

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 980 778**

51 Int. Cl.:

C12Q 1/6886 (2008.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **26.03.2015 E 20175912 (3)**

97 Fecha y número de publicación de la concesión europea: **01.05.2024 EP 3839071**

54 Título: **Detección de neoplasias colorrectales**

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:
03.10.2024

73 Titular/es:

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

72 Inventor/es:

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

74 Agente/Representante:

ISERN JARA, Jorge

Observaciones:

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

ES 2 980 778 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 neoplasias colorrectales

5 **Campo de la invención**

En la presente memoria se proporciona tecnología relacionada con la detección de neoplasias y en particular, pero no exclusivamente, con métodos para detectar neoplasias premalignas y malignas, tales como cáncer colorrectal.

10 **Antecedentes**

El cáncer colorrectal sigue siendo el 2º cáncer más común en hombres y mujeres estadounidenses juntos (Siegel R, y col., *CA Cancer J Clin* 2013;63:11-30). La biología subyacente de la progresión desde la lesión precursora hasta el cáncer se presta favorablemente a la detección (Vogelstein B, y col., *Science* 2013;339:1546-58). La evidencia apoya y las pautas respaldan cualquiera de las diversas pruebas y estrategias (Levin B, y col., *Gastroenterology* 2008;134:1570-95; Rex DK y col., *Am J Gastroenterol* 2009;104:739-50; Karl J y col., *Clin Gastroenterol Hepatol* 2008;6:1122-8). Desde una perspectiva social, la detección se considera rentable (Karl J, y col., *Clin Gastroenterol Hepatol* 2008;6:1122-8; Heitman SJ y col., *PLoS Med* 2010;7:e1000370; Parekh M y col., *Aliment Pharmacol Ther* 2008;27:697-712; Sharaf RN y col., *Am J Gastroenterol* 2013;108:120-32).

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

Aunque el cribado reduce la mortalidad por cáncer colorrectal (Mandel JS, y col., *N Engl J Med.* 1993, 328:1365-71; Hardcastle JD y col., *Lancet.* 1996, 348:1472-7; Kronborg O y col., *Scand J Gastroenterol.* 2004, 39:846-51; Winawer SJ y col., *NJ National Cancer Inst.* 1993, 85:1311-8; Singh H y col., *JAMA.* 2006, 295:2366-73), las reducciones observadas han sido modestas (Singh H, y col., *JAMA.* 2006; 295, 2366-73; Heresbach D y col., *Eur J Gastroenterol Hepatol.* 2006, 18:427-33) y más de la mitad de los adultos en los Estados Unidos no se han sometido a pruebas de detección (Meissner HI, *Cancer Epidemiol Biomarkers Prev.* 2006;15:389-94).

Un enfoque emergente para la detección del cáncer implica el análisis de las alteraciones del ADN específicas del tumor en muestras corporales de pacientes con cáncer, tales como las heces, el suero y la orina (Osborn NK, Ahlquist DA, *Gastroenterology* 2005;128:192-206; Ahlquist DA y col., *Gastroenterology* 2000;119:1219-27; Ahlquist DA, y col., *Gastroenterology* 2002;122: Suppl A40; Chen WD y col., *J. Natl Cancer Inst* 2005;97:1124-32; Zou H y col., *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 y col., *Cancer Res* 2006;66:3338-44; Itzkowitz SH y col., *Clin Gastroenterol Hepatol* 2007;5:111-7; Kann L, y col., *Clin Chem* 2006;52:2299-302). Es importante seleccionar marcadores con alta precisión si se quiere lograr la eficiencia y la eficacia en una aplicación de detección del cáncer. Debido a la heterogeneidad molecular de la neoplasia colorrectal, las altas tasas de detección a menudo requieren un panel de marcadores.

Se han detectado varios genes metilados en las muestras de heces y suero/plasma de pacientes con cáncer colorrectal (Ahlquist DA, *Gastroenterology* 2002;122: Suppl A40; Chen WD y col., *J. Natl Cancer Inst* 2005;97:1124-32; Zou HZ y col., *Clin Cancer Res* 2002;8: 188-91; Itzkowitz SH y col., *Clin Gastroenterol Hepatol* 2007;5:111-7; Petko Z y col., *Clin Cancer Res* 2005;11:1203-9; Muller HM y col., *Lancet* 2004;363:1283-5; Leung WK y col., *Clin Chem* 2004;50:2179-82; Ebert MP, y col., *Gastroenterology* 2006;131:1418-30; Grady WM y col., *Cancer Res* 2001;61:900 2/ Si bien 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, y col., *Gastroenterology* 2002;122: Suppl A40; Chen WD y col., *J. Natl Cancer Inst* 2005;97:1124-32; Zou H y col., *Cancer Epidemiol Biomarkers Prev* 2006;15:1115-9; Zou HZ, *Clin Cancer Res* 2002;8:188-91; Belinsky SA y col., *Cancer Res* 2006;66:3338-44; Itzkowitz SH y col., *Clin Gastroenterol Hepatol* 2007;5:111-7; Kann L y col., *Clin Chem* 2006;52:2299-302; Petko Z y col., *Clin Cancer Res* 2005;11:1203-9; Muller HM y col., *Lancet* 2004;363:1283-5; Leung WK y col., *Clin Chem* 2004;50:2179-82; Ebert MP y col., *Gastroenterology* 2006;131:1418-30; Grady WM y col., *Cancer Res* 2001; 61: 900-2).

Se necesitan herramientas más precisas, fáciles de usar y ampliamente distribuibles para mejorar la eficacia, la aceptabilidad y el acceso a las pruebas de detección.

Resumen

5 El ADN metilado se ha estudiado como una posible clase de biomarcadores en los tejidos de la mayoría de los tipos de tumores. En muchos casos, las ADN metiltransferasas agregan un grupo metilo al ADN en los sitios de islas de citosina-fosfato-guanina (CpG) como control epigenético de la expresión génica. En un mecanismo biológicamente atractivo, se cree que los eventos de metilación adquiridos en las regiones promotoras de los genes supresores de tumores silencian la expresión, contribuyendo así a la oncogénesis. La metilación del ADN puede ser una herramienta de diagnóstico química y biológicamente más estable que el ARN o la expresión de proteínas (Laird (2010) Nat Rev Genet 11: 191-203). Además, en cánceres tales 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 y col. (2007) Cancer Epidemiol Biomarkers Prev 16: 2686-96).

15 El análisis de las islas de CpG ha arrojado resultados importantes cuando se aplica a modelos animales y a líneas celulares humanas. Por ejemplo, Zhang y sus colegas descubrieron que los amplicones de diferentes partes de la misma isla de CpG pueden tener diferentes niveles de metilación (Zhang y col. (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 similar a un interruptor de la actividad de la ADN metiltransferasa (Zhang y col. (2009) PLoS Genet 5: e1000438). El análisis de tejidos murinos *in vivo* y de líneas celulares *in vitro* demostró que solo aproximadamente el 0,3 % de los promotores de alta densidad de CpG (HCP, definidos por tener >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 por tener <5 % de secuencia de CpG dentro de una región de 300 pares de bases) tendía a metilarse frecuentemente en un patrón dinámico específico de tejido (Meissner y col. (2008) Nature 454: 766-70). Los HCP incluyen promotores de genes constitutivos ubicuos y genes de desarrollo altamente regulados. Entre los sitios de HCP metilados en >50 % se encontraban varios marcadores establecidos, como Wnt 2, NDRG2, SFRP2, y BMP3 (Meissner y col. (2008) Nature 454: 766-70).

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

35 Zou y sus colegas, por ejemplo, evaluaron los genes frecuentemente metilados en la neoplasia colorrectal para identificar los más discriminantes (Zou, y col., 2007 Cancer Epidemiol Biomarkers Prev. 16(12):2686-2696). Se seleccionaron cuatro genes específicamente metilados en el cáncer colorrectal (la proteína morfogenética ósea 3 (BMP3), la EYA2, la homeobox-4 de tipo aristaless (ALX4) y la vimentina) de entre 41 genes candidatos y se evaluaron en 74 cánceres, 62 adenomas y 70 epitelios normales. El estado de metilación se analizó cualitativa y cuantitativamente y se confirmó mediante secuenciación genómica con bisulfito. El efecto de la metilación sobre 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 el 66 %, 66 %, 68 % y 72 % de los cánceres; 74 %, 48 %, 89 % y 84 % de los adenomas; y 7 %, 5 %, 11 % y 11 % de los 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 raramente en el epitelio normal.

45 La detección del cáncer necesita un marcador o panel de marcadores para el cáncer colorrectal que sea ampliamente informativo y muestre una alta especificidad para el cáncer colorrectal a nivel de tejido cuando se interroga en muestras tomadas de un sujeto (por ejemplo, una muestra de heces; una muestra de tejido colorrectal).

50 En consecuencia, en la presente memoria se describe tecnología para marcadores de detección de cáncer colorrectal que proporcionan una alta relación señal a ruido y un bajo nivel de fondo cuando se detectan a partir de muestras tomadas de un sujeto (por ejemplo, una muestra de heces; una muestra de tejido colorrectal; muestra de suero; sangre o producto sanguíneo).

55 La presente invención se define en las reivindicaciones adjuntas. En particular, la invención proporciona un método de detección para una neoplasia colorrectal en una muestra obtenida de un sujeto, el método comprende:

- 1) analizar un estado de metilación de un marcador en una muestra obtenida de un sujeto; y
- 60 2) identificar que el sujeto tiene una neoplasia cuando el estado de metilación del marcador comprende un aumento de la metilación en relación con el estado de metilación normal de ese marcador,

en donde el marcador comprende una base en una región metilada diferencialmente (DMR) y

65 en donde la región DMR es PPP2R5C.

En experimentos realizados durante el transcurso de las realizaciones de la presente descripción, 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, adenoma y/o pólipos serrados sésiles (SSP) con el estado de metilación de los mismos marcadores de ADN de los sujetos de control (por ejemplo, tejido normal, tal como colon normal) (véanse los ejemplos 1-2, tablas 1-5).

Por ejemplo, marcadores y/o paneles de marcadores (por ejemplo, una región cromosómica que tiene una anotación seleccionada entre 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 SLC8A33) se identificaron en un estudio de casos y controles comparando el estado de metilación de los marcadores de ADN (por ejemplo, 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 de control (por ejemplo, tejido normal tal como colon normal) (véanse el ejemplo 1 y la tabla 1).

Se identificaron marcadores y/o paneles de marcadores (por ejemplo, una región cromosómica con una anotación seleccionada entre BMP3, NDRG4, PDGFG, CHST2 y SFMBT2) en estudios de casos y controles al comparar el estado de metilación de marcadores de ADN de 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 (por ejemplo, tejido normal tal como colon normal) (véanse el ejemplo 1 y la tabla 2).

Además, se identificaron 185, 244 y 111 marcadores de metilación del ADN específicos para cánceres colorrectales, adenomas grandes y pólipos serrados sésiles, respectivamente (véanse el ejemplo 2 y las tablas 3, 4 y 5). Junto con los casos de cáncer colorrectal, los casos de adenomas grandes y los casos de pólipos serrados sésiles, se secuenció la mucosa colónica normal y el ADN normal de los glóbulos blancos.

Otros experimentos realizados durante el transcurso de las realizaciones de la presente descripción demostraron que NDRG4, BMP3, OPLAH, FLI1, PDGFD, CHST2_7889, SFMBT2_895, SFMBT2_896, SFMBT2_897, CHST2_7890, VAV3 y DTX1 son marcadores eficaces para detectar el cáncer colorrectal en muestras de heces (véanse el ejemplo 3 y las tablas 6 y 7).

Como se describe en la presente memoria, la tecnología proporciona una serie de marcadores de ADN metilado y subconjuntos de los mismos (por ejemplo, conjuntos de 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 o más marcadores) con alta discriminación para neoplasias colorrectales (por ejemplo, cáncer colorrectal, adenoma, pólipos SSP). Los experimentos aplicaron un filtro de selección a los marcadores candidatos para identificar los marcadores que proporcionan una alta relación señal a ruido y un bajo nivel de fondo para proporcionar una alta especificidad, por ejemplo, cuando se analizan medios distantes (por ejemplo, heces, sangre, orina, tejido metastásico, etc.) con fines de detección o diagnóstico de cáncer colorrectal. Como tal, la tecnología proporciona marcadores y combinaciones de marcadores específicos para la detección o el diagnóstico del cáncer colorrectal.

En algunas realizaciones, la tecnología está relacionada con la evaluación de la presencia y el estado de metilación de uno o más de los marcadores identificados en la presente memoria en una muestra biológica. Estos marcadores comprenden una o más regiones metiladas diferencialmente (DMR) según se analiza en la presente memoria, 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 en la presente memoria 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 la exploración genómica de puntos de referencia de restricción (Kawai y col. (1994) *Mol. Cell. Biol.* 14: 7421-7427) y otro ejemplo implica la PCR cebada arbitrariamente sensible a la metilación (Gonzalzo y col. (1997) *Cancer Res.* 57: 594-599). En algunas realizaciones, los cambios en los patrones de metilación en sitios de CpG específicos se controlan mediante la digestión del ADN genómico con enzimas de restricción sensibles a la metilación seguido de un análisis de Southern de las regiones de interés (método de digestión-Southern). En algunas realizaciones, el análisis de los cambios en los patrones de metilación implica un proceso basado en la PCR que implica la digestión del ADN genómico con enzimas de restricción sensibles a la metilación antes de la amplificación por PCR (Singer-Sam y col. (1990) *Nucl. Acids Res.* 18: 687). Además, se han notificado otras técnicas que utilizan el tratamiento con bisulfito del ADN como punto de partida para el análisis de metilación. Estas incluyen PCR específica de metilación (MSP) (Herman y col. (1992) *Proc. Natl. Acad. Sci. USA* 93: 9821-9826) y 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 y col. (1991) *Proc. Natl. Acad. Sci. USA* 88: 1143-1147) y cuantificación de la expresión alélica específica (Szabo y Mann (1995) *Genes Dev.* 9: 3097-3108; y Singer-Sam y col. (1992) *PCR Methods Appl.* 1: 160-163). Dichas técnicas utilizan cebadores internos, que se hibridan con un molde generado por PCR y terminan inmediatamente en el extremo 5' del único nucleótido que se va a analizar. Los métodos que utilizan un "ensayo de Ms-SNuPE cuantitativo" como se describe en la pat. US-7.037.650 se usan en algunas realizaciones.

Al evaluar un estado de metilación, el estado de metilación a menudo se expresa como la fracción o el porcentaje de hebras individuales de ADN que están metiladas en un sitio particular (por ejemplo, en un solo nucleótido, en una

región o locus particular, en una secuencia más larga de interés, por ejemplo, hasta una subsecuencia de ~100 pb, 200 pb, 500 pb, 1000 pb de un ADN o más) en relación con la población total de ADN en la muestra que comprende ese sitio en particular. Tradicionalmente, la cantidad de ácido nucleico no metilado se determina mediante PCR utilizando calibradores. A continuación, se trata una cantidad conocida de ADN con bisulfito y la secuencia específica de metilación resultante se determina utilizando una PCR en tiempo real u otra amplificación exponencial, por ejemplo, un ensayo QuARTS (por ejemplo, como se proporciona en la patente US- 8.361.720; y las pub. de solíc. de pat. estadounidense n.º 2012/0122088 y 2012/0122106).

Por ejemplo, en algunas realizaciones, los métodos comprenden generar una curva patrón para la diana no metilada mediante el uso de patrones externos. La curva patrón se construye a partir de al menos dos puntos y relaciona el valor de Ct en tiempo real del ADN no metilado con patrones cuantitativos conocidos. Posteriormente, se construye una segunda curva patrón para la diana metilada a partir de al menos dos puntos y patrones externos. Esta segunda curva patrón relaciona el Ct para el ADN metilado con patrones cuantitativos conocidos. A continuación, se determinan los valores de Ct de la muestra de prueba para las poblaciones metiladas y no metiladas y se calculan los equivalentes genómicos de ADN a partir de las curvas patrón producidas en las dos primeras etapas. 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, (número de ADN metilados) / (número de ADN metilados + número de ADN no metilados) × 100.

Según otro aspecto de la presente invención, la neoplasia de una muestra biológica se indica cuando la relación de metilación de uno o más marcadores de metilación del ADN en relación con un nivel del número de copias de ADN tratado con bisulfito de un gen de referencia es diferente, en donde los uno o más marcadores de metilación del ADN comprenden una base en una región metilada diferencialmente (DMR) como se proporciona en la presente memoria. La relación de metilación incluye la relación entre el 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. Normalmente, la relación de metilación está representada por la relación entre el 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 marcador de metilación del ADN.

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

La región de un gen de referencia puede ser cualquier región de un gen que tenga uno o más sitios o regiones que carezcan de sitios de metilación, por ejemplo, desprovistas de dinucleótidos de CpG. Por ejemplo, la región de un gen de referencia puede ser una región que tiene dos sitios de unión a cebadores para la amplificación, tales como la PCR, que carecen de dinucleótidos de CpG o una región que tiene al menos un sitio de unión a la sonda oligonucleotídica específico para la PCR en tiempo real que carece de dinucleótidos de 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 esté sujeta con frecuencia a alteraciones en el número de copias, tales como la amplificación o delección de genes.

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

En determinadas realizaciones, se describen métodos para detectar neoplasias en un sujeto. Dichos métodos comprenden, por ejemplo, obtener una muestra que comprende ADN de un sujeto; tratar el ADN obtenido con un reactivo que modifica selectivamente 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 del ADN en el ADN que se ha sometido al tratamiento de la etapa b), en donde uno o más marcadores de metilación del ADN comprenden una base en una región metilada diferencialmente (DMR) como se proporciona en la presente memoria; comparar el nivel de metilación determinado de los uno o más marcadores de metilación del ADN con las referencias del nivel de metilación para los uno o más marcadores de metilación del ADN para sujetos que no tienen neoplasia e identificar que el sujeto tiene una 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.

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 en una isla de CpG y una orilla de isla de CpG.

En algunas realizaciones, una determinación de la metilación elevada dentro de dicha isla de CpG u orilla de CpG comprende una metilación elevada dentro de una región codificante o una región reguladora del marcador de metilación del ADN.

5 En algunas realizaciones, determinar el nivel de metilación de uno o más marcadores de metilación del ADN en el ADN que se ha sometido al tratamiento de la etapa b) comprende determinar la puntuación de metilación y/o la frecuencia de metilación del uno o más marcadores de metilación del ADN.

10 En algunas realizaciones, el tratamiento de la etapa b) se logra mediante la modificación con bisulfito del ADN obtenido.

15 En algunas realizaciones, la determinación del nivel de metilación de uno o más marcadores de metilación del ADN en el ADN que se ha sometido al tratamiento de la etapa b) se logra mediante una técnica seleccionada del grupo que consiste en PCR específica de metilación, PCR cuantitativa específica de metilación, análisis de enzimas de restricción de ADN sensible a la metilación, pirosecuenciación con bisulfito cuantitativa y PCR de secuenciación genómica de bisulfito.

En algunas realizaciones, la neoplasia es cáncer colorrectal, un adenoma colorrectal grande y/o un pólipo serrado sésil.

20 En determinadas realizaciones, se proporcionan métodos para detectar neoplasias 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 del ADN en relación con el nivel del número de copias de ADN tratadas con bisulfito de un gen de referencia, en donde los uno o más marcadores de metilación del ADN comprenden una base en una región metilada diferencialmente (DMR) como se proporciona en la presente memoria, en donde el gen de referencia es MYOD o ACTB, que identifica al sujeto como poseedor de una neoplasia cuando la relación de metilación de uno o más de los marcadores de metilación del ADN es diferente a la relación de metilación del marcador respectivo analizado en un sujeto que no tiene una neoplasia.

