	20	4803201-4803703	1.0000
	RSPO3	6	127440413-127441057	1.0000
	SDC2	8	97505964-97506676	1.0000
	SFMBT2	10	7452746-7452779	1.0000
	SFMBT2	10	7452029-7452452	1.0000
40	SFMBT2	10	7450242-7450831	1.0000
	SOX5	12	24716178-24716294	1.0000
	VAV3	1	108507609-108507674	1.0000
	VAV3	1	108507074-108507498	1.0000
	ZNF132	19	58951402-58951530	1.0000
45	ADCY1	7	45613877-45613977	0,9984
	C1QL3	10	16562465-16562672	0,9984
	FLI1	11	128563956-128564209	0,9984
	MYH10	17	8533282-8534168	0,9984
	NTNG1	1	107683064-107683372	0,9984
50	ANKRD33B	5	10564655-10564807	0,9983
	MAX.chr10,22541502-22541671	10	22541502-22541671	0,9982
	MAX.chr11,44749119-44749205	11	44749119-44749205	0,9982
55	GALNT14	2	31360809-31360992	0,9980
	AKR1B1	7	134142981-134143723	0,9967
	CHST2	3	142839223-142839576	0,9967
	EYA4	6	133562229-133562380	0,9967
60	ZNF625	19	12267378-12267677	0,9967
	LMO1	11	8284746-8284871	0,9965
	ZNF469	16	88496963-88497197	0,9964
	ESR1	6	152129389-152129636	0,9951
65	KCNK12	2	47797187-47797452	0,9951

ES 2 812 753 T3

(Continuación)

	Anotación de genes	Cromosoma	Coordenada cromosómica	Área bajo la curva ROC
5	MAX.chr11,8040551-8040677	11	8040551-8040677	0,9951
	MOXD1	6	132722283-132722484	0,9951
	PPP2R5C	14	102247525-102247929	0,9951
	PPP2R5C	14	102248062-102248216	0,9949
10	CHST11	12	104850745-104851001	0,9948
	LASS4	19	8274584-8274671	0,9945
	GSG1L	16	28074472-28074761	0,9935
	MAX.chr11,120435350-120435981	11	120435350-120435981	0,9935
15	XKR6	8	10872819-10873619	0,9935
	MAX.chr1,244012766-244012875	1	244012766-244012875	0,9933
	DGKG	3	186079767-186080092	0,9931
	ITGB2	21	46351838-46352381	0,9931
20	MAX.chr19,22034447-22034696	19	22034447-22034696	0,9926
	ZNF347	19	53661526-53662618	0,9921
	MAX.chr8,30769438-30769680	8	30769438-30769680	0,9918
25	SOX5	12	24715012-24715416	0,9918
	CHST15	10	125852905-125853007	0,9913
	ODZ4	11	79150971-79151076	0,9913
	SOX21	13	95363646-95363959	0,9908
30	SEZ6L	22	26565137-26565417	0,9902
	GAS7	17	10102237-10102576	0,9899
	MAX.chr11,123301853-123301941	11	123301853-123301941	0,9889
	ELMO1	7	37487755-37488565	0,9886
35	EMPUJE	10	17270955-17271052	0,9886
	WNT3A	1	228195339-228195413	0,9886
	SLC8A3	14	70655516-70655712	0,9879
	PLEKHO1	1	150122783-150123157	0,9869
40	SLC8A3	14	70655268-70655368	0,9869
	ZNF682	19	20149796-20149923	0,9869
	ADORA2A	22	24820148-24820373	0,9857
	ELOVL2	6	11044395-11044834	0,9853
	GFRA3	5	137610023-137610333	0,9853
45	SOX5	12	24715703-24715776	0,9847
	EHD3	2	31456804-31457263	0,9837
	TMEM163	2	135476019-135476390	0,9837
	MAX.chr14,105400087-105400182	14	105400087-105400182	0,9835
50	MACROD1	11	63828346-63828436	0,9833
	ANKRD33B	5	10564406-10564551	0,9820
	MATK	19	3785828-3786371	0,9820
	NTNG1	1	107683961-107684314	0,9820
	ONECUT1	15	53082447-53083044	0,9820
55	WIPF1	2	175547056-175547390	0,9820
	GRID1	10	88125930-88126495	0,9804
	MAX.chr11,123301366-123301506	11	123301366-123301506	0,9804
	RASSF2	20	4803969-4804077	0,9804
60	TCF4	18	53255390-53255565	0,9804
	TRIB2	2	12857915-12858230	0,9804
	ZNF331	19	54024023-54024436	0,9804
	SLC35F1	6	118228394-118228979	0,9794
65	COL23A1	5	178016833-178017456	0,9788
	FAM159A	1	53098973-53099237	0,9788

ES 2 812 753 T3

(Continuación)

	Anotación de genes	Cromosoma	Coordenada cromosómica	Área bajo la curva ROC
5	GABBR2	9	101471421-101471519	0,9788
	CHST2	3	142838645-142839023	0,9775
	MAX.chr19,22034799-22034887	19	22034799-22034887	0,9775
	CHST10	2	101033758-101034005	0,9771
10	GFPT2	5	179780627-179781188	0,9771
	IKZF1	7	50343838-50344453	0,9771
	PRKCB	16	23847825-23848168	0,9766
	GRID2IP	7	6570511-6570865	0,9758
	HOMER2	15	83621302-83621657	0,9758
15	CHST1	11	45686306-45686534	0,9755
	MAX.chr18,77558550-77558609	18	77558550-77558609	0,9753
	LINGO3	19	2290273-2290393	0,9740
	LASS4	19	8274360-8274430	0,9739
20	ZEB1	10	31609049-31609227	0,9722
	ZNF43	19	22018746-22019004	0,9722
	C12orf39	12	21680721-21680828	0,9706
	MAX.chr10,22541884-22542001	10	22541884-22542001	0,9706
25	MAX.chr11,123301058-123301255	11	123301058-123301255	0,9706
	EMPUJE	10	17271896-17271978	0,9699
	AKAP12	6	151561236-151561473	0,9690
	C16orf5	16	4588091-4588817	0,9690
30	RORB	9	77111900-77112005	0,9690
	NRCAM	7	108095348-108095805	0,9689
	ELOVL4	6	80656845-80657306	0,9683
	CHST2	3	142838025-142838494	0,9673
35	ITPKB	1	226925082-226925651	0,9673
	PIK3CD	1	9711854-9711974	0,9673
	SARM1	17	26698693-26699117	0,9673
	GDF1	19	19006296-19006511	0,9671
	XKR6	8	10873760-10874271	0,9671
40	C9orf50	9	132382813-132382909	0,9659
	MAX.chr12,133484966-133485857	12	133484966-133485857	0,9657
	PURG	8	30890580-30890912	0,9643
45	AFF3	2	100721643-100721967	0,9641
	PDGFRA	4	55099106-55099473	0,9641
	MAX.chr3,44039952-44040054	3	44039952-44040054	0,9637
	TWIST1	7	19156788-19156858	0,9632
50	MEF2C	5	88185490-88185589	0,9608
	VCAN	5	82768837-82769031	0,9602
	NCAM1	11	112832731-112832815	0,9600
	CAMK4	5	110559508-110560719	0,9592
55	TBC1D30	12	65218900-65218994	0,9587
	BMP6	6	7727566-7728088	0,9585
	BAALC	8	104152806-104153145	0,9575
	GLB1L3	11	134146132-134146380	0,9575
	KRBA1	7	149411729-149411847	0,9569
60	TCF24	8	67874670-67875083	0,9550
	NTN1	17	8925482-8925838	0,9542
	CAMK2B	7	44364925-44365359	0,9516
	MAX.chr2,144694517-144695025	2	144694517-144695025	0,9510
65	SDK1	7	3339895-3340903	0,9510

ES 2 812 753 T3

(Continuación)