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

45 También se describen en la presente memoria, pero no se reivindican, composiciones y kits para poner en práctica 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 llevar a cabo un ensayo de detección (por ejemplo, enzimas, tampones, controles positivos y negativos para realizar QuARTS, PCR, secuenciación, bisulfito u otros ensayos). Como se describe en la presente memoria, pero no se reivindica, se proporcionan kits que contienen uno o más reactivos necesarios, suficientes o útiles para llevar a cabo un método. También se proporcionan mezclas de reacción que contienen los reactivos. Además, se proporcionan conjuntos de reactivos de mezcla maestra que contienen una pluralidad de reactivos que se pueden agregar entre sí y/o a una muestra de prueba para completar una mezcla de reacción.

60 En algunas realizaciones, la tecnología descrita en la presente memoria está asociada con una máquina programable diseñada para realizar una secuencia de operaciones aritméticas o lógicas según lo previsto por los métodos descritos en la presente memoria. Por ejemplo, algunas realizaciones de la tecnología están asociados con (por ejemplo, implementados en) software y/o hardware de ordenador. En un aspecto, la tecnología se relaciona con un ordenador que comprende una forma de memoria, un elemento para realizar operaciones aritméticas y lógicas y un elemento de procesamiento (p. ej., un microprocesador) para ejecutar una serie de instrucciones (p. ej., un método como se proporciona en la presente memoria) 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 una o más regiones DMR, por ejemplo como se proporciona en las tablas 1-6); comparar estados de metilación (por ejemplo, de una o más regiones DMR, por ejemplo, como se proporciona en las tablas 1-6); generar curvas patrón; determinar un valor de Ct; calcular una fracción, frecuencia o porcentaje de

metilación (por ejemplo, de una o más regiones DMR, por ejemplo, como se proporciona en las tablas 1-6); identificar una isla de CpG; determinar una especificidad y/o sensibilidad de un ensayo o marcador; calcular una curva ROC y una ABC asociada; análisis de secuencias; todo como se describe en la presente memoria o se conoce en la técnica.

5 En algunas realizaciones, un componente de software o hardware recibe los resultados de múltiples ensayos y determina un resultado de valor único para informar a un usuario que indica un riesgo de cáncer basado en los resultados de múltiples ensayos (por ejemplo, determinar el estado de metilación de múltiples regiones DMR, por ejemplo, como se proporciona en las tablas 1-6). Las realizaciones relacionadas calculan un factor de riesgo basado en una combinación matemática (por ejemplo, una combinación ponderada, una combinación lineal) de los resultados de múltiples ensayos, por ejemplo, determinando los estados de metilación de múltiples marcadores (tales como región DMR múltiple, por ejemplo, como se proporciona en las tablas 1-6). En algunas realizaciones, el estado de metilación de una región DMR define una dimensión y puede tener valores en un espacio multidimensional y la coordenada definida por los estados de metilación de múltiples regiones DMR es un resultado, por ejemplo, para informar a un usuario, por ejemplo, relacionado con un riesgo de cáncer colorrectal.

15 Algunas realizaciones comprenden un medio de almacenamiento y componentes de memoria. Los componentes de memoria (por ejemplo, memoria volátil y/o no volátil) resultan útiles para almacenar instrucciones (por ejemplo, una realización de un proceso como se proporciona en la presente memoria) y/o datos (por ejemplo, una parte de trabajo tales como mediciones de metilación, secuencias y descripciones estadísticas asociadas a las mismas). Algunas realizaciones se relacionan con sistemas que también comprenden una o más de una unidad CPU, una tarjeta gráfica y una interfaz de usuario (por ejemplo, que comprenden un dispositivo de salida, tal como una pantalla, y un dispositivo de entrada, tal como un teclado).

20 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.).

25 En algunas realizaciones, la tecnología comprende un medio de transmisión por cable (por ejemplo, cable metálico, fibra óptica) o inalámbrico para transmitir datos. Por ejemplo, algunas realizaciones se relacionan con 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 dicha red como iguales y en algunas realizaciones las máquinas programables tienen una relación cliente/servidor.

30 En algunas realizaciones, los datos se almacenan en un medio de almacenamiento legible por ordenador, tal como un disco duro, una memoria flash, un medio óptico, un disquete, etc.

35 En algunas realizaciones, la tecnología proporcionada en la presente memoria está asociada a una pluralidad de dispositivos programables que funcionan en conjunto para realizar un método como se describe en la presente memoria. Por ejemplo, en algunas realizaciones, una pluralidad de ordenadores (por ejemplo, conectados por una red) pueden trabajar en paralelo para recopilar y procesar datos, por ejemplo, en una implementación de computación en clúster o computación en red o alguna otra arquitectura de computación distribuida que se base en ordenadores completos (con unidades CPU integradas, almacenamiento, fuentes de alimentación, interfaces de red, etc.) conectados a una red (privada, pública o Internet) mediante una interfaz de red convencional, tal como Ethernet, fibra óptica o mediante tecnología de red inalámbrica.

40 Por ejemplo, algunas realizaciones proporcionan un ordenador que incluye un medio legible por ordenador. La realización incluye una memoria de acceso aleatorio (RAM) acoplada a un procesador. El procesador ejecuta instrucciones de programa ejecutables por ordenador 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 informáticos, como los procesadores de Intel Corporation de Santa Clara, California y Motorola Corporation de Schaumburg, Illinois. Dichos procesadores incluyen, o pueden estar en comunicación con, medios, por ejemplo, medios legibles por ordenador, que almacenan instrucciones que, cuando son ejecutadas por el procesador, hacen que el procesador realice los pasos descritos en la presente memoria.

45 Las realizaciones de los medios legibles por ordenador incluyen, aunque no de forma limitativa, un dispositivo de almacenamiento o transmisión electrónico, óptico, magnético o de otro tipo capaz de proporcionar a un procesador instrucciones legibles por ordenador. Otros ejemplos de medios adecuados incluyen, entre otros, un disquete, un CD-ROM, un DVD, un disco magnético, un chip de memoria, una ROM, una 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 ordenador pueda leer instrucciones. Además, varias formas diferentes de medios legibles por ordenador pueden transmitir o llevar instrucciones a un ordenador, incluidos un enrutador, una red privada o pública u otro dispositivo o canal de transmisión, tanto por cable como inalámbrico. Las instrucciones pueden incluir código en cualquier lenguaje de programación informático adecuado, incluidos, por ejemplo, C, C++, C#, Visual Basic, Java, Python, Perl y JavaScript.

50 En algunas realizaciones, los ordenadores están conectados a una red. Los ordenadores también pueden incluir varios dispositivos externos o internos, como un ratón, un CD-ROM, un DVD, un teclado, una pantalla u otros dispositivos de

5 entrada o salida. Algunos ejemplos de ordenadores son los ordenadores personales, los asistentes digitales, los asistentes digitales personales, los teléfonos celulares, los teléfonos móviles, los teléfonos inteligentes, los buscadores, las tabletas digitales, los ordenadores portátiles, los dispositivos de Internet y otros dispositivos basados en procesadores. En general, los ordenadores relacionados con aspectos de la tecnología proporcionada en la presente memoria pueden ser cualquier tipo de plataforma basada en procesador que opere en cualquier sistema operativo, tal como Microsoft Windows, Linux, UNIX, Mac OS X, etc., capaz de admitir uno o más programas que comprenden la tecnología proporcionada en la presente memoria. Algunas realizaciones comprenden un ordenador 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 Todos los componentes, ordenadores y sistemas descritos en la presente memoria como asociados con la tecnología pueden ser lógicos o virtuales.

15 En consecuencia, en la presente memoria se describe la tecnología relacionada con un método de detección de una neoplasia colorrectal en una muestra (por ejemplo, una muestra de heces, una muestra de tejido colorrectal; muestra de sangre; muestra de producto sanguíneo) obtenida de un sujeto (por ejemplo, un sujeto humano), el método comprende analizar el estado de metilación de un marcador en una muestra obtenida de un sujeto e identificar que el sujeto 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 una región DMR como se proporciona en las tablas 1-6. La tecnología también abarca determinar el estado o el estadio 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 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 una metilación aumentada 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 Además, en algunas realizaciones el marcador es una región de 100 o menos bases, el marcador es una región de 500 o menos bases, el marcador es una región de 1000 o menos bases, 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 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 Además, la tecnología no está limitada en el método utilizado para determinar el estado de metilación. En algunas realizaciones, el análisis comprende el uso de la reacción en cadena de la polimerasa específica de metilación, la secuenciación de ácido nucleico, la espectrometría de masas, la nucleasa específica de metilación, la separación basada en masas o la captura de dianas. En algunas realizaciones, el análisis comprende el uso de un oligonucleótido específico de metilación. En algunas realizaciones, la tecnología utiliza secuenciación paralela masiva (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 perlas, secuenciación de nanoporos, etc.

40 La tecnología describe, pero no reivindica específicamente, reactivos para detectar una región DMR, por ejemplo, en algunas realizaciones se proporciona un conjunto de oligonucleótidos que comprenden las secuencias proporcionadas por las Id. de sec. n.º: 1-110. En algunas realizaciones, se proporciona un oligonucleótido que comprende una secuencia complementaria a una región cromosómica que tiene una base en una región DMR, por ejemplo, un oligonucleótido sensible al estado de metilación de una región DMR.

45 La tecnología proporciona varios 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 que comprende el marcador (véase la 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 (véase la tabla 2). En algunas realizaciones, el marcador comprende una o más de las regiones cromosómicas proporcionadas en la tabla 3 (para el cáncer colorrectal), la tabla 4 (para el adenoma) y la tabla 5 (para los pólipos SSP).

- Además, las realizaciones proporcionan un método para analizar una región DMR de las tablas 1-6. Algunas realizaciones proporcionan la determinación del estado de metilación de un marcador, en donde 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/o una región cromosómica que tiene una anotación que es BMP3, NDRG4, PDGFG, CHST2 y SFMBT2 comprenden el marcador. Algunas realizaciones proporcionan la determinación del estado de metilación de un marcador, en donde una región cromosómica, tal como se proporciona en las tablas 3, 4 y/o 5, comprende el marcador.
- Las realizaciones del kit se describen en la presente memoria, pero no se reivindican, por ejemplo, un kit que comprende un reactivo de bisulfito y un ácido nucleico de control que comprende una secuencia de una región DMR seleccionada de cualquiera de las regiones cromosómicas proporcionadas en las tablas 1-6 y que tiene un estado de metilación asociado a un sujeto que no tiene cáncer. En algunas realizaciones, los kits comprenden un reactivo de bisulfito y un oligonucleótido como se describe en la presente memoria. En algunas realizaciones, los kits comprenden un reactivo de bisulfito y un ácido nucleico de control que comprende una secuencia de una región DMR seleccionada de cualquiera de las regiones cromosómicas proporcionadas en las tablas 1-6 y que tiene un estado de metilación asociado a un sujeto que tiene cáncer colorrectal, adenoma y/o pólipos SSP. Algunas realizaciones del kit comprenden un recolector 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 la presente memoria.
- La tecnología se relaciona con realizaciones de composiciones (por ejemplo, mezclas de reacción), que no se reivindican específicamente. En algunas realizaciones, se proporciona una composición que comprende un ácido nucleico que comprende una región DMR y un reactivo de bisulfito. Algunas realizaciones proporcionan una composición que comprende un ácido nucleico que comprende una región DMR y un oligonucleótido como se describe en la presente memoria. Algunas realizaciones proporcionan una composición que comprende un ácido nucleico que comprende una región DMR y una enzima de restricción sensible a la metilación. Algunas realizaciones proporcionan una composición que comprende un ácido nucleico que comprende una región DMR y una polimerasa.
- Se proporcionan realizaciones de métodos relacionados adicionales para la detección de una neoplasia colorrectal en una muestra obtenida de un sujeto, por ejemplo, un método que comprende determinar el estado de metilación de un marcador en la muestra que comprende una base en una región DMR seleccionada 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 del 90 %, 95 %, 97,5 %, 98 %, 99 %, 99,5 %, 99,9 % o 99,99 % y el valor de p es 0,1, 0,05, 0,025, 0,02, 0,01, 0,005, 0,001, o 0,0001. Algunas realizaciones de métodos proporcionan etapas para hacer reaccionar un ácido nucleico que comprende una DMR con un reactivo de bisulfito para producir un ácido nucleico que ha reaccionado con bisulfito; secuenciar el ácido nucleico que ha reaccionado con bisulfito para proporcionar una secuencia de nucleótidos del ácido nucleico que ha reaccionado con bisulfito; comparar la secuencia de nucleótidos del ácido nucleico que ha 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 que el sujeto tiene una neoplasia colorrectal cuando hay presente una diferencia.
- La tecnología proporciona, pero no reivindica, sistemas para la detección de una neoplasia colorrectal en una muestra obtenida de un sujeto. Las realizaciones ilustrativas de los sistemas incluyen, por ejemplo, un sistema para detectar 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 al cáncer. En algunas realizaciones se determina una alerta mediante un componente de software que recibe los resultados de múltiples ensayos (por ejemplo, determinando los estados de metilación de múltiples marcadores, por ejemplo, una región DMR, por ejemplo, como se proporciona en las tablas 1-6) y calculando un valor o resultado para informar en base a los múltiples resultados. Algunas realizaciones proporcionan una base de datos de parámetros ponderados asociados con cada DMR proporcionada en la presente memoria para su uso en el cálculo de un valor o resultado y/o una alerta para informar a un usuario (por ejemplo, como un médico, profesional de enfermería, profesional sanitario, etc.). En algunas realizaciones, se notifican todos los resultados de múltiples ensayos y, en algunas realizaciones, se utilizan uno o más resultados para proporcionar una puntuación, un valor o un 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.
- 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 recolectar una muestra, como un componente para recolectar 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 describen pero no se reivindican ácidos nucleicos, por ejemplo, un conjunto de ácidos nucleicos, cada ácido nucleico tiene una secuencia que comprende una región 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 de sistemas relacionados comprenden un conjunto de ácidos nucleicos como se describe y una base de datos de secuencias de ácidos nucleicos 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.

La tecnología se relaciona con composiciones (por ejemplo, mezclas de reacción). También se describe, pero no se reivindica, una composición que comprende un ácido nucleico, que comprende una región DMR (por ejemplo, una región 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 región DMR y un oligonucleótido como se describe en la presente memoria. También se describe, pero no se reivindica, una composición que comprende un ácido nucleico que comprende una región DMR y una enzima de restricción sensible a la metilación. También se describe, pero no se reivindica, una composición que comprende un ácido nucleico que comprende una región DMR y una polimerasa.

También se describen, pero no se reivindican, métodos para detectar una neoplasia colorrectal en una muestra obtenida de un sujeto que tiene una enfermedad inflamatoria intestinal, el método comprende 1)

analizar un estado de metilación de un marcador en una muestra obtenida de un sujeto y 2)

identificar que el sujeto 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.

Algunas realizaciones adicionales serán evidentes para los expertos en la técnica relevante basándose en las enseñanzas contenidas en la presente memoria.

Descripción detallada

En la presente memoria se describe la tecnología relacionada con la detección de la neoplasia colorrectal y, en particular, pero no exclusivamente, con los métodos para detectar el cáncer colorrectal premaligno y maligno. Ya que en la presente memoria se describe la tecnología, los encabezados de las secciones se utilizan con fines organizativos y no deben interpretarse de modo alguno como una limitación de la materia descrita.

En esta descripción detallada de las diversas realizaciones, con fines explicativos, se exponen numerosos detalles específicos para proporcionar una comprensión completa de las realizaciones descritas. Un experto en la técnica apreciará, sin embargo, que estas diversas realizaciones pueden practicarse con o sin estos detalles específicos. En otros casos, las estructuras y los 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 se pueden variar.

A menos que se defina de cualquier otra manera, todos los términos técnicos y científicos utilizados en la presente memoria tienen el mismo significado que entiende habitualmente un experto en la materia a la que pertenece la invención. Cuando las definiciones de los términos en las referencias parezcan diferir de las definiciones proporcionadas en las presentes enseñanzas, prevalecerá la definición proporcionada en las presentes enseñanzas.

Definiciones

Para facilitar la comprensión de la presente tecnología, a continuación se definen una serie de términos y expresiones. Se establecen definiciones adicionales a lo largo de la descripción detallada.

A lo largo de la memoria descriptiva y las reivindicaciones, los siguientes términos tienen los significados asociados explícitamente en la presente memoria, a menos que el contexto indique claramente lo contrario. La expresión “en una realización”, tal como se usa en la presente memoria, no se refiere necesariamente a la misma realización, aunque podría hacerlo. Además, la expresión “en otra realización”, tal como se usa en la presente memoria, no se refiere necesariamente a una realización diferente, aunque podría hacerlo.

Además, como se usa en la presente memoria, el término “o” es un operador “o” inclusivo y es equivalente al término “y/o” a menos que el contexto indique claramente lo contrario. La expresión “basado en” no es exclusiva y permite que esté basado en factores adicionales no descritos, a menos que el contexto indique claramente lo contrario. Además, a lo largo de la memoria descriptiva, el significado de “un”, “una” y “el” incluyen referencias en plural. El significado de “en” incluye “en” y “sobre”.

Como se usa en la presente memoria, un “ácido nucleico” o “molécula de ácido nucleico” generalmente se refiere a cualquier ácido ribonucleico o ácido desoxirribonucleico, que puede ser ADN o ARN no modificado o modificado. Los

“ácidos nucleicos” incluyen, sin limitación, ácidos nucleicos monocatenarios y bicatenarios. Como se usa en la presente memoria, la expresión “ácido nucleico” también incluye ADN como se describe anteriormente que contiene una o más bases modificadas. Por lo tanto, el ADN con una cadena principal modificada por motivos de estabilidad o por otras razones es un “ácido nucleico”. La expresión “ácido nucleico”, tal como se utiliza en la presente memoria, abarca tales formas de ácidos nucleicos modificadas química, enzimática o metabólicamente, así como las formas químicas de ADN características de virus y células, incluidas, por ejemplo, células simples y complejas.

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 normalmente más de diez. El tamaño exacto dependerá de muchos factores, que a su vez dependen de la función final o el uso del oligonucleótido. El oligonucleótido puede generarse de cualquier manera, incluida la síntesis química, la replicación del ADN, la transcripción inversa o una combinación de las mismas. Los desoxirribonucleótidos típicos del ADN son la timina, la adenina, la citosina y la guanina. Los ribonucleótidos típicos del ARN son uracilo, adenina, citosina y guanina.

Como se usa en la presente memoria, 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 de CpG, etc.

Los términos “complementario” y “complementariedad” se refieren a nucleótidos (p. ej., 1 nucleótido) o polinucleótidos (p. ej., una secuencia de nucleótidos) relacionados por las reglas de apareamiento de bases. Por ejemplo, la secuencia 5'-A-G-T-3' es complementaria a la secuencia 3'-T-C-A-5'. La complementariedad puede ser “parcial”, en la que solo algunas de las bases de los ácidos nucleicos coinciden según las reglas de emparejamiento de bases. O puede haber una complementariedad “completa” o “total” entre los ácidos nucleicos. El grado de complementariedad entre las cadenas de ácidos nucleicos afecta a la eficacia y a la fuerza de hibridación entre las cadenas de ácidos nucleicos. Esto es de particular importancia en las reacciones de amplificación y en los métodos de detección que dependen de la unión entre ácidos nucleicos.

El término “gen” se refiere a una secuencia de ácido nucleico (por ejemplo, ADN o ARN) que comprende secuencias codificantes necesarias para la producción de un ARN, o de un polipéptido o su precursor. Un polipéptido funcional puede estar codificado por una secuencia codificante de longitud completa o por cualquier parte de la secuencia codificante siempre que se conserven la actividad deseada o las propiedades funcionales (p. ej., actividad enzimática, unión de ligandos, transducción de señales, etc.) del polipéptido. El término “porción” cuando se usa en referencia a un gen se refiere a fragmentos de ese gen. Los fragmentos pueden variar en tamaño desde unos pocos nucleótidos hasta la secuencia completa del gen menos un nucleótido. Por lo tanto, “un nucleótido que comprende al menos una parte de un gen” puede comprender fragmentos del gen o el gen completo.

El término “gen” también abarca las regiones codificantes de un gen estructural e incluye secuencias ubicadas adyacentes a la región codificante en los extremos 5' y 3', por ejemplo, a una distancia de aproximadamente 1 kb en cada extremo, de modo que el gen corresponde a la longitud del ARNm de longitud completa (por ejemplo, que comprende secuencias codificantes, reguladoras, estructurales y otras). Las secuencias que están ubicadas en el extremo 5' de la región codificante y que están presentes en el ARNm se denominan secuencias 5' no traducidas o sin traducir. Las secuencias que están ubicadas en el extremo 3' o cadena abajo de la región codificante y que están presentes en el ARNm se denominan secuencias 3' no traducidas o 3' sin traducir. 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 transcriben en ARN nuclear (ARNnh); los intrones pueden contener elementos reguladores como potenciadores. Los intrones se eliminan o “cortan” del transcrito nuclear o primario; por tanto, los intrones están ausentes en el transcrito de ARN mensajero (ARNm). El ARNm funciona durante la traducción para especificar la secuencia o el orden de los aminoácidos en un polipéptido naciente.

Además de contener intrones, las formas genómicas de un gen también pueden incluir secuencias ubicadas 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 en el extremo 5' o 3' con respecto 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.

La expresión “de tipo silvestre” cuando se emplea en referencia a un gen, se refiere a un gen que tiene las características de un gen aislado de una fuente natural. La expresión “de tipo silvestre” cuando se emplea en referencia a un producto génico, se refiere a un producto génico que tiene las características de un producto génico aislado de una fuente natural. El término “natural”, aplicado a un objeto se refiere al hecho de que un objeto se puede encontrar en la naturaleza. Por ejemplo, una secuencia polipeptídica o polinucleotídica 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 silvestre 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 silvestre” del gen. Por el contrario, el término “modificado” o “mutante” cuando se hace referencia a un gen o a un producto génico se refiere,

respectivamente, a un gen o a un producto génico que presenta modificaciones en la secuencia y/o propiedades funcionales (p. ej., características alteradas) en comparación con el gen o el producto génico de tipo silvestre. Obsérvese que pueden aislarse mutantes de origen natural; estos se identifican por el hecho de que tienen características alteradas en comparación con el gen o el producto génico de tipo silvestre.

El término “alelo” se refiere a una variación de un gen; las variaciones incluyen, pero sin limitación, variantes y mutantes, locus polimórficos y locus polimórficos de un solo nucleótido, desplazamiento del marco de lectura y mutaciones de corte y empalme. Un alelo puede ocurrir naturalmente en una población o puede surgir durante la vida de cualquier individuo particular de la población.

Por tanto, los términos “variante” y “mutante”, cuando se usan en referencia a una secuencia de nucleótidos se refieren a una secuencia de ácido nucleico que difiere en uno o más nucleótidos de otra secuencia de ácido nucleico, generalmente relacionada. Una “variación” es una diferencia entre dos secuencias de nucleótidos diferentes; típicamente, una secuencia es una secuencia de referencia.

La “amplificación” es un caso especial de replicación de ácidos nucleicos que implica especificidad de la cadena molde. Debe contrastarse con la replicación de cadena molde no específica (por ejemplo, la replicación que depende de la cadena molde pero no depende de una cadena molde específica). La especificidad de la cadena molde se distingue aquí de la fidelidad de la replicación (por ejemplo, la síntesis de la secuencia de polinucleótidos adecuada) y la especificidad de los nucleótidos (ribonucleótidos o desoxirribonucleótidos). La especificidad de la cadena molde se describe frecuentemente en términos de especificidad del “objetivo”. Las secuencias diana son “dianas” en el sentido de que se busca separarlas de otros ácidos nucleicos. Las técnicas de amplificación se han diseñado principalmente para esta clasificación.

La amplificación de ácidos nucleicos generalmente se refiere a la producción de múltiples copias de un polinucleótido, o una porción del polinucleótido, típicamente a partir de una pequeña cantidad del polinucleótido (por ejemplo, una sola molécula de polinucleótido, de 10 a 100 copias de una molécula de polinucleótido, que pueden o no ser exactamente iguales), donde los productos de amplificación o amplicones son generalmente detectables. La amplificación de polinucleótidos abarca diversos procesos químicos y enzimáticos. La generación de múltiples copias de ADN a partir de una o unas pocas copias de una molécula o cadena molde de ADN diana durante una reacción en cadena de la polimerasa (PCR) o una reacción en cadena de la ligasa (LCR); véase, por ejemplo, la patente US-5.494.810 son formas de amplificación. Los tipos adicionales de amplificación incluyen, aunque no de forma limitativa, PCR específica de alelo (patente US-5.639.611), PCR de ensamblaje (patente US-5.965.408), amplificación dependiente de helicasa (patente US-7.662.594), PCR de inicio en caliente (patentes US-5.773.258 y 5.338.671), PCR intersecuencial específica, PCR inversa (Triglia y col. (1988) *Nucleic Acids Research*, 16:8186), PCR mediada por ligamiento (Guilfoyle, R. y col. *Nucleic Acids Research*, 25:1854-1858 (1997); patente US-5.508.169), PCR específica de metilación (Herman y col., (1996) *PNAS* 93(13) 9821-9826), PCR de minicebador, amplificación de sonda dependiente de ligadura multiplex (Schouten y col., (2002) *Nucleic Acids Research* 30(12): e57), PCR multiplex (Chamberlain y col., (1988) *Nucleic Acids Research* 16(23) 11141-11156; Ballabio y col., (1990) *Human Genetics* 84(6) 571-573; Hayden y col., (2008) *BMC Genetics* 9:80), PCR anidada, PCR de superposición-extensión (Higuchi y col., (1988) *Nucleic Acids Research* 16(15) 7351-7367), PCR en tiempo real (Higuchi y col., (1992) *Biotechnology* 10:413-417; Higuchi y col., (1993) *Biotechnology* 11:1026-1030), PCR de transcripción inversa (Bustin, SA (2000) *J. Molecular Endocrinology* 25:169-193), PCR en fase sólida, PCR termoasimétrica entrelazada y PCR Touchdown (Don, y col., *Nucleic Acids Research* (1991) 19(14) 4008; Roux, K. (1994) *Biotechniques* 16(5) 812-814; Hecker, y col., (1996) *Biotechniques* 20(3) 478-485). La amplificación de polinucleótidos también puede lograrse usando PCR digital (Kalinina y col., *Nucleic Acids Research*. 25; 1999-2004, (1997); Vogelstein y Kinzler, *Proc Natl Acad Sci USA*. 96; 9236-41, (1999); publicación de patente internacional n.º WO05023091A2; publicación de solicitud de patente US-20070202525).

Como se usa en la presente memoria, la expresión “ensayo de detección de ácido nucleico” se refiere a cualquier método para determinar la composición de nucleótidos de un ácido nucleico de interés. Los ensayos de detección de ácidos nucleicos incluyen, aunque no de forma limitativa, métodos de secuenciación de ADN, métodos de hibridación de sondas, ensayos de escisión específica de estructura (por ejemplo, el ensayo INVADER, Hologic, Inc.) y se describen, por ejemplo, en las patentes US-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 y col, *PNAS, USA*, 97:8272 (2000), y el documento US-2009/0253142); reacción en cadena de la polimerasa; métodos de hibridación ramificados (por ejemplo, Chiron, patentes US-5.849.481, 5.710.264, 5.124.246, y 5.624.802); replicación en círculo rodante (por ejemplo, las patentes US-6.210.884, 6.183.960 y 6.235.502); NASBA (por ejemplo, la patente US- 5.409.818); tecnología de baliza molecular (por ejemplo, la patente US- 6.150.097); tecnología de sonda cíclica (por ejemplo, las patentes US- 5.403.711, 5.011.769, y 5.660.988); reacción en cadena de la ligasa (por ejemplo, Barnay *Proc. Natl. Acad. Sci USA* 88, 189-93 (1991)); Ensayo QuARTS (por ejemplo, según lo proporcionado por la patente US- 8.361.720; y las pub. de solíc. de pat. US-2012/0122088 y 2012/0122106); y métodos de hibridación en sándwich (por ejemplo, la patente US- 5.288.609).

El término “cebador” se refiere a un oligonucleótido, ya sea de forma natural como en un digerido de restricción purificado o producido sintéticamente, que es capaz de actuar como un punto de iniciación de la síntesis cuando se coloca en condiciones en las que la síntesis de un producto de extensión del cebador que es complementario a una hebra de ácido nucleico (p. ej., en presencia de nucleótidos y un agente inductor tal como una polimerasa de ADN ya una temperatura y un pH adecuados). El cebador es preferiblemente monocatenario para máxima eficiencia en la amplificación, pero como alternativa puede ser bicatenario. Si es bicatenario, el cebador se trata en primer lugar

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 como para cebar la síntesis de los 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.

El término “sonda” se refiere a un oligonucleótido (p. ej., una secuencia de nucleótidos), ya sea de forma natural como en una digestión de restricción purificada o producida de forma sintética, recombinante o mediante amplificación por PCR, que es capaz de hibridarse con otro oligonucleótido de interés. Una sonda puede ser monocatenaria o bicatenaria. Las sondas son útiles en la detección, identificación y aislamiento de secuencias de genes particulares (p. ej., una “sonda de captura”). Se contempla que cualquier sonda utilizada en la presente invención pueda, en algunas realizaciones, marcarse con cualquier “molécula indicadora”, de modo que sea detectable en cualquier sistema de detección, incluidos, entre otros, enzimas (p. ej., ELISA, así como ensayos histoquímicos basados en enzimas), sistemas fluorescentes, radiactivos y luminiscentes. No se pretende que la presente descripción se limite a ningún sistema de detección o etiqueta en particular.

Como se usa en la presente memoria, “metilación” se refiere a la metilación de citosina en las posiciones C5 o N4 de la citosina, la posición N6 de la adenina u otros tipos de metilación de ácidos nucleicos. El ADN amplificado *in vitro* generalmente no está metilado porque los métodos típicos de amplificación de ADN *in vitro* no conservan el patrón de metilación de la cadena molde de amplificación. Sin embargo, “ADN metilado” o “ADN no metilado” también pueden referirse a ADN amplificado cuya cadena molde original está metilada o no metilada, respectivamente.

Por consiguiente, como se usa en la presente memoria, un “nucleótido metilado” o una “base de nucleótido metilada” se refiere a la presencia de un resto de metilo en una base de nucleótido, en donde el resto de metilo no está presente en una base de nucleótido típica reconocida. Por ejemplo, la citosina no contiene un resto de metilo en su anillo de pirimidina, pero la 5-metilcitosina contiene un resto de 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 la presente memoria, la timina no se considera un nucleótido metilado cuando está presente en el ADN, ya que la timina es una base nucleotídica típica del ADN.

Como se usa en la presente memoria, una “molécula de ácido nucleico metilado” se refiere a una molécula de ácido nucleico que contiene uno o más nucleótidos metilados.

Como se usa en la presente memoria, un “estado de metilación”, “perfil de metilación” y “estatus de metilación” de una molécula de ácido nucleico se refiere a la presencia o ausencia de una o más bases de nucleótido metiladas en la molécula de ácido nucleico. Por ejemplo, una molécula de ácido nucleico que contiene una citosina metilada se considera metilada (p. ej., el estado de metilación de la molécula de ácido nucleico está metilado). Una molécula de ácido nucleico que no contiene nucleótidos metilados se considera no metilada.

El estado de metilación de una secuencia de ácido nucleico en particular (p. ej., un marcador genético o región de ADN como se describe en la presente memoria) puede indicar el estado de metilación de cada base en la secuencia o puede indicar el estado de metilación de un subconjunto de bases (p. ej., de una o más citosinas) dentro de la secuencia, o puede indicar información sobre la densidad de metilación regional dentro de la secuencia proporcionando o sin proporcionar información precisa de las ubicaciones dentro de la secuencia en la que se produce la metilación.

El estado de metilación de un locus de nucleótido 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 7º nucleótido de una molécula de ácido nucleico está metilado cuando el nucleótido presente en el 7º nucleótido de la molécula de ácido nucleico es 5-metilcitosina. De manera similar, el estado de metilación de una citosina en el 7º nucleótido de una molécula de ácido nucleico no está metilado cuando el nucleótido presente en el 7º nucleótido de la molécula de ácido nucleico es citosina (y no 5-metilcitosina).

El estado de metilación se puede representar o indicar opcionalmente mediante un “valor de metilación” (p. ej., que representa una frecuencia, fracción, proporción, porcentaje, etc.). Se puede generar un valor de metilación, por ejemplo, cuantificando la cantidad de ácido nucleico intacto presente después de la digestión de restricción con una enzima de restricción dependiente de la metilación o comparando los perfiles de amplificación después de la reacción con bisulfito o comparando las secuencias de ácidos nucleicos tratados con bisulfito y no tratados. Por consiguiente, 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 particularmente útil cuando se desea comparar el estado de metilación de una secuencia en una muestra con un umbral o valor de referencia.

Como se usa en la presente memoria, “frecuencia de metilación” o “porcentaje de metilación (%)” se refiere al número de casos en los que una molécula o locus está metilado en relación con el número de casos en que la molécula o el locus no está metilado.

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, entre otras, si alguno de los restos de citosina (C) dentro de esta secuencia de ADN está metilado, la ubicación de los restos de C metilados, la frecuencia o el porcentaje de C metilado en cualquier región particular de un ácido nucleico y diferencias alélicas en la metilación debidas, por ejemplo, a diferencias 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 los restos de citosina (C) dentro de una secuencia de ácido nucleico están metilados, se puede decir que están “hipermetilados” o que tienen “metilación aumentada”, mientras que si los restos de citosina (C) dentro de una secuencia de ADN no están metilados, se puede denominar “hipometilada” o que tiene “metilación disminuida”. Asimismo, si los restos de citosina (C) dentro de una secuencia de ácido nucleico están metilados en comparación con otra secuencia de ácido nucleico (p. ej., de una región diferente o de un individuo diferente, etc.), esa secuencia se considera hipermetilada o con metilación aumentada en comparación con la otra secuencia de ácido nucleico. Como alternativa, si los restos de citosina (C) dentro de una secuencia de ADN no están metilados en comparación con otra secuencia de ácido nucleico (p. ej., de una región diferente o de un individuo diferente, etc.), esa secuencia se considera hipometilada o con metilación disminuida en comparación con la otra secuencia de ácido nucleico. Además, la expresión “patrón de metilación”, como se usa en la presente memoria, 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 o similar frecuencia de metilación o porcentaje de metilación pero tener diferentes patrones de metilación cuando el número de nucleótidos metilados y no metilados es el mismo 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 el grado (p. ej., una tiene una metilación aumentada o disminuida en relación con la otra), la frecuencia o el patrón de metilación. La expresión “metilación diferencial” se refiere a una diferencia en el nivel o patrón de metilación de ácidos nucleicos en una muestra positiva para cáncer en comparación con el nivel o patrón de metilación de ácidos nucleicos en una muestra negativa para cáncer. También puede referirse a la diferencia en niveles o patrones entre pacientes que tienen recurrencia del cáncer después de la cirugía frente a 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.

La frecuencia del estado de metilación se puede utilizar para describir una población de individuos o una muestra de un solo individuo. Por ejemplo, un locus de nucleótido que tiene una frecuencia de estado de metilación del 50 % está metilado en el 50 % de los casos y no metilado en el 50 % de los casos. Dicha frecuencia se puede utilizar, por ejemplo, para describir el grado en que un locus de nucleótido o región de ácido nucleico está metilado en una población de individuos o una colección de ácidos nucleicos. Por lo tanto, cuando la metilación en una primera población o conjunto de moléculas de ácido nucleico es diferente de la metilación en una segunda población o conjunto de moléculas de ácido nucleico, la frecuencia del estado de metilación de la primera población o conjunto será diferente de la frecuencia del estado de metilación del segundo. población o conjunto. Dicha frecuencia también puede utilizarse, por ejemplo, para describir el grado en que un locus de nucleótido o región de ácido nucleico está metilado en un solo individuo. Por ejemplo, dicha frecuencia se puede utilizar para describir el grado en que un grupo de células de una muestra de tejido está metilado o no metilado en un locus de nucleótido o región de ácido nucleico.

Como se usa en la presente memoria, un “locus de nucleótidos” se refiere a la ubicación de un nucleótido en una molécula de ácido nucleico. Un locus de nucleótido de un nucleótido metilado se refiere a la ubicación de un nucleótido metilado en una molécula de ácido nucleico.

Típicamente, la metilación del ADN humano se produce en una secuencia de dinucleótidos que incluye una guanina y una citosina adyacentes, donde la citosina está ubicada en el extremo 5' de la guanina (también denominadas secuencias de dinucleótidos de CpG). La mayoría de las citosinas dentro de los dinucleótidos de CpG están metiladas en el genoma humano, sin embargo, algunas permanecen sin metilar en regiones genómicas ricas en dinucleótidos de CpG específicas, conocidas como islas de CpG (véase, por ejemplo, Antequera y col. (1990) Cell 62: 503-514).

Como se usa en la presente memoria, una “isla de CpG” se refiere a una región de ADN genómico rica en G:C que contiene un número aumentado de dinucleótidos de CpG en relación con el ADN genómico total. Una isla de 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 el 50 % y la relación entre la frecuencia de CpG observada y la frecuencia esperada es 0,6; en algunos casos, una isla de CpG puede tener al menos 500 pares de bases de longitud, donde el contenido G:C de la región es al menos el 55 % y la relación entre la frecuencia de CpG observada y la frecuencia esperada es 0,65. La frecuencia de CpG observada sobre la frecuencia esperada puede calcularse según el método proporcionado en Gardiner-Garden y col. (1987) J. Mol. Biol. 196: 261-281. Por ejemplo, la frecuencia de CpG observada sobre la frecuencia esperada se puede calcular según la fórmula $R = (A \times B) / (C \times D)$, donde R es la relación entre la frecuencia de CpG observada y la frecuencia esperada, A es el número de dinucleótidos de CpG 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 típicamente se determina en las islas de CpG, por ejemplo, en las regiones promotoras. Sin embargo, se apreciará que otras secuencias en el genoma humano son propensas a la metilación del ADN, tales como CpA y CpT (véase Ramsahoye (2000) Proc. Natl. Acad. Sci. USA 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).

5 Como se usa en la presente memoria, 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 manera que refleje 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 etapas adicionales, si se desea, para lograr el cambio deseado de secuencia de nucleótidos. Dicho 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 en un nucleótido diferente. Dicho cambio en la secuencia de nucleótidos del ácido nucleico puede dar como resultado una molécula de ácido nucleico en la que cada nucleótido no metilado se modifica en un nucleótido diferente. Dicho cambio en la secuencia de nucleótidos del ácido nucleico puede dar como resultado una molécula de ácido nucleico en la que cada uno de los nucleótidos seleccionados que no está metilado (p. ej., cada citosina no metilada) se modifica en un nucleótido diferente. El uso de dicho reactivo para cambiar la secuencia de nucleótidos del ácido nucleico puede dar como resultado una molécula de ácido nucleico en la que cada nucleótido que es un nucleótido metilado (p. ej., cada citosina metilada) se modifica en un nucleótido diferente. Como se usa en la presente memoria, 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 modifica un nucleótido sin modificar los otros tres nucleótidos. En una realización ilustrativa, dicho reactivo modifica un nucleótido seleccionado no metilado para producir un nucleótido diferente. En otra realización ilustrativa, dicho reactivo puede desaminar nucleótidos de citosina no metilados. Un ejemplo de reactivo es bisulfito.