	Anotación de genes	Cromosoma	Coordenada cromosómica	Área bajo la curva ROC
5	SRRM3	7	75896637-75896925	0,9498
	CLSTN2	3	139654045-139654299	0,9493
	SELV	19	40005836-40005892	0,9487
	LY86-AS1	6	6546375-6546598	0,9477
	PPM1E	17	56833684-56833978	0,9477
10	TM6SF1	15	83776196-83776373	0,9462
	MAX.chr1,244013190-244013393	1	244013190-244013393	0,9461
	MAX.chr1,39269813-39270150	1	39269813-39270150	0,9461
15	MAX.chr14,100437680-100437767	14	100437680-100437767	0,9446
	EGR2	10	64575060-64575283	0,9444
	SOX7	8	10587893-10588143	0,9428
	LRRC4	7	127671918-127672318	0,9395
20	RGMA	15	93631919-93632242	0,9395
	ZNF804A	2	185463105-185463763	0,9395
	C1QL3	10	16562866-16563332	0,9373
	SFMBT2	10	7451097-7451185	0,9366
	CHFR	12	133464655-133464819	0,9360
25	JAKMIP1	4	6201350-6201560	0,9360
	ANK1	8	41754327-41754726	0,9346
	FAM126A	7	23053043-23053438	0,9346
	SFMBT2	10	7452885-7452956	0,9338
30	MME	3	154797723-154797909	0,9329
	BHLHE22	8	65494269-65494355	0,9323
	DPY19L2	12	64062131-64062443	0,9314
	VASH2	1	213124472-213124778	0,9302
	PREX1	20	47444582-47444776	0,9286
35	ARHGAP20	11	110582796-110583345	0,9248
	MAX.chr10,22765155-22765223	10	22765155-22765223	0,9242
	ZNF141	4	332064-332199	0,9221
40	DNAJC6	1	65731412-65731782	0,9216
	PCDHGA1	5	140855415-140856027	0,9216
	QKI	6	163835376-163835472	0,9213
	MAX.chr3,115231555-115231576	3	115231555-115231576	0,9199
45	GLIS1	1	54204505-54204712	0,9178
	ZEB1	10	31608394-31608690	0,9150
	BVES	6	105584685-105585220	0,9134
	LOC100131199	7	140773610-140773855	0,9134
	PAQR9	3	142682282-142682813	0,9134
50	CD109	6	74405903-74406086	0,9126
	RAX	18	56936593-56936656	0,9091
	C12orf68	12	48577334-48577557	0,9069
	SNX32	11	65601167-65601514	0,9069
	HLA-DQB1	6	32632785-32632860	0,9058
55	PRKCB	16	23846951-23847056	0,9053
	GLI3	7	42276418-42277414	0,9036
	Septin9	17	75368800-75369056	0,9031
	FAM72A	1	206137408-206137473	0,9016
60	NDRG4	16	58497979-58498250	0,9003
	FAM72B	1	120838272-120838775	0,8995
	KIF5C	2	149633039-149633137	0,8990
	JAKMIP1	4	6202051-6202410	0,8989
	CHST15	10	125852559-125852792	0,8989
65	MDGA1	6	37664238-37664539	0,8987

(Continuación)

	Anotación de genes	Cromosoma	Coordenada cromosómica	Área bajo la curva ROC
5	NGFR	17	47573986-47574084	0,8981
	ZEB1	10	31608798-31608892	0,8979
	ZNF529	19	37064200-37064435	0,8979
	ZNF287	17	16472295-16472694	0,8962
	C1orf70	1	1476065-1476127	0,8962
10	C20orf194	20	3388089-3388291	0,8932
	TIAM1	21	32930226-32930576	0,8910
	NR3C1	5	142784971-142785160	0,8895
	CBLN2	18	70211543-70211719	0,8873
	NKX6-3	8	41503949-41504137	0,8858
15	TMEM178	2	39893089-39893224	0,8856
	KHDC1	6	73972941-73973104	0,8845
	PRDM6	5	122425730-122425886	0,8824
	ST3GAL1	8	134583587-134583963	0,8824
	FAM72B	1	120839339-120839381	0,8815
20	DAB2IP	9	124461296-124461420	0,8787
	CHST15	10	125852012-125852098	0,8780
	C6orf147	6	74019826-74019955	0,8775
25	MAX.chr21,47063135-47064177	21	47063135-47064177	0,8775
	TIAM1	21	32931523-32931688	0,8772
	C1QL3	10	16563667-16563892	0,8758
	NUP210	3	13461109-13461191	0,8755
	FAM72B	1	120836675-120836768	0,8736
30	MAX.chr15,34806855-34807014	15	34806855-34807014	0,8722
	CAND2	3	12838197-12838303	0,8719
	TBX21	17	45810562-45810819	0,8702
	MGAT5B	17	74864552-74864821	0,8685
35	ZNF569	19	37960066-37960505	0,8676
	NTNG1	1	107684447-107684545	0,8672
	WASF3	13	27131683-27131757	0,8647
	LIPE	19	42905798-42906349	0,8644
40	DOCK10	2	225906664-225906922	0,8578
	LTBP4	19	41119795-41119907	0,8546
	MAX.chr19,20959229-20959691	19	20959229-20959691	0,8521
	FGR	1	27960931-27961018	0,8513

Tabla 5A - DMR de pólipos de sésiles serrados (SSP)

	Cromosoma	Coordenadas cromosómicas	Anotación de genes
50	5	10564655-10564710	ANKRD33B
	12	103351885-103351983	ASCL1
	1	203598574-203598800	ATP2B4
	6	91005003-91005091	BACH2
	6	7727026-7727129	BMP6
55	6	105584890-105584983	BVES
	10	21784521-21784567	C10orf114
	1	226737152-226737231	C1orf95
	5	110559571-110559638	CAMK4
	2	56411545-56411640	CCDC85A
60	1	158150837-158150885	CD1D
	1	158151102-158151205	CD1D
	10	90967004-90967028	CH25H
	2	101033768-101033858	CHST10
	12	104850745-104850879	CHST11
65	3	142838194-142838411	CHST2

ES 2 812 753 T3

(Continuación)

	Cromosoma	Coordenadas cromosómicas	Anotación de genes
5	9	34590231-34590344	CNTFR
	12	49484149-49484231	DHH
	15	30484732-30484813	DKFZP434L187
	1	65731412-65731530	DNAJC6
	2	225906664-225906763	DOCK10
10	2	225907515-225907632	DOCK10
	2	73520956-73520964	EGR4
	6	80656889-80656974	ELOVL4
	6	80657208-80657306	ELOVL4
	5	111754713-111754810	EPB41L4A
15	3	96532270-96532344	EPHA6
	7	23053937-23054066	FAM126A
	11	125365196-125365327	FEZ1
	13	22243643-22243727	FGF9
	7	90894876-90894960	FZD1
20	17	42907793-42907827	GJC1
	3	179169408-179169505	GNB4
	1	101005577-101005661	GPR88
	1	53068071-53068182	GPX7
25	7	6570755-6570845	GRID2IP
	16	10275378-10275472	GRIN2A
	17	14205388-14205498	HS3ST3B1
	7	23509037-23509225	IGF2BP3
	6	39281409-39281488	KCNK17
30	15	79724426-79724525	KIAA1024
	2	208031024-208031104	KLF7
	2	208031731-208031826	KLF7
	2	30454421-30454492	LBH
	2	30454871-30454977	LBH
35	2	74726179-74726265	LBX2
	5	87970308-87970374	LOC645323
	5	87970772-87970894	LOC645323
	2	170220089-170220148	LRP2
40	12	40618617-40618655	LRRK2
	1	25944147-25944152	MAN1C1
	11	123301366-123301387	MAX.chr11,123301366-123301387
	17	45867397-45867662	MAX.chr17,45867397-45867662
	19	55963254-55963329	MAX.chr19,55963254-55963329
45	2	127783352-127783403	MAX.chr2,127783352-127783403
	2	96192422-96192521	MAX.chr2,96192422-96192521
	20	1783778-1783841	MAX.chr20,1783778-1783841
	22	42310340-42310438	MAX.chr22,42310340-42310438
50	3	43935668-43935753	MAX.chr3,43935668-43935753
	4	186049639-186049660	MAX.chr4,186049639-186049660
	6	114664537-114664631	MAX.chr6,114664537-114664631
	7	127807622-127807693	MAX.chr7,127807622-127807693
	7	149745500-149745592	MAX.chr7,149745500-149745592
55	9	114074-114160	MAX.chr9,114074-114160
	9	114354-114435	MAX.chr9,114354-114435
	9	99983903-99984118	MAX.chr9,99983903-99984118
	6	37664654-37664664	MDGA1
	15	66546092-66546108	MEGF11
60	15	82339716-82339790	MEX3B
	11	30607877-30607973	MPPED2
	4	113437673-113437953	NEUROG2
	1	153651840-153651933	NPR1
	5	142785023-142785050	NR3C1
65	8	32406662-32406739	NRG1