25 Como se usa en la presente memoria, la expresión “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 de CpG.

30 La expresión “ensayo de metilación” se refiere a cualquier ensayo para determinar el estado de metilación de una o más secuencias de dinucleótidos de CpG dentro de una secuencia de un ácido nucleico.

35 El término “MS AP-PCR” (reacción en cadena de la polimerasa cebada 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 enfocarse en las regiones con mayor probabilidad de contener dinucleótidos de CpG, y descrita por Gonzalzo y col. (1997) Cancer Research 57: 594-599.

El término “MethyLight™” se refiere a la técnica de PCR en tiempo real basada en fluorescencia reconocida en la técnica descrita por Eads y col. (1999) Cancer Res. 59: 2302-2306.

40 El término “HeavyMethyl™” se refiere a un ensayo en donde las sondas de bloqueo específicas para la metilación (también denominadas en la presente memoria bloqueantes) 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.

45 El término ensayo “HeavyMethyl™ MethyLight™” se refiere a un ensayo HeavyMethyl™ MethyLight™, que es una variación del ensayo MethyLight™, en donde el ensayo MethyLight™ se combina con sondas de bloqueo específicas de metilación que cubren las posiciones de CpG entre los cebadores de amplificación.

El término “Ms-SNuPE” (extensión del cebador de un solo nucleótido sensible a la metilación) se refiere al ensayo reconocido en la técnica descrito por Gonzalzo y Jones (1997) Nucleic Acids Res. 25: 2529-2531.

50 El término “MSP” (PCR específica de metilación) se refiere al ensayo de metilación reconocido en la técnica descrito por Herman y col. (1996) Proc. Natl. Acad. Sci. USA 93: 9821-9826 y en la patente US- 5.786.146.

55 El término “COBRA” (análisis combinado de restricción de bisulfito) se refiere al ensayo de metilación reconocido en la técnica descrito por Xiong y Laird (1997) Nucleic Acids Res. 25: 2532-2534.

El término “MCA” (amplificación de las islas de CpG metiladas) se refiere al ensayo de metilación descrito por Toyota y col. (1999) Cancer Res. 59: 2307-12 y en el documento WO 00/26401A1.

60 Como se usa en la presente memoria, un “nucleótido seleccionado” se refiere a un nucleótido de los cuatro nucleótidos típicos 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 típicos (p. ej., cuando C es el nucleótido seleccionado, tanto C metilado como no metilado se incluyen en el significado de un nucleótido seleccionado), mientras que un nucleótido seleccionado metilado se refiere específicamente a un nucleótido típico metilado y un nucleótido no metilado se refiere específicamente a un nucleótido de aparición típica no metilado.

65

Las expresiones “enzima de restricción específica de metilación” o “enzima de restricción sensible a la metilación” se refieren a una enzima que digiere selectivamente un ácido nucleico dependiendo 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 lugar con una eficiencia 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 lugar con una eficiencia significativamente reducida si el sitio de reconocimiento no está metilado. Se prefieren enzimas de restricción específicas de metilación, cuya secuencia de reconocimiento contiene un dinucleótido de CG (por ejemplo, una secuencia de reconocimiento tal como CGCG o CCCGGG). Además, para algunas realizaciones se prefieren enzimas de restricción que no cortan si la citosina en este dinucleótido está metilada en el átomo de carbono C5.

Tal como se usa en la presente memoria, un “nucleótido diferente” se refiere a un nucleótido que es químicamente diferente de un nucleótido seleccionado, típicamente de tal forma que el nucleótido diferente tiene propiedades de emparejamiento de bases de Watson-Crick que difieren del nucleótido seleccionado, por lo que el nucleótido que se presenta típicamente que es complementario al nucleótido seleccionado no es lo mismo que el nucleótido típico que es complementario al nucleótido diferente. Por ejemplo, cuando C es el nucleótido seleccionado, U o T pueden ser el nucleótido diferente, lo que se ilustra por la complementariedad de C con G y la complementariedad de U o T con A. Como se usa en la presente memoria, un nucleótido que es complementario al nucleótido seleccionado o que es complementario al nucleótido diferente se refiere a un nucleótido que forma pares de bases, en condiciones de alta rigurosidad, con el nucleótido seleccionado o con un nucleótido diferente con mayor afinidad que la unión de bases del nucleótido complementario con tres de los cuatro nucleótidos típicos. Un ejemplo de complementariedad es el apareamiento de bases de Watson-Crick en el ADN (p. ej., A-T y C-G) y el ARN (p. ej., A-U y C-G). Por tanto, por ejemplo, G forma un emparejamiento de bases, en condiciones de alta rigurosidad, con mayor afinidad a C que el emparejamiento de bases de G con G, A o T y por tanto, cuando C es el nucleótido seleccionado, G es un nucleótido complementario al nucleótido seleccionado.

Como se usa en la presente memoria, la “sensibilidad” de un marcador dado se refiere al porcentaje de muestras que informan un valor de metilación del 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 intervalo asociado con la enfermedad), y un falso negativo se define como una neoplasia confirmada por histología que informa un valor de metilación del ADN por debajo del valor umbral (por ejemplo, el intervalo asociado con la ausencia de enfermedad). El valor de la sensibilidad, por lo tanto, refleja la probabilidad de que una medida de metilación del ADN para un marcador dado obtenido de una muestra de enfermedad conocida esté en el intervalo de medidas asociadas a la enfermedad. Como se define en la presente memoria, la relevancia clínica del valor de sensibilidad calculado representa una estimación de la probabilidad de que un marcador dado pueda detectar la presencia de una afección clínica cuando se aplica a un sujeto con dicha afección.

Como se usa en la presente memoria, 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 (p. ej., el intervalo asociado con ausencia de enfermedad) y un falso positivo se define como una muestra no neoplásica confirmada por histología que informa un valor de metilación del ADN por encima del valor umbral (p. ej., el intervalo asociado con la enfermedad). El valor de especificidad, por lo tanto, refleja la probabilidad de que una medida de metilación del ADN para un marcador dado obtenido de una muestra no neoplásica conocida esté en el intervalo de medidas no asociadas a enfermedades. Como se define en la presente memoria, la relevancia clínica del valor de especificidad calculado representa una estimación de la probabilidad de que un marcador determinado pueda detectar la ausencia de una afección clínica cuando se aplica a un paciente sin dicha afección.

El término “ABC”, como se usa en la presente memoria, es una abreviatura de “área bajo la curva”. Concretamente, se refiere al área bajo una curva de características operativas del receptor (ROC). La curva ROC es una gráfica de la tasa de verdaderos positivos frente a la tasa de falsos positivos para los diferentes puntos de corte posibles de una prueba de diagnóstico. Muestra el compromiso entre sensibilidad y especificidad según el punto de corte seleccionado (cualquier aumento de la sensibilidad irá acompañado de una disminución de la especificidad). El área bajo una curva ROC (ABC) es una medida de la precisión de una prueba de diagnóstico (cuanto mayor sea el área, mejor; el valor óptimo es 1; una prueba aleatoria tendría una curva ROC en diagonal con un área de 0,5; por referencia: J. P. Egan. (1975) *Signal Detection Theory and ROC Analysis*, Academic Press, Nueva York).

Como se usa en la presente memoria, el término “neoplasia” se refiere a “una masa anormal de tejido, cuyo crecimiento excede y no se coordina con el de los tejidos normales”. Véase, por ejemplo, Willis RA, “*The Spread of Tumors in the Human Body*”, Londres, Butterworth & Co, 1952.

Como se usa en la presente memoria, el término “adenoma” se refiere a un tumor benigno de origen glandular. Aunque estos crecimientos son benignos, con el tiempo pueden progresar y volverse malignos.

El término “precanceroso” o “preneoplásico” y sus equivalentes se refieren a cualquier trastorno proliferativo celular que esté experimentando una transformación maligna.

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 localiza la neoplasia, adenoma, cáncer, etc.

5 Como se usa en la presente memoria, una aplicación de prueba de “diagnóstico” incluye la detectar o identificar una patología o afección en un sujeto, determinar la probabilidad de que un sujeto contraerá una determinada enfermedad o afección, determinar la probabilidad de que un sujeto con una enfermedad o afección responderá a la terapia, determinar el pronóstico de un sujeto con una enfermedad o afección (o su probable progresión o regresión), y determinar 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 (p. ej., un producto farmacéutico, p. ej., un fármaco) u otro tratamiento.

10 El término “marcador”, como se usa en la presente memoria, se refiere a una sustancia (por ejemplo, un ácido nucleico o una región de un ácido nucleico) que es capaz de diagnosticar un cáncer al distinguir las células cancerosas de las células normales, por ejemplo, en función de su estado de metilación.

15 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 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 configuración que es diferente de aquella en la 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 (p. ej., un gen) que se encuentra en el cromosoma de la célula hospedadora en la proximidad de los genes vecinos; las secuencias de ARN, como una secuencia de ARNm específica que codifica una proteína específica, se encuentran en la célula como una mezcla con muchos 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 células que normalmente expresan la proteína, donde el ácido nucleico se encuentra en una ubicación cromosómica diferente a la de las células naturales, o está flanqueado por una secuencia de ácido nucleico diferente que la que se encuentra 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 codificante o con sentido (es decir, el oligonucleótido puede ser monocatenario), pero puede contener las cadenas con sentido y las antisentido (es decir, el oligonucleótido puede ser bicatenario). Un ácido nucleico aislado puede, después del aislamiento de su entorno natural o típico, combinarse con otros ácidos nucleicos o moléculas. Por ejemplo, un ácido nucleico aislado puede estar presente en una célula hospedadora en la que se ha colocado, por ejemplo, para la expresión heteróloga.

20 El término “purificado” se refiere a moléculas, ya sean secuencias de ácidos nucleicos o de aminoácidos que se extraen, se aíslan o se separan de su entorno natural. Por lo tanto, una “secuencia de ácido nucleico aislada” puede ser una secuencia de ácido nucleico purificada. Las moléculas “sustancialmente purificadas” están libres al menos en un 60 %, preferiblemente al menos en un 75 % y más preferiblemente en al menos un 90 % de otros componentes con los que están asociadas de forma natural. Como se usa en la presente memoria, los términos “purificado” o “para 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 hospedadoras de plantas, bacterias, levaduras o mamíferos y los polipéptidos se purifican mediante la eliminación de proteínas de células hospedadoras; el porcentaje de polipéptidos recombinantes aumenta de ese modo en la muestra.

25 La expresión “composición que comprende” una secuencia polinucleotídica o polipéptido dado se refiere, en términos generales, a cualquier composición que contenga la secuencia polinucleotídica o el polipéptido dado. La composición puede comprender una solución acuosa que contenga 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.).

30 El término “muestra” se utiliza en su sentido más amplio. En cierto sentido, puede referirse a una célula o tejido animal. En otro sentido, se entiende que incluye un espécimen 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.

35 Como se usa en la presente memoria, 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 la 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 del tejido colorrectal), la muestra es una muestra remota.

40 Como se usa en la presente memoria, los términos “paciente” o “sujeto” se refieren a organismos que se someterán a diversas pruebas proporcionadas por la tecnología. El término “sujeto” incluye animales, preferiblemente mamíferos, incluidos los seres humanos. En una realización preferida, el sujeto es un primate. En una realización aún más preferida, el sujeto es un ser humano.

Como se usa en la presente memoria, el término “kit” se refiere a cualquier sistema de administración para administrar materiales. En el contexto de los ensayos de reacción, dichos sistemas de administración incluyen sistemas que permiten el almacenamiento, transporte o administración de reactivos de reacción (p. ej., oligonucleótidos, enzimas, etc. en los contenedores apropiados) y/o materiales de apoyo (p. ej., tampones, instrucciones 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 relevantes y/o los materiales de apoyo. Como se usa en la presente memoria, el término “kit fragmentado” se refiere a sistemas de administración que comprenden dos o más contenedores separados, cada uno de los cuales contiene una subporción de los componentes totales del kit. Los contenedores se pueden entregar al destinatario previsto juntos o por separado. Por ejemplo, un primer contenedor puede contener una enzima para usar en un ensayo, mientras que un segundo contenedor contiene oligonucleótidos. El término “kit fragmentado” pretende abarcar kits que contienen reactivos específicos de analitos (ASR) regulados según la sección 520(e) de la Ley Federal de Alimentos, Medicamentos y Cosméticos, pero no se limita los mismos. De hecho, cualquier sistema de administración que comprenda dos o más contenedores separados, cada uno de los cuales contiene una subporción de los componentes totales del kit, se incluye en la expresión “kit fragmentado”. Por el contrario, un “kit combinado” se refiere a un sistema de administración que contiene todos los componentes de un ensayo de reacción en un solo contenedor (p. ej., en una sola caja que contiene cada uno de los componentes deseados). El término “kit” incluye tanto kits fragmentados como combinados.

20 Realizaciones de la tecnología

En conjunto, los cánceres gastrointestinales son responsables de más mortalidad por cáncer que cualquier otro sistema de órganos. El cáncer colorrectal (CCR) es el segundo cáncer más mortal a nivel nacional, con > 600 000 muertes al año. Si bien los cánceres colorrectales se examinan en los Estados Unidos, el cumplimiento es deficiente debido al costo, las molestias y la invasividad de la colonoscopia y al pésimo rendimiento de la gama actual de análisis de sangre fecal no invasivos. Para reducir la carga que representa el CCR para las personas y la sociedad, se necesitan nuevas estrategias de prueba que sean eficaces y fáciles de usar para los pacientes. Una prueba molecular precisa y no invasiva que utilice biomarcadores ampliamente informativos puede proporcionar un enfoque racional. Un ensayo basado en heces es un ejemplo de ello. Los cánceres colorrectales y los precánceres expulsan células y ADN al torrente digestivo y, en última instancia, se excretan en las heces. Se han utilizado ensayos de alta sensibilidad para detectar marcadores genéticos y epigenéticos en las heces de pacientes con cáncer CCR y pólipos precancerosos. En un estudio multicéntrico reciente, estos marcadores se incorporaron a un análisis de heces que mostró una sensibilidad para el CCR esencialmente equivalente a la de la colonoscopia. Los nuevos marcadores serían de mayor valor si se demostrara que son más sensibles, más específicos o más predictivos del sitio de la lesión que los marcadores existentes. Además, los marcadores detectarían idealmente las lesiones precancerosas críticas, pólipos adenomatosos y pólipos serrados, además del CCR cuando se aplican a una aplicación de detección.

Los mecanismos genómicos subyacentes al cáncer colorrectal implican anomalías genéticas y cromosómicas, incluidas las mutaciones de una sola base, la aneuploidía, las deleciones, las translocaciones, las alteraciones del número de copias y los cambios en la expresión. Todos estos eventos se están estudiando intensamente utilizando tecnologías genómicas más nuevas, incluida la secuenciación masiva en paralelo. Sin embargo, las alteraciones genéticas están demostrando ser muy heterogéneas y, en algunos aspectos, aleatorias, más que recurrentes. El gen APC, por ejemplo, es el gen más mutado en las lesiones del CCR (~ 90 %), pero los sitios de mutación están repartidos en 15 exones codificantes que requieren un análisis de todo el gen. Esto añade complejidad al desarrollo de ensayos con los niveles de rendimiento requeridos.

La metilación epigenética del ADN en los sitios de las islas de citosina-fosfato-guanina (CpG) por parte de las 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 las regiones promotoras de los genes supresores de tumores silencian la expresión, lo que contribuye a la oncogénesis. La metilación del ADN puede ser una herramienta de diagnóstico química y biológicamente más estable que el ARN o la expresión de proteínas. Además, los marcadores de metilación aberrantes son más informativos y sensibles que las mutaciones individuales del ADN y ofrecen una especificidad excelente.

Las aplicaciones clínicas de los marcadores altamente discriminantes podrían tener un gran impacto. Por ejemplo, el ensayo de dichos marcadores en medios distantes, como las heces o la sangre, resulta útil en ensayos de detección o diagnóstico precisos para la detección de la neoplasia colorrectal.

En experimentos realizados durante el transcurso de las realizaciones de la presente descripción, 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, adenoma y/o pólipos serrados sésiles (SSP) con el estado de metilación de los mismos marcadores de ADN de los sujetos de control (por ejemplo, tejido normal, tal como colon normal) (véanse los ejemplos 1-2, tablas 1-5).

Por ejemplo, marcadores y/o paneles de marcadores (por ejemplo, una región cromosómica que tiene una anotación seleccionada entre 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 (por ejemplo, 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 de control (por ejemplo, tejido normal, tal como colon normal) (véanse el ejemplo 1 y las tablas 1A y 1B).

Se identificaron marcadores y/o paneles de marcadores (por ejemplo, una región cromosómica con una anotación seleccionada entre BMP3, NDRG4, PDGFG, CHST2 y SFMBT2) en estudios de casos y controles al comparar el estado de metilación de marcadores de ADN de 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 (por ejemplo, tejido normal tal como colon normal) (véanse el ejemplo 1 y la tabla 2).

Además, se identificaron 185, 244 y 111 marcadores de metilación del ADN específicos para cánceres colorrectales, adenomas grandes y pólipos serrados sésiles, respectivamente (véanse el ejemplo 2 y las tablas 3, 4 y 5). Junto con los casos de cáncer colorrectal, los casos de adenomas grandes y los casos de pólipos serrados sésiles, se secuenció la mucosa colónica normal y el ADN normal de los glóbulos blancos.

Otros experimentos realizados durante el transcurso de las realizaciones de la presente descripción demostraron que NDRG4, BMP3, OPLAH, FLI1, PDGFD, CHST_7889, SFMBT2_895, SFMBT2_896, SFMBT2_897, CHST2_7890, VAV3 y DTX1 son marcadores eficaces para detectar el cáncer colorrectal en muestras de heces (véanse el ejemplo 3 y las tablas 6 y 7).

En consecuencia, en la presente memoria se describe tecnología para marcadores de detección de cáncer colorrectal que proporcionan una alta relación señal a ruido y un bajo nivel de fondo cuando se detectan a partir de muestras tomadas de un sujeto (por ejemplo, muestra de heces; una muestra de tejido colorrectal). Los marcadores se identificaron en estudios de casos y controles al comparar el estado de metilación de los marcadores de ADN de tejido colorrectal de sujetos con neoplasia y/o adenoma colorrectal con el estado de metilación de los mismos marcadores de ADN de sujetos de control (por ejemplo, tejido normal, tal como colon normal) (véanse los ejemplos 1-3 y las tablas 1-6).

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

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 relacionan con marcadores que comprenden una región metilada diferencialmente (región DMR, por ejemplo, una región DMR como se proporciona en las tablas 1-6) que se usan para el diagnóstico o la detección de trastornos proliferativos celulares neoplásicos (por ejemplo, cáncer colorrectal), incluida la detección temprana durante las etapas precancerosas de la enfermedad. Además, los marcadores se utilizan para diferenciar las neoplasias de los trastornos proliferativos celulares 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.

Los marcadores de la presente tecnología son particularmente eficaces para detectar o distinguir entre trastornos proliferativos colorrectales, por lo que se proporcionan medios mejorados para la detección temprana, la clasificación y el tratamiento del cáncer colorrectal.