ES 2 812 753 T3

(Continuación)

Cromosoma	Coordenadas cromosómicas	Anotación de genes
1	40137384-40137471	NT5C1A
1	107684212-107684314	NTNG1
1	107683064-107683130	NTNG1
5	140855796-140855883	PCDHGA1
1	9711931-9711974	PIK3CD
2	198669944-198670044	PLCL1
1	150122951-150122989	PLEKHO1
14	102247811-102247881	PPP2R5C
9	33677215-33677313	PTENP1
18	9708795-9708891	RAB31
18	9708515-9708598	RAB31
1	167599730-167599772	RCSD1
1	44872395-44872487	RNF220
9	94712910-94712961	ROR2
9	77111911-77112005	RORB
17	1928103-1928210	RTN4RL1
1	101702045-101702063	S1PR1
1	860904-860978	SAMD11
22	42949849-42949919	SERHL2
10	7450571-7450659	SFMBT2
1	220101492-220101587	SLC30A10
6	118228394-118228493	SLC35F1
12	24715012-24715060	SOX5
12	24715174-24715255	SOX5
10	73847865-73847982	SPOCK2
18	52989026-52989191	TCF4
13	43148769-43148861	TNFSF11
9	135285696-135285788	TTF1
6	149069140-149069222	UST
3	55521770-55521861	WNT5A
10	31608625-31608690	ZEB1
19	58666209-58666308	ZNF329
19	20149832-20149923	ZNF682
19	53073640-53073729	ZNF701
7	6655558-6655640	ZNF853

5

10

15

20

25

30

35

40

45

50

55

60

65

Tabla 5B: DMR de pólipos de sésiles serrados (SSP) clasificados por área bajo la curva ROC

	Anotación de genes	Cromosoma	Coordenada cromosómica	Área bajo la curva ROC
5	CAMK4	5	110559571-110559638	1.0000
	FGF9	13	22243643-22243727	1.0000
	GJC1	17	42907793-42907827	1.0000
	GPX7	1	53068071-53068182	1.0000
10	GRIN2A	16	10275378-10275472	1.0000
	IGF2BP3	7	23509037-23509225	1.0000
	MAX.chr19,55963254-55963329	19	55963254-55963329	1.0000
15	MAX.chr2,127783352-127783403	2	127783352-127783403	1.0000
	MAX.chr4,186049639-186049660	4	186049639-186049660	1.0000
	NRG1	8	32406662-32406739	1.0000
	NTNG1	1	107684212-107684314	1.0000
20	PIK3CD	1	9711931-9711974	1.0000
	PTENP1	9	33677215-33677313	1.0000
	RAB31	18	9708795-9708891	1.0000
	RNF220	1	44872395-44872487	1.0000
25	SLC30A10	1	220101492-220101587	1.0000
	ZNF853	7	6655558-6655640	1.0000
	LBH	2	30454871-30454977	0,9967
	DOCK10	2	225906664-225906763	0,9958
	MEX3B	15	82339716-82339790	0,9951
30	PCDHGA1	5	140855796-140855883	0,9926
	FZD1	7	90894876-90894960	0,9916
	GNB4	3	179169408-179169505	0,9916
	MAN1C1	1	25944147-25944152	0,9916
	ZNF701	19	53073640-53073729	0,9916
35	DKFZP434L187	15	30484732-30484813	0,9902
	RTN4RL1	17	1928103-1928210	0,9902
	MAX.chr20,1783778-1783841	20	1783778-1783841	0,9890
	SPOCK2	10	73847865-73847982	0,9853
40	MAX.chr7,149745500-149745592	7	149745500-149745592	0,9804
	HS3ST3B1	17	14205388-14205498	0,9748
	LRP2	2	170220089-170220148	0,9748
	MAX.chr9,114354-114435	9	114354-114435	0,9748
45	ELOVL4	6	80656889-80656974	0,9706
	ELOVL4	6	80657208-80657306	0,9706
	MPPED2	11	30607877-30607973	0,9706
	MAX.chr17,45867397-45867662	17	45867397-45867662	0,9664
50	CD1D	1	158150837-158150885	0,9618
	ATP2B4	1	203598574-203598800	0,9559
	C10orf114	10	21784521-21784567	0,9496
	KCNK17	6	39281409-39281488	0,9485
	FEZ1	11	125365196-125365327	0,9429
55	PLCL1	2	198669944-198670044	0,9363
	CD1D	1	158151102-158151205	0,9314
	NT5C1A	1	40137384-40137471	0,9286
	MAX.chr9,99983903-99984118	9	99983903-99984118	0,9191
60	GPR88	1	101005577-101005661	0,9069
	RAB31	18	9708515-9708598	0,9069
	NR3C1	5	142785023-142785050	0,9048
	S1PR1	1	101702045-101702063	0,9044
	SOX5	12	24715012-24715060	0,9044
65	EPB41L4A	5	111754713-111754810	0,9034

(Continuación)

	Anotación de genes	Cromosoma	Coordenada cromosómica	Area bajo la curva ROC
5	LOC645323	5	87970772-87970894	0,9034
	CHST11	12	104850745-104850879	0,8971
	ROR2	9	94712910-94712961	0,8946
	MAX.chr22,42310340-42310438	22	42310340-42310438	0,8918
10	RORB	9	77111911-77112005	0,8905
	ZEB1	10	31608625-31608690	0,8897
	NEUROG2	4	113437673-113437953	0,8856
	TCF4	18	52989026-52989191	0,8807
15	MAX.chr3,43935668-43935753	3	43935668-43935753	0,8750
	KLF7	2	208031731-208031826	0,8739
	ZNF682	19	20149832-20149923	0,8739
	DOCK10	2	225907515-225907632	0,8732
	MDGA1	6	37664654-37664664	0,8725
20	LBH	2	30454421-30454492	0,8718
	WNT5A	3	55521770-55521861	0,8718
	BVES	6	105584890-105584983	0,8706
	SAMD11	1	860904-860978	0,8706
	MEGF11	15	66546092-66546108	0,8701
25	LRRK2	12	40618617-40618655	0,8697
	BACH2	6	91005003-91005091	0,8687
	CHST10	2	101033768-101033858	0,8676
	CHST2	3	142838194-142838411	0,8676
30	MAX.chr9,114074-114160	9	114074-114160	0,8664
	EGR4	2	73520956-73520964	0,8627
	NTNG1	1	107683064-107683130	0,8613
	TTF1	9	135285696-135285788	0,8613
	SOX5	12	24715174-24715255	0,8611
35	ASCL1	12	103351885-103351983	0,8603
	KLF7	2	208031024-208031104	0,8600
	CNTFR	9	34590231-34590344	0,8592
	MAX.chr6,114664537-114664631	6	114664537-114664631	0,8592
40	SFMBT2	10	7450571-7450659	0,8585
	TNFSF11	13	43148769-43148861	0,8571
	DNAJC6	1	65731412-65731530	0,8550
	GRID2IP	7	6570755-6570845	0,8550
45	BMP6	6	7727026-7727129	0,8474
	MAX.chr11,123301366-123301387	11	123301366-123301387	0,8471
	PPP2R5C	14	102247811-102247881	0,8466
	SERHL2	22	42949849-42949919	0,8464
50	RCSD1	1	167599730-167599772	0,8456
	MAX.chr2,96192422-96192521	2	96192422-96192521	0,8431
	ANKRD33B	5	10564655-10564710	0,8407
	NPR1	1	153651840-153651933	0,8384
	EPHA6	3	96532270-96532344	0,8267
55	FAM126A	7	23053937-23054066	0,8262
	CH25H	10	90967004-90967028	0,8235
	KIAA1024	15	79724426-79724525	0,8235
	LOC645323	5	87970308-87970374	0,8235
	ZNF329	19	58666209-58666308	0,8224
60	UST	6	149069140-149069222	0,8199
	CCDC85A	2	56411545-56411640	0,8193
	PLEKH01	1	150122951-150122989	0,8176
	SLC35F1	6	118228394-118228493	0,8082
65	LBX2	2	74726179-74726265	0,8059

(Continuación)

Anotación de genes	Cromosoma	Coordenada cromosómica	Área bajo la curva ROC
DHH	12	49484149-49484231	0,8046
MAX.chr7,127807622-127807693	7	127807622-127807693	0,8007
C1orf95	1	226737152-226737231	0,8000

Ejemplo 3.

[0227] Este ejemplo demuestra NDRG4, BMP3, OPLAH, FLI1, PDGFD, CHST_7889, SFMBT2_895, SFMBT2_896, SFMBT2_897, CHST27890, VAV3 y DTX1 como marcadores efectivos para detectar el cáncer colorrectal en muestras de heces.

[0228] Las secuencias de cebador y sonda directa e inversa para NDRG4, BMP3, OPLAH, FLI1, PDGFD, CHST_7889, SFMBT2_895, SFMBT2_896, SFMBT2_897, CHST2_7890, VAV3 y DTX1 se proporcionan en la Tabla 6A. La Tabla 6B proporciona información sobre el marcador metilado, el cromosoma y las coordenadas genómicas de DMR proporcionadas en la Tabla 6A.

[0229] Las sondas de captura para cada marcador fueron diseñadas para tener una temperatura de fusión de 75-80°C y longitudes entre 25-35 bases (ver, Tabla 6C). Además, la región de hibridación de la sonda de captura se seleccionó para estar dentro de la huella de QuARTS postbisulfito. La Tabla 6C proporciona el marcador de metilación y las secuencias de sonda de captura respectivas.

Tabla 6A: Secuencias de cebador directo, sonda, cebador inverso para marcadores utilizados en el ejemplo 3

Marcador de metilación	Cebador directo		Secuencia de la sonda		Imprimación inversa	
	Secuencia	NO: (SEQ ID)	Secuencia	NO: (SEQ ID)	Secuencia	NO: (SEQ ID)
VAV3	TCGGAGTCGA GTTTAGCGC (SEQ ID NO: 54)		CCACGGACG- CGGCGTTCGGA/3C C6/ (SEQ ID NO: 55)		CGAAATCGAAAAACAAA AACCGC (SEQ ID NO: 56)	
CHST2_7889	CGAGTTCGGT AGTTGTACGT AGA (SEQ ID NO: 57)		CGCCGAGG- TCGTCGATACCG/3C 6/ (SEQ ID NO: 58)		CGAAATACGAACGCGAA ATCTAAAACT (SEQ ID NO: 59)	
SFMBT2_897	GTcGTcGTTcG AGAGGGTA (SEQ ID NO: 60)		CCACGGACG- ATCGGTTTCGGTT/3C 6/ (SEQ ID NO: 61)		CGAACAAAAACGAACGA ACGAA (SEQ ID NO: 62)	
SFMBT2_896	GCGTTTAGGT TGGTCGGAGA (Version1) (SEQ ID NO: 63) GCGTTTAGGT TGGTCGGAG (Version 2) (SEQ ID NO: 90)		CGCCGAGG- CTACGAACCGAA/3C 6/ (Version 1) (SEQ ID NO: 64) CGCCGAGGCCGAA AAACTAC/3C6/ (Version 2) (SEQ ID NO: 91)		CCTAACCAACGCACTCA ACC (Version 1) (SEQ ID NO: 65) ACGCACTCAACCTACGA AC (Version 2) (SEQ ID NO: 92)	

(Continuación)

Marcador de metilación	Cebador directo	Secuencia de la sonda	Imprimación inversa
SFMBT2_895	TTAGCGAcGTA GTcGTcGTTG (Version 1) (SEQ ID NO: 66) GCGACGTAGT CGTCGTTGT (Version 2) (SEQ ID NO: 93)	CCACGGACG- CGAAAACGGGAA/3C 6/ (Version 1) (SEQ ID NO: 67) CCACGGACGGAAAA CGCGAAA/3C6/ (Version 2) (SEQ ID NO: 94)	CCCAACGCGAAAAAAC GC (Version 1) (SEQ ID NO: 68) CCAACGCGAAAAAACG CG (Version 2) (SEQ ID NO: 95)
CHST2_7890	GTATAGCGCG ATTTTCGTAGcG (SEQ ID NO: 69)	CGCCGAGG- CGAACATCCTCC/3C 6/ (SEQ ID NO: 70)	AATTACCTACGCTATCCG CCC (SEQ ID NO: 71)
OPLAH	cGTcGcGTTTT TcGGTTATACG (SEQ ID NO: 72)	CCACGGACG- GCACCGTAAAAC/3C 6/ (SEQ ID NO: 73)	CGCGAAAACATAAAC CGCG (SEQ ID NO: 74)

(Continuación)

Marcador de metilación	Cebador directo	Secuencia de la sonda	Imprimación inversa
PDGFD	AAACGTTAATT TGTTGTTTGTT TCGTT (Version 1) (SEQ ID NO: 75) GCGAATAAAT AAACGTTAATT TGTTGTTTGTT TCG (Version 2) (SEQ ID NO: 96)	ACTTCCGACGCG TATAAATACC (Version 1) (SEQ ID NO: 76) CCACGGACGCGCA CTTCCTTA/3C6/ (Version 2) (SEQ ID NO: 97)	GCGAATAAATAACGTTA ATTTGTTGTTTTCG (Version 1) (SEQ ID NO: 77) ACTTCCGACGCGTATA AATACC (Version 2) (SEQ ID NO: 98)
FLI1	GTTGcGAGGT TAGGTTGTAAT CG (SEQ ID NO: 78)	CGCCGAGG- CGTCCATTAAAC/3C 6/ (SEQ ID NO: 79)	CGCCGCTTACCCTTAATAA TCCC (SEQ ID NO: 80)
DTX1	GAGTCGCGGT TTCGTTTTTC (SEQ ID NO: 81)	CGCCGAGG- CGCGTTCGTTTT /3C6/(SEQ ID NO: 82)	GACGCGACGACCGAAAA AC (SEQ ID NO: 83)
NDRG4	CGGTTTTTCGT TCGTTTTTTTCG (SEQ ID NO: 84)	CCACGGACG GTTTCGTTTATCG/3C 6/ (SEQ ID NO: 85)	CCGCCTTCTACGCGACT A (SEQ ID NO: 86)

(Continuación)

Marcador de metilación	Cebador directo	Secuencia de la sonda	Imprimación inversa
BMP3	GTTTAAATTTTC GGTTTCGTTCG TC (SEQ ID NO: 87)	CGCCGAGG CGGTTTTTGGCG/3C 6/ (SEQ ID NO: 88)	CGCTACGAAACACTCCG A (SEQ ID NO: 89)

Tabla 6B. Marcador metilado, cromosoma y coordenadas genómicas DMR.

Marcador metilado	Cromosoma	Coordenadas genómicas DMR
BMP3	4	81031173 - 81031262
NDRG4	16	58463478 - 58463588
VAV3	1	107964966 - 107965057
CHST2_7889	3	143119424 - 143119583
SFMBT2_897	10	7410903 - 7411014
SFMBT2_896	10	7410764 - 7410837
SFMBT2_895	10	7410331 - 7410490
CHST2_7890	3	143119999 - 143120158
OPLAH	8	144051847 - 144052006
PDGFD	11	104164082 - 104164186
FLI1	11	128694158 - 128694317
DTX1	12	113056762 - 113056895

Tabla 6C. Marcador de metilación y secuencias de sonda de captura respectivas.