Además de las realizaciones en donde se analiza el análisis de metilación de al menos un marcador, una región de un marcador o una base de un marcador que comprende una región DMR (por ejemplo, una región DMR como se proporciona en las tablas 1-6) proporcionados en la presente memoria, la tecnología también proporciona paneles de marcadores que comprenden al menos un marcador, región de un marcador o base de un marcador que comprende una región DMR con utilidad para la detección de cánceres colorrectales. Además de las realizaciones en donde se analiza el análisis de metilación de al menos un marcador, una región de un marcador o una base de un marcador que comprende una región DMR (por ejemplo, una región DMR como se proporciona en las tablas 1-6) proporcionados en la presente memoria, la tecnología también proporciona paneles de marcadores que comprenden cualquier tipo o clase de marcadores (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.).

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 una DMR.

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 las secuencias de dinucleótidos de CpG dentro de al menos un marcador que comprende una región DMR (por ejemplo, una región DMR como se proporciona en las tablas 1-6). Los dinucleótidos de CpG genómicos pueden estar metilados o no metilados (conocidos como

alternativa como hiper e hipometilados, respectivamente). Sin embargo, los métodos son adecuados para el análisis de muestras biológicas de naturaleza heterogénea, por ejemplo, una baja concentración de células tumorales o materiales biológicos de las mismas, dentro de un fondo de una muestra remota (por ejemplo, sangre, efluentes de órganos o heces). Por consiguiente, cuando se analiza el estado de metilación de una posición de CpG dentro de dicha muestra, 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.

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

Combinaciones de marcadores

En algunas realizaciones, la tecnología se relaciona con la evaluación del estado de metilación de combinaciones de marcadores que comprenden una región 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 una región DMR. En algunas realizaciones, evaluar el estado de metilación de más de un marcador aumenta la especificidad y/o sensibilidad de una detección o diagnóstico para identificar una neoplasia colorrectal en un sujeto. Además de las realizaciones en donde se analiza el análisis de metilación de al menos un marcador, una región de un marcador o una base de un marcador que comprende una región DMR (por ejemplo, una región DMR como se proporciona en las tablas 1-6) proporcionados en la presente memoria, la tecnología también proporciona paneles de marcadores que comprenden cualquier tipo o clase de marcadores (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.).

Diversos tipos de cáncer se predicen mediante diversas combinaciones de marcadores, por ejemplo, identificados 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

Un método para analizar un ácido nucleico en busca de 5-metilcitosina se basa en el método del bisulfito descrito por Frommer y col. para la detección de 5-metilcitosinas en el ADN (Frommer y col. (1992) Proc. Natl. Acad. Sci. USA 89: 1827-31) o variaciones del mismo. El método del bisulfito para cartografiar 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 según las siguientes etapas: en primer lugar, la citosina reacciona con el 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 la adenina (por lo que se comporta como la timina), mientras que la 5-metilcitosina se empareja con la guanina (por lo que se comporta como la citosina). Esto hace posible la discriminación de citosinas metiladas de citosinas no metiladas, p. ej., 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 US-5.786.146.

Algunas tecnologías convencionales se relacionan con métodos que consisten en encerrar el ADN a analizar en una matriz de agarosa, impidiendo así la difusión y renaturalización del ADN (el bisulfito solo reacciona con el ADN monocatenario), y reemplazando las etapas de precipitación y purificación por una diálisis rápida (Olek A, y col. (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, lo que ilustra la utilidad y la sensibilidad del método. Se proporciona una descripción general de los métodos convencionales para detectar la 5-metilcitosina en Rein, T., y col. (1998) Nucleic Acids Res. 26: 2255.

La técnica del bisulfito típicamente implica amplificar fragmentos específicos, cortos de un ácido nucleico conocido después de un tratamiento con bisulfito, y posteriormente analizar el producto mediante secuenciación (Olek y Walter (1997) Nat. Genet. 17: 275-6) o una reacción de extensión del cebador (Gonzalzo y Jones (1997) Nucleic Acids Res. 25: 2529-31; documento WO 95/00669; La patente US-6.251.594) para analizar posiciones individuales de citosina. Algunos métodos usan 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 y col, documento WO 99/28498). De forma adicional, 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 y col. (1997) Hum Mol Genet. 6: 387-95; Feil y col. (1994) Nucleic Acids Res. 22: 695; Martin y col. (1995) Gene 157: 261-4; el documento WO 9746705; documento WO 9515373).

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

Por ejemplo, la secuenciación genómica se ha simplificado para el análisis de patrones de metilación y distribuciones de 5-metilcitosina usando el tratamiento con bisulfito (Frommer y col. (1992) Proc. Natl. Acad. Sci. USA 89: 1827-1831). De forma adicional, la digestión con enzimas de restricción de productos de PCR amplificados a partir de ADN convertido con bisulfito resulta útil para evaluar el estado de metilación, por ejemplo, como describen Sadri y Hornsby (1997) Nucl. Acids Res. 24: 5058-5059 o como se incorpora en el método conocido como COBRA (análisis combinado de restricción de bisulfito) (Xiong y Laird (1997) Nucleic Acids Res. 25: 2532-2534).

El análisis COBRA™ es un ensayo de metilación cuantitativa útil para determinar los niveles de metilación del ADN en locus específicos 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 un tratamiento estándar con bisulfito según el procedimiento descrito por Frommer y col. (Proc. Natl. Acad. Sci. USA 89:1827-1831, 1992). A continuación, se realiza la amplificación por PCR del ADN convertido con bisulfito usando cebadores específicos para las islas de CpG de interés, seguida de digestión con endonucleasas de restricción, electroforesis en gel y detección usando sondas de hibridación marcadas y 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 cuantitativa lineal en un amplio espectro de niveles de metilación del ADN. Además, esta técnica se puede aplicar de forma fiable al ADN obtenido a partir de muestras de tejido embebido en parafina microdisecionado.

Los reactivos típicos (por ejemplo, como se pueden encontrar en un kit basado en COBRA™ típico) para el análisis COBRA™ pueden incluir, aunque no de forma limitativa: cebadores de PCR para locus específicos (por ejemplo, genes específicos, marcadores, DMR, regiones de genes, regiones de marcadores, secuencia de ADN tratada con bisulfito, isla de CpG, etc.); enzima de restricción y tampón adecuado; oligonucleótido de hibridación génica; oligonucleótido de hibridación de control; kit de marcaje de cinasa para sonda de oligonucleótidos; y nucleótidos marcados. Además, los reactivos de conversión de bisulfito pueden incluir: tampón de desnaturalización de ADN; tampón de sulfonación; reactivos o kits de recuperación de ADN (p. ej., precipitación, ultrafiltración, columna de afinidad); tampón de desulfonación; y componentes de recuperación de ADN.

Preferiblemente, los ensayos tales como “MethylLight™” (una técnica de PCR en tiempo real basada en fluorescencia) (Eads y col, Cancer Res. 59:2302-2306, 1999), Ms-SNuPE™ (Methylation-sensitive Single Nucleotide Primer Extension) reactions (Gonzalzo & Jones, Nucleic Acids Res. 25:2529-2531, 1997), PCR específica de metilación (“MSP”; Herman y col., Proc. Natl. Acad. Sci. USA 93:9821-9826, 1996; La patente US-5.786.146), y amplificación de las islas de CpG metiladas (“MCA”; Toyota y col., Cancer Res. 59:2307-12, 1999) se usan solos o en combinación con uno o más de estos métodos.

La técnica de ensayo “HeavyMethyl™” es un método cuantitativo para evaluar las diferencias de metilación en base a la amplificación específica de metilación del ADN tratado con bisulfito. Las sondas de bloqueo específicas de metilación (“bloqueadores”) que cubren las posiciones de CpG entre los cebadores de amplificación o están cubiertas por ellos permiten la amplificación selectiva específica de la metilación de una muestra de ácido nucleico.

El término ensayo “HeavyMethyl™ MethylLight™” se refiere a un ensayo HeavyMethyl™ MethylLight™, que es una variación del ensayo MethylLight™, en donde el ensayo MethylLight™ se combina con sondas de bloqueo específicas de metilación que cubren las posiciones de CpG entre los cebadores de amplificación. El ensayo HeavyMethyl™ también puede utilizarse junto con cebadores de amplificación específicos de metilación.

Los reactivos típicos (p. ej., como se pueden encontrar en un kit basado en MethylLight™ típico) para el análisis HeavyMethyl™ pueden incluir, aunque no de forma limitativa: cebadores de PCR para locus específicos (por ejemplo, genes específicos, marcadores, DMR, regiones de genes, regiones de marcadores, secuencia de ADN tratada con bisulfito, isla de CpG o secuencia de ADN tratada con bisulfito o isla de CpG, etc.); oligonucleótidos bloqueantes; tampones de PCR optimizados y desoxinucleótidos; y Taq polimerasa.

La MSP (PCR específica de metilación) permite evaluar el estado de metilación de prácticamente cualquier grupo de sitios de CpG dentro de una isla de CpG, independientemente del uso de enzimas de restricción sensibles a la metilación (Herman y col. Proc. Natl. Acad. Sci. USA 93:9821-9826, 1996; La patente US-5.786.146). Brevemente, el ADN se modifica con bisulfito de sodio, que convierte las citosinas no metiladas, pero no a las metiladas, en uracilo, y los productos se amplifican posteriormente con cebadores específicos para ADN metilado frente a no metilado. MSP requiere solo pequeñas cantidades de ADN, es sensible al 0,1 % de alelos metilados de un locus de isla de CpG determinado y se puede realizar en ADN extraído de muestras embebidas en parafina. Los reactivos típicos (p. ej., como se pueden 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 locus específicos (por ejemplo, genes específicos, marcadores, DMR, regiones de genes, regiones de marcadores, secuencia de ADN tratada con bisulfito, isla de CpG, etc.); tampones de PCR y desoxinucleótidos optimizados, y sondas específicas.

El ensayo MethylLight™ es un ensayo de metilación cuantitativo de alto rendimiento que utiliza PCR en tiempo real basada en fluorescencia (por ejemplo, TaqMan®) que no requiere manipulaciones adicionales tras la etapa de PCR (Eads y col,

Cancer Res. 59:2302-2306, 1999). Brevemente, el proceso de MethylLight™ comienza con una muestra mixta de ADN genómico que se convierte, en una reacción de bisulfito de sodio, en un grupo mixto de diferencias de secuencia dependientes de la metilación según los procedimientos estándar (el proceso de bisulfito convierte los residuos de citosina no metilados en uracilo). A continuación, se realiza una PCR basada en fluorescencia en una reacción “sesgada”, por ejemplo, con cebadores de PCR que se solapan con dinucleótidos de CpG conocidos. La discriminación de secuencias se produce tanto a nivel del proceso de amplificación como a nivel del proceso de detección de fluorescencia.

El ensayo MetyLight™ 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 secuencias se produce al nivel de la 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 supuesto sitio de metilación particular. Una reacción en la que ni los cebadores ni la sonda se superponen a ningún dinucleótido de CpG proporciona un control no sesgado de la cantidad de ADN de entrada. Como alternativa, se logra una prueba cualitativa para la metilación genómica sondando el grupo 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.

El proceso de MetyLight™ se utiliza 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 dos conjuntos de reacciones de PCR usando sondas TaqMan®, por ejemplo, con cebadores MSP y/u oligonucleótidos bloqueadores de HeavyMethyl y una sonda TaqMan®. La sonda TaqMan® tiene doble marcaje con moléculas fluorescentes “indicadoras” e “inactivadoras” y está diseñada para ser específica para una región con un contenido en GC relativamente alto, de modo que se funde a una temperatura aproximadamente 10 °C más alta en el ciclo de PCR que los cebadores directos o inversos. Esto permite que la sonda TaqMan® permanezca completamente hibridada durante la etapa de hibridación/extensión de la PCR. A medida que la polimerasa Taq sintetiza enzimáticamente una nueva hebra durante la PCR, eventualmente llegará a la sonda TaqMan® hibridada. La actividad endonucleasa 5' a 3' de la polimerasa Taq desplazará la sonda TaqMan® al digerirla para liberar la molécula indicadora fluorescente para la detección cuantitativa de su señal ahora sin inactivar utilizando un sistema de detección fluorescente en tiempo real.

Los reactivos típicos (por ejemplo, como se pueden encontrar en un kit basado en MethylLight™ típico) para el análisis MethylLight™ pueden incluir, aunque no de forma limitativa: cebadores de PCR para locus específicos (por ejemplo, genes específicos, marcadores, DMR, regiones de genes, regiones de marcadores, secuencia de ADN tratada con bisulfito, isla de CpG, etc.); sondas TaqMan® o Lightcycler®; tampones de PCR optimizados y desoxinucleótidos; y Taq polimerasa.

El ensayo QM™ (metilación cuantitativa) es una prueba cuantitativa alternativa para patrones de metilación en muestras de ADN genómico, en donde la discriminación de secuencias se produce al nivel de hibridación de sonda. En esta versión cuantitativa, la reacción de PCR proporciona una amplificación sin sesgo en presencia de una sonda fluorescente que se superpone a un supuesto sitio de metilación particular. Una reacción en la que ni los cebadores ni la sonda se superponen a ningún dinucleótido de CpG proporciona un control no sesgado de la cantidad de ADN de entrada. Como alternativa, se logra una prueba cualitativa para la metilación genómica sondando el grupo 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 posibles sitios de metilación.

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 marcaje con moléculas fluorescentes “indicadoras” e “inactivadoras” y está diseñada para ser específica para una región con un contenido en GC relativamente alto, de modo que se funde, a una temperatura aproximadamente 10 °C más alta en el ciclo de PCR que la de los cebadores directos o inversos. Esto permite que la sonda TaqMan® permanezca completamente hibridada durante la etapa de hibridación/extensión de la PCR. A medida que la polimerasa Taq sintetiza enzimáticamente una nueva hebra durante la PCR, eventualmente llegará a la sonda TaqMan® hibridada. La actividad endonucleasa 5' a 3' de la polimerasa Taq desplazará la sonda TaqMan® al digerirla para liberar la molécula indicadora fluorescente para la detección cuantitativa de su señal ahora sin inactivar utilizando un sistema de detección fluorescente en tiempo real. Los reactivos típicos (por ejemplo, como se pueden encontrar en un kit basado en QM™ típico) para el análisis QM™ pueden incluir, aunque no de forma limitativa: cebadores de PCR para locus específicos (por ejemplo, genes específicos, marcadores, DMR, regiones de genes, regiones de marcadores, secuencia de ADN tratada con bisulfito, isla de CpG, etc.); sondas TaqMan® o Lightcycler®; tampones de PCR optimizados y desoxinucleótidos; y Taq polimerasa.

La técnica Ms-SNuPE™ es un método cuantitativo para evaluar las diferencias de metilación en sitios de CpG específicos basados en el tratamiento con bisulfito del ADN, seguido de la extensión del cebador de un solo nucleótido (Gonzalvo y 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. A continuación, se realiza la amplificación de la secuencia diana deseada utilizando cebadores de PCR específicos para el ADN convertido con bisulfito, y el producto resultante se aísla y se utiliza como plantilla para el análisis de metilación en el sitio de CpG de interés. Se

pueden analizar pequeñas cantidades de ADN (p. ej., secciones de anatomopatología diseccionadas) y evita la utilización de enzimas de restricción para determinar el estado de metilación en los sitios de CpG.

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

La secuenciación con bisulfito de representación reducida (RRBS) comienza con el tratamiento con bisulfito del ácido nucleico para convertir todas las citosinas no metiladas en uracilo, seguido de la digestión con enzimas de restricción (p. ej., una enzima que reconoce un sitio que incluye una secuencia CG como MspI) y la secuenciación completa de los fragmentos después del acoplamiento a un ligando adaptador. La elección de la enzima de restricción enriquece los fragmentos para las regiones densas de CpG, lo que reduce el número de secuencias redundantes que pueden asignarse a múltiples posiciones de genes durante el análisis. Como tal, la RRBS reduce la complejidad de la muestra de ácido nucleico mediante la selección de un subconjunto (p. ej., mediante selección de tamaño mediante electroforesis en gel preparativa) de fragmentos de restricción para la secuenciación. A diferencia de la secuenciación con bisulfito del genoma completo, cada fragmento producido por la digestión con enzimas de restricción contiene información de metilación del ADN para al menos un dinucleótido de CpG. Como tal, la RRBS enriquece la muestra en promotores, islas de 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 locus genómicos.

Un protocolo típico de RRBS comprende las etapas de digerir una muestra de ácido nucleico con una enzima de restricción como MspI, rellenar salientes y añadir las colas de A, ligar adaptadores, conversión de bisulfito y PCR. Véase, p. ej., y col. (2005) *Nat Methods* 7: 133-6; Meissner y col. (2005) *Nucleic Acids Res.* 33: 5868-77.

En algunas realizaciones, se usa un ensayo de amplificación de señal y diana en tiempo real específico de alelo cuantitativo (QuARTS) para evaluar el estado de metilación. Tres reacciones ocurren 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 de 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 solapa se une débilmente al amplicón. La presencia del oligonucleótido invasivo específico en el sitio de unión diana hace que la escisión libere la secuencia del flap cortando entre la sonda de detección y la secuencia del flap. La secuencia del flap es complementaria a una parte sin horquilla de un casete de FRET correspondiente. En consecuencia, la secuencia del flap funciona como un oligonucleótido invasivo en el casete de FRET y efectúa una escisión entre el fluoróforo del casete de FRET y un inactivador, que produce una señal fluorescente. La reacción de escisión puede cortar múltiples sondas por diana y, por lo tanto, liberar múltiples fluoróforos por flap, lo que proporciona una amplificación exponencial de la señal. QuARTS puede detectar múltiples objetivos en un solo pocillo de reacción mediante el uso de casetes de FRET con diferentes colorantes. Véase, por ejemplo, en Zou y col. (2010) "Sensitive quantification of methylated markers with a novel methylation specific technology" *Clin Chem* 56: A199; La patente US-8.361.720 y en la solc. estadounidense con n.º de serie 13/594.674, 12/946.745, y 12/946.752.

En algunas realizaciones, el ácido nucleico diana se aísla de una muestra mediante, por ejemplo, una captura génica directa. 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 de centrifugación), la captura de un ácido nucleico diana (si está presente) de la muestra aclarada con un reactivo de captura para formar un complejo de captura, el aislamiento del complejo de captura de la muestra clarificada, la recuperación del ácido nucleico diana (si está presente) del complejo de captura en una solución de ácido nucleico y, opcionalmente, repetición para el aislamiento de diferentes dianas (véase, por ejemplo, las solicitudes de patente estadounidenses con n.º 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).

En algunas realizaciones, los fragmentos del ADN tratado se amplifican utilizando conjuntos de oligonucleótidos cebadores (por ejemplo, véase la tabla 2) y una enzima de amplificación. La amplificación de varios segmentos de ADN puede realizarse simultáneamente en un mismo recipiente de reacción. Típicamente, la amplificación se lleva a cabo mediante una reacción en cadena de la polimerasa (PCR). Los amplicones tienen normalmente una longitud de 100 a 2000 pares de bases.

En otra realización del método, se puede detectar el estado de metilación de las posiciones de CpG dentro o cerca de un marcador que comprende una región DMR (por ejemplo, una región DMR como se proporciona en las tablas 1-6) mediante el uso de oligonucleótidos cebadores específicos de metilación. Esta técnica (MSP) ha sido descrita en la pat. US- 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 de MSP contienen al menos un cebador que se hibrida con un dinucleótido de CpG tratado con bisulfito. Por tanto, la secuencia de dichos cebadores comprende al menos un dinucleótido de CpG. Los cebadores de MSP específicos para el ADN no metilado contienen una "T" en la posición de la posición de C en el CpG.

5 Los fragmentos obtenidos mediante la amplificación pueden llevar un marcador directa o indirectamente detectable. En algunas realizaciones, los marcadores son marcadores fluorescentes, radionúclidos o fragmentos de moléculas separables que tienen una masa típica que puede detectarse en un espectrómetro de masas. Cuando dichos marcadores son marcadores de masa, algunas realizaciones prevén que los amplicones marcados tengan una única carga neta positiva o negativa, lo que permite una mejor detectabilidad en el espectrómetro de masas. La detección
10 puede llevarse a cabo y visualizarse por medio de, por ejemplo, espectrometría de masas de ionización/desorción láser asistida por matriz (MALDI) o usando espectrometría de masas de electropulverización (ESI).

En algunas realizaciones, los métodos para aislar el ADN comprenden el aislamiento de ácidos nucleicos como se describe en la solc. de pat. estadounidense con n.º de serie 13/470.251 ("Isolation of Nucleic Acids").

15 Métodos

En algunas realizaciones de la tecnología, se describen métodos que comprenden las siguientes etapas:

20 1) poner en contacto un ácido nucleico (por ejemplo, ADN genómico, por ejemplo, aislado de un fluido corporal, tal como una muestra de heces o tejido colorrectal) obtenido del sujeto con al menos un reactivo o una serie de reactivos que distinga entre dinucleótidos de CpG metilados y no metilados dentro de al menos un marcador DMR (por ejemplo, una región DMR como se proporciona en las tablas 1-6) y

25 2) detectar una neoplasia colorrectal o un trastorno proliferativo (por ejemplo, cáncer colorrectal, adenoma grande, pólipos SSP) (por ejemplo, con una sensibilidad superior o igual al 80 %) y una especificidad superior o igual al 80 %).

En algunas realizaciones de la tecnología, se describen métodos que comprenden las siguientes etapas:

30 1) poner en contacto un ácido nucleico (por ejemplo, ADN genómico, por ejemplo, aislado de un fluido corporal tal como una muestra de heces o tejido colorrectal) obtenido del sujeto con al menos un reactivo o una serie de reactivos que distinga entre dinucleótidos de 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

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

40 Preferiblemente, la sensibilidad es de aproximadamente el 70 % a aproximadamente el 100 %, o de aproximadamente el 80 % a aproximadamente el 90 %, o de aproximadamente el 80 % a aproximadamente el 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 %, de 91 % a 99 %, de 93 % a 97 %, de 94 % a 96 %, de 95 % a 99 %, de 96 % a 99,5 %, de 97 % a 99,9 %, etc.).

El ADN genómico puede aislarse por cualquier medio, incluido el uso de kits comercialmente disponibles. Brevemente, en donde el ADN de interés está encapsulado en una membrana celular, la muestra biológica debe romperse y lisarse por medios enzimáticos, químicos o mecánicos. A continuación, la solución de ADN puede eliminarse de proteínas y otros contaminantes, por ejemplo, mediante digestión con proteinasa K. A continuación, el ADN genómico se recupera de la solución. Esto puede llevarse a cabo por medio de diversos métodos que incluyen la precipitación salina, 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, el coste y la cantidad requerida de ADN. Todos los tipos de muestras que comprenden materia neoplásica o materia preneoplásica son adecuados para su uso en el presente método, por ejemplo, líneas celulares, portaobjetos de histología, biopsias, tejido embebido en parafina, fluidos corporales, heces, efluentes colónicos, orina, plasma sanguíneo, suero sanguíneo, sangre completa, células sanguíneas aisladas, células aisladas de la sangre y combinaciones de los mismos.

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 la captura génica directa, por ejemplo, como se detalla en las solicitudes de patente estadounidense con n.º 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.

65

A continuación, la muestra de ADN genómico se trata con al menos un reactivo o una serie de reactivos, que distingue entre dinucleótidos de CpG metilados y no metilados dentro de al menos un marcador que comprende una región DMR (por ejemplo, una región DMR como se proporciona en las tablas 1-6).

5 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.

10 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 sea 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.

15 El ácido nucleico tratado se analiza después para determinar el estado de metilación de las secuencias del gen diana (al menos un gen, secuencia genómica o nucleótido de un marcador que comprende una región DMR, por ejemplo, al menos una región DMR como se proporciona en las tablas 1-6). En algunas realizaciones, el método de análisis es QUARTS y/o MSP, tal como se describe en la presente memoria.

20 La metilación aberrante, más específicamente la hipermetilación de un marcador que comprende una región DMR (por ejemplo, una región DMR como se proporciona en las tablas 1-6) se asocia con un cáncer colorrectal.

25 La tecnología se relaciona con el análisis de cualquier muestra asociada a un cáncer colorrectal. Por ejemplo, en algunas realizaciones la muestra comprende un tejido y/o fluido biológico obtenido de un paciente. En algunas realizaciones, el detergente comprende SDS. 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 desprendidas en la luz gastrointestinal y/o células gastrointestinales recuperadas de heces. En algunas realizaciones, el sujeto es un ser humano. Estas muestras pueden proceder de la porción alta del tubo digestivo, la porción baja del tubo digestivo o comprender células, tejidos y/o secreciones tanto de la parte alta como de la parte baja del tubo digestivo. 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 fluido celular, ascitis, orina, heces, fluido pancreático, fluido obtenido durante una endoscopia, sangre, moco o saliva. En algunas realizaciones, la muestra es una muestra de heces.

35 Dichas muestras se pueden obtener mediante cualquier variedad de técnicas. Por ejemplo, las muestras de orina y heces se pueden obtener fácilmente, mientras que las muestras de sangre, ascitis, suero o líquido colorrectal o pancreático se pueden obtener por vía parenteral utilizando, por ejemplo, una aguja y una jeringa. Las muestras libres de células o sustancialmente libres de células se pueden obtener sometiendo la muestra a diversas técnicas que incluyen, aunque no de forma limitativa, centrifugación y filtración. Aunque generalmente se prefiere que no se utilicen técnicas invasivas para obtener la muestra, aún puede ser preferible obtener muestras como homogeneizados de tejido, secciones de tejido y muestras de biopsia

45 La tecnología se puede usar para tratar a un paciente (por ejemplo, un paciente con cáncer colorrectal, con cáncer colorrectal en estadio temprano o que puede desarrollar cáncer colorrectal), el método comprende determinar el estado de metilación de una o más regiones DMR como se proporcionan en la presente memoria y administrar un tratamiento al paciente basado en los resultados de la determinación del estado de metilación. El tratamiento puede consistir en realizar una colonoscopia, administrar un compuesto farmacéutico, una vacuna, realizar una cirugía, obtener imágenes del paciente, realizar otra prueba. Preferiblemente, dicho uso es en un método de detección clínica, un método de evaluación de pronóstico, un método de seguimiento de los resultados del tratamiento, un método para identificar a los pacientes con mayor probabilidad de responder a un tratamiento terapéutico particular, un método de obtención de imágenes de un paciente o sujeto y un método para el cribado y desarrollo de fármacos.

55 En algunas realizaciones, el pronóstico clínico del cáncer incluye determinar la agresividad del cáncer y la probabilidad de recurrencia del tumor para planificar la terapia más eficaz. Si se puede hacer un pronóstico más preciso o incluso se puede evaluar un riesgo potencial de desarrollar el cáncer, se puede elegir la terapia adecuada y, en algunos casos, una terapia menos agresiva para el paciente. La evaluación (por ejemplo, determinar el estado de metilación) de los biomarcadores del cáncer es útil para separar sujetos con buen pronóstico y/o bajo riesgo de desarrollar cáncer que no necesitarán terapia o una terapia limitada de aquellos con mayor probabilidad de desarrollar cáncer o sufrir una recurrencia del cáncer que podrían beneficiarse de tratamientos o seguimientos más intensivos.

60 En algunas realizaciones de la materia descrita actualmente, pueden realizarse múltiples determinaciones de los biomarcadores a lo largo del tiempo para facilitar el diagnóstico y/o el 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 una realización de este tipo, por ejemplo, se podría esperar ver un cambio en el estado de metilación de uno o más biomarcadores (p. ej., DMR) descritos en la presente

65

memoria (y potencialmente uno o más biomarcadores adicionales, si se controlan) en una muestra biológica a lo largo del tiempo durante el curso de una terapia eficaz.

5 Los métodos y composiciones de la presente descripción se pueden emplear para el tratamiento o diagnóstico de enfermedades en un estadio temprano, por ejemplo, antes de que aparezcan los síntomas de la enfermedad.

10 En algunas realizaciones, un análisis estadístico asocia un indicador de pronóstico con una predisposición a un resultado adverso. Por ejemplo, en algunas realizaciones, un estado de metilación diferente al de una muestra de control normal obtenida de un paciente que no tiene cáncer puede indicar que es más probable que un sujeto sufra cáncer que sujetos con un nivel que es más similar al estado de metilación en la muestra de control, determinado por un nivel de significación estadística. Además, un cambio en el estado de metilación desde un nivel inicial (por ejemplo, "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 acontecimientos adversos. La significancia estadística a menudo se determina comparando dos o más poblaciones y determinando un intervalo de confianza y/o un valor de p. Véase, 15 por ejemplo, Dowdy y Wearden, *Statistics for Research*, John Wiley & Sons, Nueva York, 1983. Los intervalos de confianza ilustrativos de la presente materia son 90 %, 95 %, 97,5 %, 98 %, 99 %, 99,5 %, 99,9 % y 99,99 %, mientras que los valores p ilustrativos son 0,1, 0,05, 0,025, 0,02, 0,01, 0,005, 0,001, y 0,0001.

20 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 la presente memoria (por ejemplo, una DMR), y simplemente se compara el grado de cambio en el estado de metilación del biomarcador en una muestra biológica 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 en la presente memoria es de aproximadamente el 5 %, aproximadamente el 10 %, aproximadamente el 15 %, aproximadamente el 20 %, 25 aproximadamente el 25 %, aproximadamente el 30 %, aproximadamente el 50 %, aproximadamente el 75 %, 25 aproximadamente el 100 % y aproximadamente el 150 %. En otras realizaciones más, se puede establecer un "nomograma", mediante el cual un 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. El experto en la materia está familiarizado con el uso de dichos 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 30 porque se hace referencia a mediciones de muestras individuales, no a promedios de población.

35 En algunas realizaciones, una muestra de control se analiza simultáneamente con la muestra biológica, de manera que los resultados obtenidos de la muestra biológica pueden compararse con los resultados obtenidos de la muestra de control. Además, se contempla que se puedan proporcionar curvas patrón con las que se puedan comparar los resultados del ensayo para la muestra biológica. Estas curvas patrón presentan los estados de metilación de un biomarcador en función de las unidades de ensayo, por ejemplo, la intensidad de la señal fluorescente, si se utiliza un marcador fluorescente. Usando muestras tomadas de múltiples donantes, se pueden proporcionar curvas patrón para controlar los estados de metilación de uno o más biomarcadores en tejido normal, así como para los niveles "en riesgo" de uno o más biomarcadores en tejido tomado de donantes con metaplasia o de donantes con un cáncer gastrointestinal. 40 En ciertas realizaciones del método, se identifica que un sujeto tiene metaplasia al identificar un estado de metilación aberrante de una o más DMR proporcionadas en la presente memoria 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 se identifique que el sujeto tiene cáncer.

45 El análisis de marcadores se puede realizar por separado o simultáneamente con marcadores adicionales dentro de una muestra de prueba. Por ejemplo, se pueden combinar varios marcadores en una prueba para el procesamiento eficiente de múltiples muestras y para proporcionar potencialmente una mayor precisión de diagnóstico y/o pronóstico. Además, un experto en la técnica reconocería el valor de probar múltiples muestras (por ejemplo, en puntos de tiempo sucesivos) del mismo sujeto. Dicha prueba de muestras en serie puede permitir la identificación de cambios en los estados de metilación del marcador a lo largo del tiempo. Los cambios en el estado de metilación, 50 así como la ausencia de cambios en el estado de metilación pueden proporcionar información útil sobre el estado de la enfermedad que incluye, aunque no de forma limitativa, identificar el tiempo aproximado desde el inicio del suceso, la presencia y la cantidad de tejido recuperable, la idoneidad de las terapias farmacológicas, la eficacia de varias terapias y la identificación del resultado del sujeto, incluido el riesgo de sucesos futuros.

55 El análisis de biomarcadores se puede llevar a cabo en una variedad de formatos físicos. Por ejemplo, se puede utilizar el uso de placas de microtitulación o la automatización para facilitar el procesamiento de un gran número de muestras de prueba. Como alternativa, se pueden usar formatos de muestra única para facilitar el tratamiento y diagnóstico inmediatos de manera oportuna, por ejemplo, en transporte ambulatorio o salas de urgencias.

60 En algunas realizaciones, al sujeto se le diagnostica cáncer colorrectal si, en comparación con un estado de metilación de control, existe una diferencia mensurable 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 que el sujeto no tiene cáncer colorrectal, que no tiene riesgo de cáncer o que tiene un riesgo 65 bajo de cáncer. A este respecto, los sujetos que tienen cáncer o riesgo del mismo pueden diferenciarse de sujetos que tienen cáncer o riesgo de cáncer bajo o sustancialmente nulo. Aquellos sujetos que tengan un riesgo de

desarrollar un cáncer colorrectal pueden someterse a un programa de detección más intensivo y/o regular, que incluye la vigilancia endoscópica. Por otro lado, aquellos sujetos que tienen un riesgo bajo o sustancialmente nulo pueden evitar someterse a una endoscopia, hasta que un examen futuro, por ejemplo, un examen realizado según la tecnología actual, indique que ha aparecido en estos sujetos un riesgo de cáncer colorrectal.

5 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 a un sujeto que tiene, o corre el riesgo de desarrollar, un
10 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 con respecto a 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 en las que se analiza una muestra de control simultáneamente con la muestra biológica, el estado de metilación predeterminado es el estado de metilación en la muestra de control. En otras
15 realizaciones del método, el estado de metilación predeterminado se basa y/o se identifica mediante una curva patrón. En otras realizaciones del método, el estado de metilación predeterminado es un estado o rango de estado específico. Como tal, se puede elegir el estado de metilación predeterminado, dentro de límites aceptables que serán evidentes para los expertos en la técnica, basándose en parte en la realización del método que se practica y la especificidad deseada, etc.

20 El objeto de la presente invención incluye además un sistema para diagnosticar un cáncer gastrointestinal en un sujeto. El sistema se puede proporcionar, por ejemplo, como un kit comercial que se puede utilizar 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 ilustrativo proporcionado según la presente tecnología incluye evaluar el estado de metilación de una región DMR, como se proporciona en las tablas 1-6.

25 Ejemplos

Ejemplo 1.

30 Este ejemplo describe estudios tisulares independientes de casos y controles que identificaron marcadores de metilación altamente discriminantes para la neoplasia colorrectal. Los ejemplos utilizaron la secuenciación RRBS para la fase de descubrimiento y la PCR específica de la metilación (MSP) y la amplificación cuantitativa de señal y diana en tiempo real específica de alelos (QUARTS) para la fase de validación.

35 Las muestras de tejido se identificaron a partir de los registros de cáncer existentes. La población accesible incluyó a aquellos que se sometieron a colectomía abierta o laparoscopia o biopsia de colon con una muestra archivada. Todos los tejidos fueron revisados por un anatomopatólogo gastrointestinal experto para confirmar la clasificación correcta. Los tejidos de los casos de neoplasia colorrectal incluyeron cánceres colorrectales en estadios I-IV (recién congelados), adenomas avanzados >1 cm de tamaño (recién congelados) y pólipos serrados sésiles del lado derecho (FFPE). Se estudiaron dos grupos de control. El primer grupo de control incluyó 18 tejidos epiteliales de colon de pacientes que se confirmó que
40 estaban libres de neoplasia de colon. El segundo grupo incluyó 18 muestras de capa leucocitaria normal de pacientes sin cáncer. Los casos y ambos controles se compararon según el sexo y la edad. Además, las cohortes de CCR y 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 utilizando una técnica de fenol-cloroformo, lo que produjo al menos 500 ng de ADN. La identificación de casos, la comparación y la extracción de ADN fueron realizadas por personal independiente para mantener el enmascaramiento del personal de laboratorio sobre el estado de casos y
45 controles.

50 El ADN genómico (300 ng) se fragmentó mediante digestión con 10 unidades de MspI, una enzima de restricción específica de metilación que reconoce motivos que contienen CpG. Esto enriquece las muestras en contenido de CpG y elimina áreas redundantes del genoma. Los fragmentos digeridos se repararon en los extremos y se formaron colas en 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 contenían una de las cuatro secuencias de código de barras (para vincular cada fragmento a su ID de muestra). La selección de tamaño de fragmentos de 160-340 pb (insertos de 40-220 pb) se realizó usando microesferas/tampón Agencourt AMPure XP SPRI (Beckman Coulter, Brea CA). Los límites del tampón fueron de 0,7X a 1,1X de volúmenes de muestra de microesferas/tampón. El volumen de elución final fue de 22 ul (tampón EB - Qiagen, Germantown MD). Se usó qPCR para medir la eficiencia de la ligadura y la calidad de los fragmentos en una pequeña alícuota de muestra. A continuación, las muestras se sometieron a conversión de bisulfito (dos veces) utilizando un protocolo EpiTect modificado (Qiagen). La qPCR y la PCR convencional (PfuTurbo Cx hotstart - Agilent, Santa Clara CA) seguidas de la evaluación Bioanalyzer 2100 (Agilent) en alícuotas de muestras convertidas determinaron el número óptimo de ciclos de PCR antes de
55 la amplificación de la biblioteca final. Condiciones para la PCR final: 50 ul de rxn: 5 ul de tampón 10X, 1,25 ul de cada dNTP 10 mM, 5 ul de cóctel de cebadores (~5 uM), 15 ul de cadena molde (muestra), 1 ul de PfuTurbo Cx hotstart, 22,75 de agua. 95 C-5 min; 98 C-30 s; 16 ciclos de 98 C-10 s, 65 C-30 s, 72 C-30 s; 72 C-5 min; 4 C Las muestras se combinaron (equimolarmente) en bibliotecas de 4 unidades de plexado según el esquema de aleatorización y se analizaron con el bioanalizador para verificar el tamaño final y con qPCR usando patrones phiX y cebadores específicos de adaptador.
60
65

Las muestras se cargaron en carriles de celdas de flujo según una asignación aleatoria de carriles con carriles adicionales reservados para controles de ensayo internos. La secuenciación se realizó por el Next Generation Sequencing Core en el Medical Genome Facility de la Clínica Mayo en un instrumento Illumina HiSeq 2000. Las lecturas fueron unidireccionales durante 101 ciclos. Cada carril de celda de flujo generó 100-120 millones de lecturas, suficiente para una mediana de cobertura de 30 a 50 veces la profundidad de secuenciación (número de lecturas por CpG) para secuencias alineadas. Se usó el software de flujo de trabajo de Illumina estándar para realizar las llamadas de bases y la generación de lecturas de secuencias en formato fastq. Para la alineación de secuencias y la extracción de metilación, como anteriormente (véase, por ejemplo, Sun, y col., 2012 *Bioinformatics* 28(16):2180-1), se usó SAAP-RRBS, un proceso de anotación y análisis optimizado para la secuenciación de bisulfito de representación reducida.