Marcador de metilación	Secuencia de captura de sonda
VAV3	/5AmMC6/GATCGAGGGAGCAGGAGCCGCGGCTGACGGGTC GCG (SEQ ID NO: 99)
CHST2_7889	/5AmMC6/CGGTGCCGAGAGCTGCCAGAGAGTTGGATTCTGC G (SEQ ID NO: 100)
SFMBT2_897	/5AmMC6/GCGAGCGGGCAAGGGCGGGCGAGC (SEQ ID NO: 101)
SFMBT2_896	/5AmMC6/ACCTGCGGGCCGAAGGGCTGCTCTCCGG (SEC ID 102)
SFMBT2_895	/5AmMC6/AGGAGACGCGGGAGCGCGGGGTAGGTAGC (SEC 103)
CHST2_7890	/5AmMC6/GGCATCCTCCCGGTGATGGAAGCAGCCGCCG C (SEQ ID NO: 104)
OPLAH	/5AmMC6/GGAAGGCGCGGCGCTCGGTCAGCACTGACAGCA G (SEQ ID NO: 105)
PDGFD	/5AmMC6/TCGCCGAGCTCTCCCAAACCTTCTGCATGCTGAA CTTT (SEQ ID NO: 106)
FLI1	/5AmMC6/CCGTCCATTTGGCCAAGTCTGCAGCCGAGCC (SEQ ID NO: 107)
DTX1	/5AmMC6/CTGCGTCCGTCCGTCCGCGCCGGCAGTCTGTCCA (SEQ ID NO: 108)
NDRG4	/5AmMC6/TCCCTCGCGCGTGGCTTCCGCCTTCTGCGCGGCT GGGGTGCCCGGTGG (SEQ ID NO: 109)
BMP3	/5AmMC6/GCGGGACACTCCGAAGGCGCAAGGAG (SEC ID N.º: 110)

[0230] Cada sonda de captura se sintetizó con una modificación 5'-NH₂ para permitir el acoplamiento a partículas magnéticas que son -COOH modificado a través de la química de acoplamiento de carbodiimida estándar. Además, un oligonucleótido complementario a la sonda de captura se sintetizó para contener una etiqueta 5'-Cy₃. Esta sonda complementaria se utilizó para confirmar el acoplamiento de la sonda de captura a partículas magnéticas.

[0231] Para probar la eficacia de captura de cada sonda, así como evaluar el rendimiento del marcador, se hicieron dos agrupaciones de heces de pacientes normales y con cáncer. El grupo de heces de cáncer provino de 6 pacientes (3 son CRC, 1 es un AA y 2 incógnitas). Del mismo modo, las heces normales provienen de 6 pacientes normales sin CRC. Las heces se prepararon mezclando el sobrenadante después de la centrifugación homogeneizada. El sobrenadante agrupado se dividió en alícuotas en muestras de captura única que contenían sobrenadantes de 14 ml.

[0232] Las sondas de captura fueron diseñadas para tener una temperatura de fusión de 75-80°C y longitudes entre 25-35 bases. Además, la región de hibridación de la sonda de captura se seleccionó para estar dentro de la huella QuARTS posterior al bisulfito.

[0233] Para realizar la captura, las bolas de captura (partículas magnéticas con sondas de captura unidas

covalentemente) para dos marcadores más las bolas de captura de ACTB se agruparon para formar un grupo de bolas de captura triple.

5 **[0234]** La captura se realizó en triplicados de 14 ml de sobrenadantes de heces positivos y normales (la excepción fue VAV3 y DTX1 se realizaron por duplicado ya que el grupo se estaba agotando).

10 **[0235]** Después de completar la captura, el ADN de las heces se eluyó de las perlas de captura con 0,1 N NaOH a 42°C durante 20 minutos seguido de conversión de bisulfito a 56°C durante 1 hora usando bisulfito de amonio. Luego se desulfonó el ADN de las heces y se purificó usando perlas magnéticas recubiertas de sílice y se eluyó en 70 µl de Tris-HCl 10 mM, pH 8, EDTA 0,1 mM. Luego se ensayaron 10 uL de eluyente y se cuantificaron en ensayos QuARTS.

15 **[0236]** Para poder cuantificar las muestras, se usaron plásmidos pUC57 con insertos de ADN correspondientes a las huellas QuARTS. Los insertos de ADN se flanquearon con sitios EcoRI para permitir la linealización y cuantificación usando absorbancia a 260 nm.

[0237] Para permitir el cálculo inverso de las hebras en las muestras eluidas, se realizaron ensayos QuARTS en 10 µl de eluyentes y diluciones en serie de los plásmidos digeridos.

20 **[0238]** La Tabla 6D muestra los resultados obtenidos para cada uno de los marcadores probados. Estos resultados muestran que los marcadores de metilación tenían grandes diferencias de cadena entre el grupo de heces positivo y normal, lo que indica que estos marcadores son candidatos para la detección de CRC en las heces.

Tabla 6D.

Marcador de metilación	Hebras medias de agrupaciones positivas	Hebras promedio de la agrupación normal	Diferencias de fold (Positivo/Normal)
NDRG4	1.568	140	11,2
BMP3	395	8	51,4
OPLAH	840	279	3,0
FLI1	1,715	167	10,2
PDGFD	843	67	12,6
CHST2_7889	945	17	56,4
SFMBT2_895	837	5	152,3
SFMBT2_896	856	150	5,7
SFMBT2_897	844	45	18,7
CHST2_7890	1.396	62	22,6
VAV3	367	21	17,9
DTX1	751	105	7,2

Ejemplo 4.

45 **[0239]** Procedimiento ejemplar para la detección de un cáncer colorrectal en un sujeto humano.

50 **[0240]** Contactar con un ácido nucleico (por ejemplo, ADN genómico, por ejemplo, aislado de fluidos corporales como una muestra de heces o tejido colorrectal) obtenido de un sujeto humano con al menos un reactivo o serie de reactivos que distingue entre dinucleótidos CpG metilados y no metilados dentro de al menos un marcador DMR seleccionado de:

- FLI1, OPLAH, DTX1, MATK y SFMBT2 región 2; (como se indica en la Tabla 1);
- BMP3, NDRG4, PDGFG, CHST2 y SFMBT2 (como se indica en la Tabla 2);
- SFMBT2, LIFR, OPLAH, CRHR2, AGLB4, ZEB2, ZNF788, SFMBT2, ESR1, ANKRD33B, CHST2 y FNDC1 (como se indica en la Tabla 3); y
- NDRG4, BMP3, OPLAH, FLI1, PDGFD, CHST_7889, SFMBT2_895, SFMBT2_896, SFMBT2_897, CHST27890, VAV3 y DTX1 (como se indica en la Tabla 6).

60 **[0241]** Identificar al sujeto que tiene un cáncer colorrectal cuando el estado de metilación del marcador es diferente al estado de metilación del marcador analizado en un sujeto que no tiene una neoplasia.

Procedimiento ejemplar para la detección de un adenoma grande basado en colorrectal en un sujeto humano.

65 **[0242]** Contactar con un ácido nucleico (p. ej., ADN genómico, p. ej., aislado de fluidos corporales como una muestra de heces o tejido colorrectal) obtenido de un sujeto humano con al menos un reactivo o serie de reactivos que distingue entre dinucleótidos CpG metilados y no metilados dentro de al menos un marcador DMR seleccionado entre:

- ADD2, AGL4, AKAP12, ANKRD33B, ASCL1, Clorf70, CHST11, CHST15, DTX1, ECEL1, EYA4, FLI1, FLJ43390, FLT3, GRASP, ITGA4, K100N2, L100N2, MGAT3, OPLAH (145106742-145106921), OPLAH (145106349-145106456), PDE8B, PDGFD, PKIA, POU3F1, QKI, RASSF2, RSPO3, SDC2, SFMBT2 (7452746-7452779), SFMBT2 (452) -7450831), SOX5, VAV3 (108507609-108507674), VAV3 (108507074-108507498) y ZNF132 (como se indica en la Tabla 4).

[0243] Identificar al sujeto que tiene un adenoma grande basado en colorrectal cuando el estado de metilación del marcador es diferente al estado de metilación del marcador analizado en un sujeto que no tiene una neoplasia.

Procedimiento ejemplar para la detección de pólipos de sésiles serrados (SSP) en un sujeto humano.