Se realizaron dos estudios de validación basados en MSP en conjuntos de muestras ampliados para confirmar la precisión y la reproducibilidad de los candidatos metilados diferencialmente observados. El primero, un estudio de *validación interna* (MSP), se llevó a cabo en muestras coincidentes con enmascaramiento usando réplicas biológicas y técnicas de neoplasia colorrectal, colon normal y leucocitos normales. Esta etapa se realizó para garantizar que los sitios de metilación diferencial identificados por la filtración de datos de 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, un experimento de *validación externa*, utilizó la tecnología QuARTs para probar los mejores candidatos en muestras de neoplasia colorrectal independientes, asignadas al azar, coincidentes, con enmascaramiento y muestras de colon normal.

Los cebadores para cada marcador se diseñaron para que se dirigiesen 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 de tratamiento con bisulfito y entrada de ADN. El diseño se realizó mediante el software Methprimer (Universidad de California, San Francisco CA) o mediante métodos semimanuales. A continuación, los ensayos se probaron y optimizaron mediante qPCR con colorantes SYBR Green (Life Technologies, Grand Island, NY) en diluciones de controles de ADN genómico universalmente metilados y no metilados.

Las reacciones de MSP se realizaron en ADN extraído de tejido como se ha descrito anteriormente (véase, por ejemplo, Kisiel, y col., 2012 *Cancer* 118(10):2623-2631). Brevemente, el ADN se trató con bisulfito utilizando el kit de metilación de ADN EZ (Zymo Research, Orange, CA) y se eluyó en tampón. Se usó 1 μ l de ADN tratado con bisulfito como molde 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 LightCycler (Indianápolis, IN), donde se usó ADN metilado universal CpGenome tratado con bisulfito (Millipore, Billerica, MA) como control positivo y se diluyó en serie para crear curvas patrón para todas las placas. Las secuencias de oligonucleótidos y las temperaturas de hibridación están disponibles previa petición.

La principal comparación de interés fue la diferencia de metilación entre los casos y los controles de colon en cada CpG cartografiada. Las islas de CpG se definen bioquímicamente por una relación de CpG observada con respecto a la esperada superior a 0,6 (31). Sin embargo, para este modelo, se crearon unidades de mosaico de análisis de CpG de "región metilada diferencialmente (DMR)" en función de la distancia entre las ubicaciones de los sitios de CpG para cada cromosoma. Como la distancia entre cualquier CpG excedía la ubicación anterior o siguiente en más de 100 pb, se creaba 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 era ≥ 200 lecturas (lo que equivale aproximadamente a un promedio de 10 lecturas por sujeto) y la variación del % de metilación era mayor que cero (se excluyeron sitios de CpG no informativos con varianza 0). 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 que el tamaño de la muestra de individuos para cada grupo era 18.

La significación estadística se determinó mediante regresión logística sobre el % de metilación por región DMR (utilizando 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 sobredisperso, donde el parámetro de dispersión se estimó utilizando la estadística de χ -cuadrado de Pearson de los residuos del modelo ajustado. Para evaluar la metilación específica de la hebra, las regiones directa e inversa se analizaron por separado. A continuación, las regiones DMR se clasificaron según 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 ≥ 10 % en los casos. Cada DMR significativa se consideró como un marcador candidato.

Para el estudio de validación interna, el resultado principal fue el área bajo la curva de características operativas del receptor (ABC) para cada marcador. Esto se calculó mediante regresión logística (JMP versión 9.0.1, SAS Institute, Cary NC) para modelar la intensidad del número de copias corregida por mediana de cada marcador con neoplasia colorrectal en comparación con colon normal y leucocitos normales. Para el estudio de validación externa, se seleccionaron los marcadores con los valores de ABC más altos y la relación más amplia de la mediana del número de copias del marcador entre casos y controles. El resultado principal para el experimento de validación externa fue la ABC de cada marcador representado frente a la intensidad de la señal de cada marcador, medida por el logaritmo de la relación de la mediana del %

de metilación corregido en los casos en comparación con los controles. Con dieciocho casos, hay una potencia >80 % para detectar un área bajo la curva de 0,85 o superior a partir de la hipótesis nula de 0,5 con un nivel de significación bilateral de 0,05.

5 Descubrimiento de marcadores RRBS Se secuenciaron mediante RRBS extractos de ADN coincidentes, con enmascaramiento y asignados al azar de 18 cánceres colorrectales, 18 adenomas avanzados (> 1 cm), 18 pólipos serrados sésiles, 18 tejidos epiteliales de colon normales y 18 controles de leucocitos derivados de la capa leucocitaria normal. La mediana de edad era de 65 años (rango intercuartílico 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 de CpG donde se cumplían los
10 criterios de varianza y cobertura de grupo, se consideró un rango aproximado de 2-3 millones de sitios de CpG para el análisis adicional. Las muestras de la SSA, que se derivaron de bloques de tejido FFPE tenían un número menor de CpG (200 000-1,2 millones) debido a su menor calidad inherente). Las regiones DMR 1068 (CCR), 1200 (adenoma avanzado) y 268 (SSA) cumplieron con los criterios de significación para la metilación diferencial. Estas se agruparon en 185, 244 y 109 regiones candidatas con suficientes firmas de metilación para el diseño de cebadores de MSP. La longitud de las regiones DMR osciló entre 30 y más de 1000 bases. Las firmas de metilación contenían de 6 a 69 CpG contiguos. El 84 % de las regiones DMR se refieren a las regiones reguladoras (promotoras) 5' de los genes, la mayoría de las cuales se asocian a islas CpG más grandes. En el caso de las regiones DMR del CCR, aproximadamente el 15 % tiene asociaciones previas con la neoplasia colorrectal. El 50 % se ha informado 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.) Algunos de los sitios anotados contienen múltiples regiones DMR.

Validación interna En función del número de CpG vecinos en cada firma de metilación del gen candidato, se diseñaron cebadores para los 50 candidatos principales de los 185 marcadores de CCR. La clasificación se realizó con referencia a las métricas de regresión logística y a la relación entre el número de copias de casos/controles. 35 de ellas superaron el control de calidad (CC) interno y 15 fueron rechazadas. A continuación, se utilizó la MSP para analizar los 35 candidatos en muestras de ADN de otras cohortes coincidentes, con enmascaramiento y parcialmente independientes, que incluían 36 muestras de lesiones del CCR, 36 adenomas avanzados y 36 muestras de epitelio colónico normal. Además, se extrajeron 31 muestras de ADN de pacientes con enfermedad inflamatoria intestinal (EII) (18 con CCR, 2 con displasia de alto grado y 13 controles normales). La *β-actina* se amplificó en todas las muestras. De los 35 ensayos de MSP, 18 marcadores candidatos tuvieron un ABC > 0,88, relaciones de número de copias de caso/control > 50 y metilación de fondo < 1 % (véanse las tablas 1A y 1B). Estos fueron seleccionados para su inclusión en una validación externa independiente.

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

Marcador	AUC	Mediana S/N	Metilación de fondo (%)	Cromosoma	Coordenadas genómicas de DMR
FLI1	0,986	>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 región 2	0,971	90	0,71	10	7452746-7452779
KCNK12	0,963	>1000	0	2	47797187-47797452
VAV3 región 1	0,954	135	0,27	1	108507074-108507498
SFMBT2 región 3	0,952	>1000	0	10	7452885-7452956
PPP2R5C	0,949	>1000	0,1	14	102247525-102247929
CHST2 región 2	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 región 1	0,931	390	0,07	3	142838025-142838494
SFMBT2 región 1	0,931	>1000	0	10	7452029-7452452
QKI	0,921	50	0,61	6	163834534-163834925
VAV3 región 2	0,892	>1000	0	1	108507609-108507674
SLC8A3	0,883	596	0,1	14	70655516-70655712

ES 2 980 778 T3

Tabla 1B: Marcadores más discriminatorios en tejidos independientes por MSP, incluidos los cebadores de MSP directos y los cebadores de MSP inversos

Marcador	Cebador de MSP directo	Cebador de MSP inverso
FLI1	GGGAGTGAGGGTAGGGCGTTC (Id. de sec. n.º: 1)	CTCGCAACCCCTTCGAATTAACCCG (Id. de sec. n.º: 2)
OPLAH	TGCGTAGGTGATAGGGAGGGGTTAC (Id. de sec. n.º: 3)	ACAAAACACATCCTATTAACGCGAA (Id. de sec. n.º: 4)
DTX1	GAGTCGCGGTTTCGTTTTTC (Id. de sec. n.º: 5)	GACGCGACGACCGAAAAAC (Id. de sec. n.º: 6)
MATK	TGCACACCCCGAGGCGGTCCCGG (Id. de sec. n.º: 7)	CGCCCCAAAATAAAAAACGAA (Id. de sec. n.º: 8)
SFMBT2 región 2	GCGTTTAGGTTGGTCCGAGA (Id. de sec. n.º: 10)	CCTAACCAACGCACTCAACC (Id. de sec. n.º: 11)
KCNK12	CGTACGTGGCGTTTTAGCGC (Id. de sec. n.º: 12)	TCGAAAACCCCGACGAAACGAAACG (Id. de sec. n.º: 13)
VAV3 región 1	GCGTAAGGTCGAATATTGAGTCGA (Id. de sec. n.º: 14)	AAAATACTACCCACCAACCACCGAA (Id. de sec. n.º: 15)
SFMBT2 región 3	GTCGTCGTTCCGAGAGGGTA (Id. de sec. n.º: 16)	CGAACAAAAACGAACGAACGAA (Id. de sec. n.º: 17)
PPP2R5C	TCGATTTTAAITTTTGTGTUGTTGTAGAT TCGC (Id. de sec. n.º: 18)	GAAAAAACTAAAAACGACAAAAAAACC CGACG (Id. de sec. n.º: 19)
CHST2 región 2	GGAACGAGTGATAGTCGGATAGTTTCGTC (Id. de sec. n.º: 20)	CGCCCGAAAACGACCCCG (Id. de sec. n.º: 21)
PKIA	CGGGGATGATTTTATGTAGTCGGAGTTT CGC (Id. de sec. n.º: 22)	CCCGCCGAATACTCGATCAACTCG (Id. de sec. n.º: 23)
PDGFD	GCGAATAAATAAACGTTAATTTGTTGTTT GTTTC (Id. de sec. n.º: 24)	CCGAACGCGTATAAATACCGCACTT (Id. de sec. n.º: 25)
ELOVL2	CGGTTTTATTATTATGATTCGTAGCGG (Id. de sec. n.º: 26)	CGACTACCCTAAACAACGCATCGC (Id. de sec. n.º: 27)
CHST2 región 1	CGAGTTCGGTAGTTGTACGTAGA (Id. de sec. n.º: 28)	CGAAATACGAACGCGAAATCTAAACT (Id. de sec. n.º: 29)
SFMBT2 región 1	GCGACGTAGTCGTCGTTGT (Id. de sec. n.º: 30)	CCAACGCGAAAAAACGCG (Id. de sec. n.º: 31)
QKI	GAGCGGACGTCGCGGTAC (Id. de sec. n.º: 32)	CGCCACGACGCGAATCTTAACTACG (Id. de sec. n.º: 33)
VAV3 región 2	GGATCGAGGGAGTAGGAGTCGC (Id. de sec. n.º: 34)	CGAAACCGAACCTAACGCGACG (Id. de sec. n.º: 35)
SLC8A3	AGTTTTTTCGCGCGTTTTTTTTGO (Id. de sec. n.º: 36)	GCCGAATCTCCGCTTACACG (Id. de sec. n.º: 37)

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

Validación externa El ADN coincidente, con enmascaramiento y asignado al azar de 40 CCR, 24 adenomas avanzados y 40 muestras epiteliales de colon normales se analizaron mediante QUARTs para los 18 principales candidatos. La mediana de edad de este subconjunto fue de 60 años (rango intercuartílico 52-67). El equilibrio de género fue de 50:50. Lo mismo para los porcentajes de lesiones distales:proximales en los casos. La β -actina se amplificó en todas las muestras. Los 18 marcadores demostraron un rendimiento acorde con los resultados de la validación interna. En el análisis del CCR, 9 marcadores mostraron un rendimiento superior tanto en términos de las características del ABC como de las relaciones de metilación entre los casos y los controles. Como se muestra en la siguiente tabla (tabla 1C), estos 9 mostraron una excelente asociación con el cáncer colorrectal.

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

Marcador	ABC ROC	Error estándar	Valor Z	Pr (> z)	Inferior, 95	Superior, 95	% medio de metilación: caso/control
vav3_877 (VAV3 región 2)	0,8718	0,0354	10,5	0	0,8024	0,9412	5046
chst2_7889 (CHST2 región 1)	0,8462	0,0374	9,25	0	0,7728	0,9195	10127
sfmbt2_896 (sfmbt2 región 2)	0,9526	0,0251	18,05	0	0,9034	1,0017	139
sfmbt2_895 (sfmbt2 región 1)	0,9439	0,0266	16,72	0	0,8919	0,996	580
pdgfd	0,8814	0,0351	10,85	0	0,8125	0,9503	1351
dtx1	0,9606	0,0222	20,77	0	0,9171	1,004	883
fli1	0,9744	0,0179	26,51	0	0,9393	1,0094	1131
sfmbt2_897 (sfmbt2 región 3)	0,9311	0,0293	14,73	0	0,8737	0,988 5	323
chst2_7890 (chst2 región 2)	0,9231	0,0293	14,46	0	0,8657	0,9804	6970

Para el estudio de validación de las heces de EII, se eligieron tres marcadores (PDGFD, CHST2 7889 y SFMBT2 896) y se compararon con el BMP3 y el NDRG4.

Se descubrieron marcadores de ADN metilado y se pusieron a prueba para medir la detección del CRN asociado a la EII (IBD-CRN: displasia de bajo grado [LGD], displasia de alto grado [HGD] y cáncer colorrectal [CCR]).

5 Los marcadores se identificaron y probaron en 3 etapas separadas: descubrimiento; validación biológica; y pilotaje clínico. En primer lugar, un experimento de descubrimiento identificó los marcadores mediante la secuenciación con bisulfito de representación reducida (RRBS) en el ADN extraído de muestras de tejido congelado de archivos de CCR y adenomas esporádicos. En segundo lugar, los marcadores candidatos se validaron mediante un ensayo de PCR específica de metilación (MSP) en el ADN extraído de tejidos de archivo de pacientes con IBD-CCR y controles de IBD sin CRN. En tercer lugar, las heces archivadas de casos independientes de IBD-CRN y controles de IBD se analizaron mediante amplificación cuantitativa de señal y diana en tiempo real específica de alelos (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 las variables clínicas se probó mediante la suma de rangos de Chi-square y Wilcoxon para obtener datos categóricos y continuos, respectivamente.

15 Mediante RRBS se secuenciaron 18 muestras de CCR esporádico, 18 de adenomas avanzados y 18 muestras de colon normal; los 20 mejores candidatos fueron analizados por MSP en muestras independientes, incluidas 18 IBD-CRC y 13 controles de IBD. Se seleccionaron tres marcadores (PDGFD, CHST2, SFMBT2) para compararlos con BMP3 y NDRG4; todos fueron analizados mediante QuARTS en muestras de heces de 33 casos de IBD-CRN (8 CCR, 8 HGD, 8 LGD \geq 1 cm, 10 LGD $<$ 1 cm) y 50 controles de IBD. Se excluyeron cuatro controles por insuficiencia de $[\beta]$ -actina. La mediana de la duración de la EII fue de 23 años (rango intercuartílico [IQR] 9-35) en los casos y de 13 (8-20) años en los controles ($p=0,0009$). No se observaron otras diferencias significativas al comparar la edad, el sexo, la gravedad de la inflamación, el alcance de la EII o la colangitis esclerosante primaria comórbida. Los niveles de PDGFD, CHST2, SFMBT2 se vieron moderadamente influenciados por la duración de la enfermedad ($p=0,04, 0,02, 0,04$), pero no así BMP3 y NDRG4. Otras variables no fueron significativas. Se presentan tasas de detección con una especificidad del 90 % (tabla 2A, 2B y 2C).

Para las muestras de CCR ($n = 1$) y HGD ($n = 2$) que fueron negativas, se revisaron los bloques H&E. Una muestra de HGD se reclasificó como cambio reactivo. El ADN se extrajo de los tejidos restantes de HGD y CCR, y se analizó mediante QUARTS para cada marcador anterior. Estas fueron muy positivas.

Los ensayos de QUARTS se repitieron en las muestras de heces correspondientes con un cambio mínimo en el % de metilación; sin embargo, el número de copias sin procesar aumentó para cada marcador, de modo que BMP3, NDRG4, PDGFG, CHST27889 pudieron detectar 8/8 CCR y 6/7 HGD (14/15 CCR+HGD, 93 % de sensibilidad) con un rango de especificidad de 89-93 %.

Tabla 2A: Detección de neoplasias colorrectales asociadas a la EII mediante ADN fecal metilado con una especificidad del 90 %

Marcador metilado	CCR	CCR + HGD (n=16)	HGD	LGD \geq 1 cm	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 %)

Tabla 2B: Detección de neoplasias colorrectales asociadas a la EII mediante ADN fecal metilado con una especificidad del 90 %, incluidas las coordenadas genómicas de la región DMR

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

Tabla 2C: Detección de neoplasias colorrectales asociadas a la EII mediante ADN fecal metilado con una especificidad del 90 %, incluidos los cebadores de MSP directos, la secuencia de sonda y los cebadores de MSP inversos

Marcador metilado	Cebador directo	Secuencia de la sonda	Cebador inverso
BMP3	GTTTAATTTTCGGTTTCGTCGT C (Id. de sec. n.º: 38)	CGCCGAG GCGGTTTT TTGCG (Id. de sec. n.º: 39)	CGCTACGAAACTCCGA (Id. de sec. n.º: 40)
NDRG4	CGGTTTTCGTTTCGTTTTTCG (Id. de sec. n.º: 41)	CCACGGA CGGTTTCGT TTATCG (Id. de sec. n.º: 42)	CCGCCTTCTACGCGACTA (Id. de sec. n.º: 43)
PDGFG	GCGAATAAATAAACGTTAATTT G TTGTTTGTTC (Id. de sec. n.º: 44)	CCACGGA CGCGCACT TCCTTA (Id. de sec. n.º: 45)	CCGAACGCGTATAAATACCG CACTT (Id. de sec. n.º: 46)
CHST2	CGAGTTCGGTAGTTGTACGTAG A (Id. de sec. n.º: 47)	CGCCGAG GTCGTCTGA TACCG (Id. de sec. n.º: 48)	CGAAATACGAACGCGAAATC TAAAACT (Id. de sec. n.º: 49)
SFMBT2	GCGACGTAGTCGTCGTTGT (Id. de sec. n.º: 50)	CCACGGA CGGAAAAC GCGAAA (Id. de sec. n.º: 51)	CCAACGCGAAAAAACGCG (Id. de sec. n.º: 52)

Ejemplo 2.

Este ejemplo describe la identificación de marcadores de ADN metilado para la detección del cáncer colorrectal y el precáncer.

En la presente memoria se describe 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 serrados sésiles (SSP) (tablas 5A y 5B), todos identificados a partir de datos generados por el enriquecimiento de islas de CpG junto con la secuenciación masiva en paralelo de conjuntos de muestras de tejido de control-casos. Los adenomas y los pólipos SSP son las lesiones precancerosas críticas para el CCR 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 secuenciación con bisulfito de representación reducida (RRBS). Los análisis terciarios y cuaternarios son únicos e integrales para el proceso de selección de marcadores. La etapa terciaria implica analizar los datos en términos de límites de cobertura, excluir todos los sitios no informativos, contrastar el % de metilación entre los subgrupos de diagnóstico mediante regresión logística, crear “islas” de CpG in silico basadas en agrupaciones definidas de sitios de metilación contiguos y el análisis de las características operativas del receptor. La etapa cuaternaria filtra los datos del % de metilación (tanto los CpG individuales como los grupos in silico) para seleccionar marcadores que maximicen las relaciones señal a ruido, minimicen la metilación de fondo, tengan en cuenta la heterogeneidad del tumor y enfatizan el rendimiento de la ROC. Además, este enfoque analítico asegura la identificación de los CpG de puntos críticos en una longitud de ADN definida para facilitar el desarrollo y el diseño óptimo de los ensayos de marcadores posteriores: PCR específica de metilación, secuenciación profunda de fragmentos pequeños, etc.