[0244] Contactar con un ácido nucleico (por ejemplo, ADN genómico, por ejemplo, aislado de fluidos corporales como una muestra de heces o tejido colorrectal) obtenido de un sujeto humano con al menos un reactivo o serie de reactivos que distingue entre metilado y no dinucleótidos de CpG metilados dentro de al menos un marcador DMR seleccionado entre:

- CAMK4, FGF9, GJC1, GPX7, GRIN2A, IGF2BP3, MAX.chr19,55963254-55963329, MAX.chr2,127783352-127783403, MAX.chr4,18496, 3939401, NR18,160601, NTNG1, PIK3CD, PTENP1, RAB31, RNF220, SLC30A10 y ZNF853 (como se indica en la Tabla 5).

[0245] Identificar al sujeto que tiene pólipos de sésiles serrados cuando el estado de metilación del marcador es diferente al estado de metilación del marcador analizado en un sujeto que no tiene una neoplasia.

Procedimiento ejemplar para la detección de un cáncer colorrectal en un sujeto humano con enfermedad inflamatoria intestinal.

[0246] Contactar con un ácido nucleico (por ejemplo, ADN genómico, por ejemplo, aislado de fluidos corporales como una muestra de heces o tejido colorrectal) obtenido de un sujeto humano que tiene enfermedad inflamatoria intestinal con al menos un reactivo o una serie de reactivos que distingue entre dinucleótidos CpG metilados y no metilados dentro de al menos un marcador DMR seleccionado de:

- BMP3, NDRG4, PDGFG, CHST2 y SFMBT2 (como se indica en la Tabla 2).

[0247] Identificar al sujeto que tiene un cáncer colorrectal cuando el estado de metilación del marcador es diferente del estado de metilación del marcador analizado en un sujeto que no tiene una neoplasia.

LISTA DE SECUENCIAS

[0248]

<110> CORPORACIÓN DE CIENCIAS EXACTAS
FUNDACIÓN MAYO PARA LA EDUCACIÓN MÉDICA Y LA INVESTIGACIÓN

<120> DETECCIÓN DE NEOPLASMA COLORECTAL

<130> EXCTM-33841/WO-2/ORD

<150> US 61/977,954
<151> 2014-04-10

<150> US 61/972,942
<151> 2014-03-31

<160> 110

<170> Patente en la versión 3,5

<210> 1
<211> 21
<212> ADN
<213> Secuencia artificial

<220>
<223> Sintético

<400> 1
 gggagtgagg gtagggcgtt c 21
 5 <210> 2
 <211> 25
 <212> ADN
 <213> Secuencia artificial
 10 <220>
 <223> sintético
 <400> 2
 ctcgcaacc cttgaatta acccg 25
 15 <210> 3
 <211> 25
 <212> ADN
 <213> Secuencia artificial
 20 <220>
 <223> sintético
 <400> 3
 tgcgtaggtg atagggaggg gttac 25
 25 <210> 4
 <211> 25
 <212> ADN
 <213> Secuencia artificial
 30 <220>
 <223> sintético
 <400> 4
 acaaacaca tcctattaac gcgaa 25
 35 <210> 5
 <211> 19
 <212> ADN
 <213> Secuencia artificial
 40 <220>
 <223> sintético
 <400> 5
 gagtgcggt ttcgtttc 19
 45 <210> 6
 <211> 19
 <212> ADN
 <213> Secuencia artificial
 50 <220>
 <223> sintético
 <400> 6
 gacgacga ccgaaaaac 19
 55 <210> 7
 <211> 23
 <212> ADN
 <213> Secuencia artificial
 60 <220>
 <223> sintético
 65

<400> 7
 tgcacacccc gaggcggtcc cgg 23
 5 <210> 8
 <211> 23
 <212> ADN
 <213> Secuencia artificial
 10 <220>
 <223> sintético

 <400> 8
 cgccccaaa ataaaaaac gaa 23
 15 <210> 9
 <211> 0
 <212> ADN
 <213> Secuencia artificial
 20 <220>
 <223> sintético

 <400> 9
 000
 25 <210> 10
 <211> 20
 <212> ADN
 <213> Secuencia artificial
 30 <220>
 <223> sintético

 <400> 10
 gcgttaggt tggcggaga 20
 35 <210> 11
 <211> 20
 <212> ADN
 <213> Secuencia artificial
 40 <220>
 <223> sintético

 <400> 11
 cctaaccaac gcactcaacc 20
 45 <210> 12
 <211> 21
 <212> ADN
 <213> Artificial secuencia
 50 <220>
 <223> sintético

 <400> 12
 cgtagcgtgg cgtttagcg c 21
 55 <210> 13
 <211> 26
 <212> ADN
 <213> secuencia artificial
 60 <220>
 <223> sintético
 65

<400> 13
 tcgaaaaccc cgacgaaacg aaaacg 26

5

<210> 14
 <211> 25
 <212> ADN
 <213> Secuencia artificial

10

<220>
 <223> sintético

<400> 14
 gcgtaaggtc gaatattga gtcga 25

15

<210> 15
 <211> 25
 <212> ADN
 <213> Secuencia artificial

20

<220>
 <223> sintético

<400> 15
 aaaatactac ccaccaacca ccgaa 25

25

<210> 16
 <211> 19
 <212> ADN
 <213> Secuencia artificial

30

<220>
 <223> sintético

<400> 16
 gtcgctgctc gagaggta 19

35

<210> 17
 <211> 22
 <212> ADN
 <213> Secuencia artificial

40

<220>
 <223> sintético

<400> 17
 cgaacaaaa cgaacgaacg aa 22

45

<210> 18
 <211> 33
 <212> ADN
 <213> Secuencia artificial

50

<220>
 <223> sintético

<400> 18
 tcgatttat tttgtgtc gttgtagatt cgc 33

55

<210> 19
 <211> 33
 <212> ADN
 <213> Artificial secuencia

60

<220>
 <223> sintético

65

ES 2 812 753 T3

<400> 19
gaaaaaacta aaaaacgaca aaaaacccg acg 33

5 <210> 20
<211> 28
<212> ADN
<213> Secuencia artificial

10 <220>
<223> sintético

<400> 20
ggaacgagtg atagtcggat agc

15 <210> 21
<211> 18
<212> ADN
<213> Secuencia artificial

20 <220>
<223> sintético

<400> 21
cgccccgaaa cgaccccg 18

25 <210> 22
<211> 31
<212> ADN
<213> Secuencia artificial

30 <220>
<223> sintético

<400> 22
cggggatgat ttatgtagt cggagtctg c 31

35 <210> 23
<211> 24
<212> ADN
<213> Secuencia artificial

40 <220>
<223> sintético

45 <400> 23
ccgcccgaat actcgatcaa ctcg 24

50 <210> 24
<211> 34
<212> ADN
<213> Secuencia artificial

55 <220>
<223> sintético

<400> 24
gccaataaat aaacgtaat ttgtgttg ttc 34

60 <210> 25
<211> 25
<212> ADN
<213> Secuencia artificial

65 <220>
<223> sintético

<400> 25
 ccgaacgcgt ataaatacc 25

5

<210> 26
 <211> 28
 <212> ADN
 <213> Secuencia artificial

10

<220>
 <223> sintético

<400> 26
 cggttttatt tattatgatt cgtagcgg 28

15

<210> 27
 <211> 24
 <212> ADN
 <213> Secuencia artificial

20

<220>
 <223> sintético

<400> 27
 cgactaccct aaacaacgca tcgc 24

25

<210> 28
 <211> 23
 <212> ADN
 <213> Secuencia artificial

30

<220>
 <223> sintético

<400> 28
 cgagttcggg agttgtacgt aga 23

35

<210> 29
 <211> 27
 <212> ADN
 <213> Secuencia artificial

40

<220>
 <223> sintético

45

<400> 29
 cgaaatacga acgcgaaatc taaact 27

50

<210> 30
 <211> 19
 <212> ADN
 <213> Secuencia artificial

55

<220>
 <223> sintético

<400> 30
 gcgacgtagt cgtcgtgt 19

60

<210> 31
 <211> 19
 <212> ADN
 <213> Secuencia artificial

65

<220>
 <223> sintético

<400> 31
 ccaacgcgaa aaaaacgcg 19
 5 <210> 32
 <211> 19
 <212> ADN
 <213> Secuencia artificial
 10 <220>
 <223> sintético

 <400> 32
 gaggcggacg tcgcggtac 19
 15 <210> 33
 <211> 25
 <212> ADN
 <213> Secuencia artificial
 20 <220>
 <223> sintético

 <400> 33
 cgccacgacg cgaatcttaa ctacg 25
 25 <210> 34
 <211> 22
 <212> ADN
 <213> Secuencia artificial
 30 <220>
 <223> sintético

 <400> 34
 ggatcgaggg agtaggagtc gc 22
 35 <210> 35
 <211> 22
 <212> ADN
 <213> Secuencia artificial
 40 <220>
 <223> sintético

 <400> 35
 cgaaaccgaa cctaacgcga cg 22
 45 <210> 36
 <211> 24
 <212> ADN
 <213> Secuencia artificial
 50 <220>
 <223> sintético

 <400> 36
 agtttttcg cgcgttttt ttgc 24
 55 <210> 37
 <211> 21
 <212> ADN
 <213> Secuencia artificial
 60 <220>
 <223> sintético
 65 <220>
 <223> sintético

<400> 37
 gccgaatctc cgccttacac g 21

5

<210> 38
 <211> 23
 <212> ADN
 <213> Secuencia artificial

10

<220>
 <223> sintético

<400> 38
 gttaatttt cggttcgtc gtc 23

15

<210> 39
 <211> 20
 <212> ADN
 <213> Secuencia artificial

20

<220>
 <223> sintético

<400> 39
 cgccgaggcg gtttttgcg 20

25

<210> 40
 <211> 18
 <212> ADN
 <213> Secuencia artificial

30

<220>
 <223> sintético

<400> 40
 cgctacgaaa cactccga 18

35

<210> 41
 <211> 21
 <212> ADN
 <213> Secuencia artificial

40

<220>
 <223> sintético

45

<400> 41
 cggtttcgt tcgttttc g 21

50

<210> 42
 <211> 21
 <212> ADN
 <213> Secuencia artificial

55

<220>
 <223> sintético

<400> 42
 ccacggacgg ttcgtttatc g 21

60

<210> 43
 <211> 18
 <212> ADN
 <213> Secuencia artificial

65

<220>
 <223> sintético

<400> 43
 ccgccttcta cgcgacta 18

5

<210> 44
 <211> 34
 <212> ADN
 <213> Secuencia artificial

10

<220>
 <223> sintético

<400> 44
 gcgaataaat aaacgtaaat ttgtgttg ttc 34

15

<210> 45
 <211> 21
 <212> ADN
 <213> Secuencia artificial

20

<220>
 <223> sintético

<400> 45
 ccacggacgc gcacttcctt a 21

25

<210> 46
 <211> 25
 <212> ADN
 <213> Secuencia artificial

30

<220>
 <223> sintético

<400> 46
 ccgaacgcgt ataaataccg cactt 25

35

<210> 47
 <211> 23
 <212> ADN
 <213> Secuencia artificial

40

<220>
 <223> sintético

45

<400> 47
 cgagttcgg t agttgtacgt aga 23

50

<210> 48
 <211> 20
 <212> ADN
 <213> Artificial secuencia

55

<220>
 <223> sintético

<400> 48
 cgccgaggtc gtcgataccg 20

60

<210> 49
 <211> 27
 <212> ADN
 <213> Secuencia artificial

65

<220>
 <223> sintético

<400> 49
 cgaaatacga acgcgaaatc taaaact 27

5

<210> 50
 <211> 19
 <212> ADN
 <213> Secuencia artificial

10

<220>
 <223> sintético

<400> 50
 gcgacgtagt cgctgtgt 19

15

<210> 51
 <211> 21
 <212> ADN
 <213> Secuencia artificial

20

<220>
 <223> sintético

<400> 51
 ccacggacgg aaaacgcgaa a 21

25

<210> 52
 <211> 19
 <212> ADN
 <213> Secuencia artificial

30

<220>
 <223> sintético

<400> 52
 ccaacgcgaa aaaaacgcg 19

35

<210> 53
 <211> 0
 <212> ADN
 <213> Secuencia artificial

40

<220>
 <223> sintético

45

<400> 53
 000

<210> 54
 <211> 19
 <212> ADN
 <213> Secuencia artificial

50

<220>
 <223> sintético

<400> 54
 tcggagtcga gtttagcgc 19

55

<210> 55
 <211> 21
 <212> ADN
 <213> Secuencia artificial

60

<220>
 <223> sintético

65

<400> 55
 ccacggacgc ggcgttcgcg a 21

5 <210> 56
 <211> 24
 <212> ADN
 <213> Secuencia artificial

10 <220>
 <223> sintético

<400> 56
 cgaaatcgaa aaaacaaaaa ccgc 24

15 <210> 57
 <211> 23
 <212> ADN
 <213> Secuencia artificial

20 <220>
 <223> sintético

<400> 57
 cgagttcggc agttgtacgt aga 23

25 <210> 58
 <211> 20
 <212> ADN
 <213> Secuencia artificial

30 <220>
 <223> sintético

<400> 58
 cgccgaggtc gtcgataccg 20

35 <210> 59
 <211> 20
 <212> ADN
 <213> Secuencia artificial

40 <220>
 <223> sintético

<400> 59
 cgccgaggtc gtcgataccg 20

45 <210> 60
 <211> 19
 <212> ADN
 <213> Secuencia artificial

50 <220>
 <223> sintético

<400> 60
 gtcgtcgttc gagaggta 19

55 <210> 61
 <211> 21
 <212> ADN
 <213> Secuencia artificial

60 <220>
 <223> sintético

<400> 60
 gtcgtcgttc gagaggta 19

65 <210> 61
 <211> 21
 <212> ADN
 <213> Secuencia artificial

70 <220>
 <223> sintético

<400> 61
 ccacggacga tcggttcgt t 21
 5 <210> 62
 <211> 22
 <212> ADN
 <213> Secuencia artificial
 10 <220>
 <223> sintético
 <400> 62
 cgaacaaaaa cgaacgaacg aa 22
 15 <210> 63
 <211> 20
 <212> ADN
 <213> Secuencia artificial
 20 <220>
 <223> sintético
 <400> 63
 gcgtaggtg tggtcggaga 20
 25 <210> 64
 <211> 20
 <212> ADN
 <213> Secuencia artificial
 30 <220>
 <223> sintético
 <400> 64
 cgccgaggct acgaaccgaa 20
 35 <210> 65
 <211> 20
 <212> ADN
 <213> Secuencia artificial
 40 <220>
 <223> sintético
 <400> 65
 cctaaccaac gactcaacc 20
 45 <210> 66
 <211> 21
 <212> ADN
 <213> Secuencia artificial
 50 <220>
 <223> sintético
 <400> 66
 ttagcgacgt agtcgctgtt g 21
 55 <210> 67
 <211> 21
 <212> ADN
 <213> Secuencia artificial
 60 <220>
 <223> sintético
 65

<400> 67
 ccacggacgc gaaaacgcga a 21

5 <210> 68
 <211> 19
 <212> ADN
 <213> Secuencia artificial

10 <220>
 <223> sintético

15 <400> 68
 cccaacgcga aaaaaacgc 19

20 <210> 69
 <211> 21
 <212> ADN
 <213> Secuencia artificial

25 <220>
 <223> sintético

30 <400> 69
 gtatagcgcg atttcgtagc g 21

35 <210> 70
 <211> 20
 <212> ADN
 <213> Secuencia artificial

40 <220>
 <223> sintético

45 <400> 70
 cgccgaggcg aacatcctcc 20

50 <210> 71
 <211> 21
 <212> ADN
 <213> Secuencia artificial

55 <220>
 <223> sintético

60 <400> 71
 aattacctac gctatccgcc c 21

65 <210> 72
 <211> 22
 <212> ADN
 <213> Secuencia artificial

70 <220>
 <223> sintético

75 <400> 72
 cgtcgcggtt ttcggtata cg 22

80 <210> 73
 <211> 21
 <212> ADN
 <213> Secuencia artificial

85 <220>
 <223> sintético

<400> 73
 ccacggacgg caccgtaaaa c 21
 5 <210> 74
 <211> 21
 <212> ADN
 <213> Secuencia artificial
 10 <220>
 <223> sintético

 <400> 74
 cgcgaaaact aaaaaaccgc g 21
 15 <210> 75
 <211> 27
 <212> ADN
 <213> Secuencia artificial
 20 <220>
 <223> sintético

 <400> 75
 aaacgtaat ttgtgttg ttcggt 27
 25 <210> 76
 <211> 24
 <212> ADN
 <213> Secuencia artificial
 30 <220>
 <223> sintético

 <400> 76
 actttccgaa cgcgtataaa tacc 24
 35 <210> 77
 <211> 35
 <212> ADN
 <213> secuencia artificial
 40 <220>
 <223> sintético

 <400> 77
 gcgaataaat aaacgtaat ttgtgttg ttcg 35
 45 <210> 78
 <211> 23
 <212> ADN
 <213> Secuencia artificial
 50 <220>
 <223> sintético

 <400> 78
 gttgcgaggt taggtgtaa tcg 23
 55 <210> 79
 <211> 20
 <212> ADN
 <213> Secuencia artificial
 60 <220>
 <223> sintético
 65 <220>
 <223> sintético

<400> 79
 cgccgaggcg tccattaac 20

5

<210> 80
 <211> 22
 <212> ADN
 <213> Secuencia artificial

10

<220>
 <223> sintético

<400> 80
 cgccgcttac ctaataatc cc 22

15

<210> 81
 <211> 19
 <212> ADN
 <213> Secuencia artificial

20

<220>
 <223> sintético

<400> 81
 gagtcgcggt ttcgtttc 19

25

<210> 82
 <211> 20
 <212> ADN
 <213> Secuencia artificial

30

<220>
 <223> sintético

<400> 82
 cgccgaggcg cgttcgtttt 20

35

<210> 83
 <211> 19
 <212> ADN
 <213> Secuencia artificial

40

<220>
 <223> sintético

45

<400> 83
 gacgacga cggaaaaac 19

50

<210> 84
 <211> 21
 <212> ADN
 <213> Secuencia artificial

55

<220>
 <223> sintético

<400> 84
 cggtttcgt tcgttttc g 21

60

<210> 85
 <211> 21
 <212> ADN
 <213> Secuencia artificial

65

<220>
 <223> sintético

<400> 85
 ccacggacgg ttcgtttatc g 21
 5 <210> 86
 <211> 18
 <212> ADN
 <213> Secuencia artificial
 10 <220>
 <223> sintético
 <400> 86
 ccgccttcta cgcgacta 18
 15 <210> 87
 <211> 23
 <212> ADN
 <213> Secuencia artificial
 20 <220>
 <223> sintético
 <400> 87
 gtttaatttt cggtttcgtc gtc 23
 25 <210> 88
 <211> 20
 <212> ADN
 <213> Secuencia artificial
 30 <220>
 <223> sintético
 <400> 88
 cgccgaggcg gtttttgcg 20
 35 <210> 89
 <211> 18
 <212> ADN
 <213> Secuencia artificial
 40 <220>
 <223> sintético
 <400> 89
 cgctacgaaa cactccga 18
 45 <210> 90
 <211> 19
 <212> ADN
 <213> Secuencia artificial
 50 <220>
 <223> sintético
 <400> 90
 gcgttaggt tggtcggag 19
 55 <210> 91
 <211> 20
 <212> ADN
 <213> Secuencia artificial
 60 <220>
 <223> sintético
 65

<400> 91
 cgccgaggcc gaaaaactac 20
 5 <210> 92
 <211> 19
 <212> ADN
 <213> Secuencia artificial
 10 <220>
 <223> sintético

 <400> 92
 acgcactcaa cctacgaac 19
 15 <210> 93
 <211> 19
 <212> ADN
 <213> Secuencia artificial
 20 <220>
 <223> sintético

 <400> 93
 gcgacgtagt cgtcgttgt 19
 25 <210> 94
 <211> 21
 <212> ADN
 <213> Secuencia artificial
 30 <220>
 <223> sintético

 <400> 94
 ccacggacgg aaaacgcgaa a 21
 35 <210> 95
 <211> 19
 <212> ADN
 <213> Secuencia artificial
 40 <220>
 <223> sintético

 <400> 95
 ccaacgcgaa aaaaacgcg 19
 45 <210> 96
 <211> 35
 <212> ADN
 <213> Secuencia artificial
 50 <220>
 <223> sintético

 <400> 96
 gcaataaat aaacgtaat ttgtgttg ttctg 35
 55 <210> 97
 <211> 21
 <212> ADN
 <213> Secuencia artificial
 60 <220>
 <223> sintético
 65

<400> 97
 ccacggacgc gcacttcctt a 21

5

<210> 98
 <211> 24
 <212> ADN
 <213> Secuencia artificial

10

<220>
 <223> sintético

<400> 98
 actttccgaa cgcgtataaa tacc 24

15

<210> 99
 <211> 36
 <212> ADN
 <213> Secuencia artificial

20

<220>
 <223> sintético

<400> 99
 gatcgagggga gcaggagccg cggctgacgg gtcgcg 36

25

<210> 100
 <211> 35
 <212> ADN
 <213> Secuencia artificial

30

<220>
 <223> sintético

<400> 100
 cggtgccgag agctgccaga gagtggatt ctgcg 35

35

<210> 101
 <211> 24
 <212> ADN
 <213> Secuencia artificial

40

<220>
 <223> sintético

45

<400> 101
 gcgagcgggc aaggcgggc gagc 24

50

<210> 102
 <211> 28
 <212> ADN
 <213> Secuencia artificial

55

<220>
 <223> sintético

<400> 102
 acctgcgggc cgaaggctg ctctccgg 28

60

<210> 103
 <211> 29
 <212> ADN
 <213> Secuencia artificial

65

<220>
 <223> sintético

<400> 103
 aggagacgcg ggagcgcggg gtaggtagc 29

5

<210> 104
 <211> 34
 <212> ADN
 <213> Secuencia artificial

10

<220>
 <223> sintético

<400> 104
 ggcatcctcc cggatgatgga agcagccgcc gccg 34

15

<210> 105
 <211> 34
 <212> ADN
 <213> Secuencia artificial

20

<220>
 <223> sintético

<400> 105
 ggaaggcgcg gcgctcggtc agcactgaca gcag 34

25

<210> 106
 <211> 39
 <212> ADN
 <213> Secuencia artificial

30

<220>
 <223> sintético

<400> 106
 tcgccgagct ctcccaaac ttctgcatg ctgaactt 39

35

<210> 107
 <211> 31
 <212> ADN
 <213> Secuencia artificial

40

<220>
 <223> sintético

<400> 107
 ccgtccattt ggccaagtct gcagccgagc c 31

45

<210> 108
 <211> 33
 <212> ADN
 <213> Secuencia artificial

50

<220>
 <223> sintético

<400> 108
 ctgctccgt ccgtcggccg ggcagtctgt cca 33

55

<210> 109
 <211> 48
 <212> ADN
 <213> secuencia artificial

60

<220>
 <223> sintético

65

ES 2 812 753 T3

<400> 109
tccctcgcgc gtggctccg cctctcgcgc ggctggggtg cccggtgg 48

5

<210> 110
<211> 26
<212> ADN
<213> secuencia artificial

10

<220>
<223> sintético

15

<400> 110
gcgggacact ccgaaggcgc aaggag 26

20

25

30

35

40

45

50

55

60

65

REIVINDICACIONES

5 **1.** Un kit que comprende un colector de muestras de heces para obtener una muestra de heces de un sujeto; reactivos para aislar un ácido nucleico de la muestra; un reactivo de bisulfito; y oligonucleótidos que se unen específicamente a una región genética que tiene las coordenadas del cromosoma 3 143119999-143120158.

10 **2.** El kit de la reivindicación 1, en donde los oligonucleótidos son un conjunto de cebadores que consisten en SEQ ID NOS: 69 y 71.

15

20

25

30

35

40

45

50

55

60

65