Las muestras de CCR, adenoma y control arrojaron aproximadamente 2-3 millones de CpG de alta calidad, lo que tras el análisis y el filtrado dio como resultado aproximadamente 1068 (CCR) y 1200 (adenoma) sitios individuales altamente discriminatorios. Estas se agruparon en 185 regiones localizadas de metilación diferencial para el CCR y 244 para el adenoma, algunas con una extensión de 30-40 bases y otras de más de una kilobase. Según una revisión superficial de la bibliografía, menos del 15 % de las regiones DMR parecen tener una asociación previa con el 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 muy débiles o nulas con el cáncer, aunque muchas están implicadas en vías funcionales que pueden ser relevantes para la tumorigénesis. 30 regiones DMR (CCR) y 42 (adenoma) no tenían ninguna anotación ni referencias en ninguna parte de la bibliografía. Esas se denominaron MAX seguido de la ubicación cromosómica y las coordenadas. Todas las regiones DMR tenían ABC de 0,85 o más (algunas demostraron una discriminación perfecta entre casos y controles con un ABC de 1,0). Además de alcanzar el umbral ABC de 0,85, los marcadores de cáncer de colon y adenoma que se identificaron tenían que mostrar una densidad de metilación $>$ de 50 veces mayor en el tumor que en la mucosa gástrica o del colon normal y una metilación $<$ 1,0 % en la mucosa colónica normal. Basándose en la experiencia previa en el descubrimiento y la validación de marcadores del cáncer de páncreas, estas características de los marcadores en el tejido predicen una alta discriminación en medios distantes, tales como las heces o la sangre.

Las muestras de pólipos SSP se derivaron de bloques FFPE y eran de menor calidad. Para estas, solo la mitad tuvo un número adecuado de lecturas (>100 000). Como tal, la rigurosidad de filtrado se redujo ligeramente en comparación con las utilizadas con las muestras congeladas. Tras la clasificación, había 268 sitios discriminatorios que se agruparon en 111 regiones localizadas.

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

Tabla 3A: Regiones DMR del cáncer colorrectal

ES 2 980 778 T3

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
5	10564655-10564807	ANKRD33B
11	110582796-110583345	ARHGAP20
11	110582170-110582657	ARHGAP20
11	110581912-110582039	ARHGAP20
12	103351885-103352327	ASCL1
13	25946118-25946206	ATP8A2
8	104152806-104153145	BAALC
9	96715112-96715603	BARX1
8	65493937-65494105	BHLHE22
6	7727566-7728088	BMP6
6	105584685-105585220	BVES
17	43339242-43339498	C17orf46
1	1475560-1475650	C1orf70
1	1476065-1476127	C1orf70
10	16562866-16563332	C1QL3
10	16563667-16563892	C1QL3
10	16562465-16562672	C1QL3
9	132382813-132382909	C9orf50
18	70211543-70211719	CBLN2
3	128720801-128720885	CCDC48
3	128719995-128720631	CCDC48
2	101033758-101034005	CHST10
12	104850745-104851001	CHST11
3	142838025-142838494	CHST2
3	142838645-142839023	CHST2
3	142839223-142839576	CHST2
5	178016833-178017456	COL23A1
7	30721941-30722028	CRHR2
9	124461296-124461420	DAB2IP
12	64062131-64062443	DPY19L2
12	113494586-113494957	DTX1
10	64575060-64575283	EGR2
2	31456804-31457263	EHD3
7	37487755-37488565	ELMO1
7	37487539-37487623	ELMO1
6	11044395-11044834	ELOVL2
6	80656845-80657306	ELOVL4
6	152129389-152129636	ESR1
6	133562485-133562878	EYA4
15	48938056-48938252	FBN1
11	128563956-128564209	FLI1
11	128562780-128563522	FLI1

ES 2 980 778 T3

Cromosoma	Coordenadas cromosómicas	Anotación de genes
6	159590083-159590220	FNDC1
9	101471421-101471519	GABBR2
2	31360809-31360992	GALNT14
2	31360542-31360640	GALNT14
11	134146132-134146380	GLB1L3
7	42276418-42277414	GLI3
12	52400569-52400726	GRASP
12	52400919-52401166	GRASP
16	28074472-28074761	GSG1L
17	1959348-1959370	HIC1
7	50343838-50344453	IKZF1
2	182321830-182321983	ITGA4
1	226925082-226925651	ITPKB
4	6201350-6201560	JAKMIP1
2	47797187-47797452	KCNK12
2	149633039-149633137	KIF5C
7	149411729-149411847	KRBA1
19	8274584-8274671	LASS4
2	30455594-30455705	LBH
5	38556357-38556743	LIFR
19	2290273-2290393	LINGO3
19	2290645-2290738	LINGO3
19	42905798-42906349	LIPE
7	140772542-140772873	LOC100131199
2	100938402-100939005	LONRF2
7	127671918-127672318	LRRC4
19	41119795-41119907	LTBP4
6	6546375-6546598	LY86-AS1
19	3785828-3786371	MATK
10	22541502-22541671	MAX.chr1 0.22541502-22541671
10	22541884-22542001	MAX.chr1 0.22541884-22542001
10	22765155-22765223	MAX.chr1 0.22765155-22765223
11	123301058-123301255	MAX.chr1 1.123301058-123301255
11	123301366-123301506	MAX.chr1 1.123301366-123301506
11	123301853-123301941	MAX.chr1 1.123301853-123301941
11	57250528-57250611	MAX.chr1 1.57250528-57250611
12	8171360-8171769	MAX.chr2 2.8171360-8171769
14	100437680-100437767	MAX.chr4 4.100437680-100437767
19	22034447-22034696	MAX.chr9 9.22034447-22034696
19	22034799-22034887	MAX.chr9 9.22034799-22034887
19	42444999-42445053	MAX.chr9 9.42444999-42445053
2	144694517-144695025	MAX.chr2 2.144694517-144695025
20	44936022-44936246	M AX.chr20.44936022-44936246
3	13324501-13324623	MAX.chr3 3.13324501-13324623
3	13324760-13324864	MAX.chr3 3.13324760-13324864
3	44039952-44040054	M AX.chr3.44039952-44040054
7	142494755-142494915	MAX.chr7 7.142494755-142494915
8	30769438-30769680	M AX.chr8.30769438-30769680
9	99983730-99984118	M AX.chr9.99983730-99984118
6	41606074-41606126	MDF1
3	150804938-150804971	MED12L
5	88185490-88185589	MEF2C
22	39853199-39853295	MGAT3

ES 2 980 778 T3

Cromosoma	Coordenadas cromosómicas	Anotación de genes
2	220416703-220417434	MIR3132
6	132722283-132722484	MOXD1
8	72755813-72756349	MSC
16	58497251-58497370	NDRG4
19	3361105-3361330	NFIC
17	47573986-47574084	NGFR
7	108095348-108095805	NRCAM
8	32406662-32406901	NRG1
17	8925482-8925838	NTN1
15	53082447-53083044	ONECUT1
8	145106742-145106921	OPLAH
8	145106349-145106456	OPLAH
5	76506245-76506578	PDE8B
11	104034769-104034920	PDGFD
22	45405722-45405819	PHF21B
8	79428485-79428684	PKIA
1	150122783-150123157	PLEKHO1
1	38510915-38511213	POU3F1
17	56833684-56833978	PPM1E
20	37434246-37434800	PPP1R16B
14	102248062-102248216	PPP2R5C
14	102247525-102247929	PPP2R5C
16	23847825-23848168	PRKCB
19	47778181-47778372	PRR24
6	163834534-163834925	QKI
18	9708397-9709392	RAB31
20	4803201-4803703	RASSF2
2	161263880-161264733	RBMS1
6	127440413-127441057	RSPO3
17	26698693-26699117	SARM1
8	97505964-97506676	SDC2
17	75369224-75369327	Septin9
17	75368800-75369056	Septin9
10	7452746-7452779	SFMBT2
10	7452885-7452956	SFMBT2
10	7452029-7452452	SFMBT2
10	7450242-7450831	SFMBT2
10	7451097-7451185	SFMBT2
14	70655516-70655712	SLC8A3
5	101632152-101632237	SLCO4C1
7	128829103-128829184	SMO
11	65601167-65601514	SNX32
13	95363646-95363959	SOX21
12	24715703-24715776	SOX5
12	24715012-24715416	SOX5
12	24716178-24716294	SOX5
17	70114081-70114176	SOX9
7	75896637-75896925	SRRM3
6	166581771-166582044	T
12	65218900-65218994	TBC1D30
12	65218335-65218778	TBC1D30
8	67874670-67875083	TCF24
18	53255390-53255565	TCF4

ES 2 980 778 T3

Cromosoma	Coordenadas cromosómicas	Anotación de genes
5	1294873-1295322	TERT
21	32930226-32930576	TIAM1
21	32716063-32716545	TIAM1
2	74741941-74742264	TLX2
15	83776196-83776373	TM6SF1
2	135476019-135476390	TMEM163
7	19156788-19156858	TWIST1
1	108507609-108507674	VAV3
1	108507074-108507498	VAV3
5	82768837-82769031	VCAN
2	175547056-175547390	WIPF1
8	10873760-10874271	XKR6
8	10872819-10873619	XKR6
10	31609049-31609227	ZEB1
2	145274517-145274600	ZEB2
2	145274704-145275062	ZEB2
19	58951402-58951530	ZNF132
19	54024023-54024436	ZNF331
19	53661526-53662618	ZNF347
19	22018452-22018639	ZNF43
16	88496963-88497197	ZNF469
19	37407197-37407365	ZNF568
19	12267378-12267677	ZNF625
19	12203466-12203641	ZNF788
2	185463105-185463763	ZNF804A
19	53970869-53971374	ZNF813

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

Anotación de genes	Cromosoma	Cromosoma Coordenada	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.chr1 0.22541884-22542001	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

ES 2 980 778 T3

LINGO3	19	2290273-2290393	0,9830
MATK	19	3785828-3786371	0,9830
RSP03	6	127440413-127441057	0,9830
MGAT3	22	39853199-39853295	0,9828
GRASP	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
MAX.chr1.123301058-123301255	11	123301058-123301255	0,9673
SOX9	17	70114081-70114176	0,9665
MAX.chr20.44936022-44936246	20	44936022-44936246	0,9659
TWIST1	7	19156788-19156858	0,9652
ELMO1	7	37487539-37487623	0,9644
NFIC	19	3361105-3361330	0,9644
PLEKHO1	1	150122783-150123157	0,9644
POU3F1	1	38510915-38511213	0,9644
NGFR	17	47573986-47574084	0,9633
LINGO3	19	2290645-2290738	0,9624
AEBP1	7	44143993-44144413	0,9613
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
GABBR2	9	101471421-101471519	0,9542
PRKCB	16	23847825-23848168	0,9539
RAB31	18	9708397-9709392	0,9536
VAV3	1	108507074-108507498	0,9536
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
ZNF568	19	37407197-37407365	0,9505
C1orf70	1	1475560-1475650	0,9497

ES 2 980 778 T3

ITGA4	2	182321830-182321983	0,9495
GLB1L3	11	134146132-134146380	0,9493
MAX.chr12.8171360-8171769	12	8171360-8171769	0,9489
MAX.chr14.100437680-100437767	14	100437680-100437767	0,9481
ZNF132	19	58951402-58951530	0,9479
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
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
MED12L	3	150804938-150804971	0,9387
MDFI	6	41606074-41606126	0,9381
MAX.chr1.123301366-123301506	11	123301366-123301506	0,9381
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
TCF24	8	67874670-67875083	0,9334
MAX.chr1.123301853-123301941	11	123301853-123301941	0,9321
SDC2	8	97505964-97506676	0,9319
TERT	5	1294873-1295322	0,9319
ELOVL2	6	11044395-11044834	0,9303
GALNT14	2	31360809-31360992	0,9298
Septin9	17	75369224-75369327	0,9289
GLI3	7	42276418-42277414	0,9288
SOX5	12	24715703-24715776	0,9286
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
CCDC48	3	128719995-128720631	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
ZNF347	19	53661526-53662618	0,9173
MAX.chr1.22034447-22034696	19	22034447-22034696	0,9164
NRCAM	7	108095348-108095805	0,9164
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
BVES	6	105584685-105585220	0,9118
CBLN2	18	70211543-70211719	0,9102
MAX.chr1.42444999-42445053	19	42444999-42445053	0,9080
BARX1	9	96715112-96715603	0,9071
GRASP	12	52400569-52400726	0,9071

ES 2 980 778 T3

PRR24	19	47778181-47778372	0,9040
MAX.chr3.13324501-13324623	3	13324501-13324623	0,9034
MAX.chr1.57250528-57250611	11	57250528-57250611	0,9031
C1QL3	10	16563667-16563892	0,9025
MAX.chr3.13324760-13324864	3	13324760-13324864	0,9003
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
SLC8A3	14	70655516-70655712	0,8922
NTN1	17	8925482-8925838	0,8916
MAX.chr1.0.22765155-22765223	10	22765155-22765223	0,8916
MAX.chr1.0.22541502-22541671	10	22541502-22541671	0,8884
MEF2C	5	88185490-88185589	0,8870
TBC1D30	12	65218335-65218778	0,8870
MAX.chr9.99983730-99984118	9	99983730-99984118	0,8854
MAX.chr9.22034799-22034887	19	22034799-22034887	0,8839
SOX21	13	95363646-95363959	0,8839
ZNF804A	2	185463105-185463763	0,8839
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
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
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
RBMS1	2	161263880-161264733	0,8715
ARHGAP20	11	110581912-110582039	0,8684
LIPE	19	42905798-42906349	0,8684
SRRM3	7	75896637-75896925	0,8684
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
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
ZNF331	19	54024023-54024436	0,8545
PHF21B	22	45405722-45405819	0,8541
EGR2	10	64575060-64575283	0,8529
ZNF813	19	53970869-53971374	0,8522
CHST11	12	104850745-104851001	0,8514
SOX5	12	24715012-24715416	0,8514
JAKMIP1	4	6201350-6201560	0,8506

ES 2 980 778 T3

Tabla 4A: Regiones DMR de adenomas grandes

Cromosoma	Coordenadas cromosómicas	Anotación de genes
7	45613877-45613977	ADCY1
2	70994754-70995045	ADD2
22	24820148-24820373	ADORA2A
2	100721643-100721967	AFF3
1	49242089-49242514	AGBL4
6	151561236-151561473	AKAP12
6	151561598-151561873	AKAP12
7	134142981-134143723	AKR1B1
8	41754327-41754726	ANK1
5	10564406-10564551	ANKRD33B
5	10564655-10564807	ANKRD33B
5	10565042-10565191	ANKRD33B
11	110582796-110583345	ARHGAP20
12	103351885-103352327	ASCL1
8	104152806-104153145	BAALC
8	65494269-65494355	BHLHE22
6	7727566-7728088	BMP6
6	105584685-105585220	BVES
12	21680721-21680828	C12orf39
12	48577334-48577557	C12orf68
16	4588091-4588817	C16orf5
1	1475560-1475650	C1orf70
1	1476065-1476127	C1orf70
10	16562866-16563332	C1QL3
10	16563667-16563892	C1QL3
10	16562465-16562672	C1QL3
20	3388089-3388291	C20orf194
6	74019826-74019955	C6orf147
9	132382813-132382909	C9orf50
7	44364925-44365359	CAMK2B
5	110559508-110560719	CAMK4
3	12838197-12838303	CAND2
18	70211543-70211719	CBLN2
6	74405903-74406086	CD109
12	133464655-133464819	CHFR
11	45686306-45686534	CHST1
2	101033758-101034005	CHST10
12	104851372-104851465	CHST11
12	104850745-104851001	CHST11
10	125852012-125852098	CHST15
10	125852559-125852792	CHST15
10	125852905-125853007	CHST15
10	125851544-125851700	CHST15
3	142838025-142838494	CHST2
3	142838645-142839023	CHST2
3	142839223-142839576	CHST2
3	139654045-139654299	CLSTN2
5	178016833-178017456	COL23A1
9	124461296-124461420	DAB2IP
3	186079767-186080092	DGKG
1	65731412-65731782	DNAJC6

ES 2 980 778 T3

2	225906664-225906922	DOCK10
12	64062131-64062443	DPY19L2
12	113494586-113494957	DTX1
2	233352345-233352431	ECEL1
10	64575060-64575283	EGR2
2	31456804-31457263	EHD3
7	37487755-37488565	ELMO1
6	11044395-11044834	ELOVL2
6	80656845-80657306	ELOVL4
6	152129389-152129636	ESR1
6	133562229-133562380	EYA4
6	133562485-133562878	EYA4
7	23053043-23053438	FAM126A
1	53098973-53099237	FAM159A
1	206137408-206137473	FAM72A
1	120839339-120839381	FAM72B
1	120838272-120838775	FAM72B
1	120836675-120836768	FAM72B
1	27960931-27961018	FGR
11	128563956-128564209	FLI1
11	128562780-128563522	FLI1
14	62584068-62584109	FLJ43390
13	28674199-28674862	FLT3
9	101471421-101471519	GABBR2
2	31360809-31360992	GALNT14
17	10102237-10102576	GAS7
19	19006296-19006511	GDF1
5	179780627-179781188	GFPT2
5	137610023-137610333	GFRA3
11	134146132-134146380	GLB1L3
7	42276418-42277414	GLI3
1	54204505-54204712	GLIS1
12	52400919-52401166	GRASP
10	88125930-88126495	GRID1
7	6570511-6570865	GRID2IP
16	28074472-28074761	GSG1L
6	32632785-32632860	HLA-DQB1
15	83621302-83621657	HOMER2
7	50343838-50344453	IKZF1
2	182321830-182321983	ITGA4
21	46351838-46352381	ITGB2
1	226925082-226925651	ITPKB
4	6202051-6202410	JAKMIP1
4	6201350-6201560	JAKMIP1
2	47797187-47797452	KCNK12
20	62103225-62103324	KCNQ2
6	73972941-73973104	KHDC1
2	149633039-149633137	KIF5C
7	149411729-149411847	KRBA1
19	8274360-8274430	LASS4
19	8274584-8274671	LASS4
19	2290273-2290393	LINGO3
19	42905798-42906349	LIPE
11	8284746-8284871	LMO1

ES 2 980 778 T3

7	140773610-140773855	LOC100131199
7	140772542-140772873	LOC100131199
2	100938402-100939005	LONRF2
7	127671918-127672318	LRRC4
19	41119795-41119907	LTBP4
6	6546375-6546598	LY86-AS1
11	63828346-63828436	MACROD1
19	3785828-3786371	MATK
1	244012766-244012875	MAX.chrl.244012766-244012875
1	244013190-244013393	MAX.chrl.244013190-244013393
1	39269813-39270150	MAX.chrl.39269813-39270150
10	22541502-22541671	MAX.chrl 0.22541502-22541671
10	22541884-22542001	MAX.chrl 0.22541884-22542001
10	22765155-22765223	MAX.chrl 0.22765155-22765223
11	120435350-120435981	MAX.chrl 1.120435350-120435981
11	123301058-123301255	MAX.chrl 1.123301058-123301255
11	123301366-123301506	MAX.chrl 1.123301366-123301506
11	123301853-123301941	MAX.chrl 1.123301853-123301941
11	44749119-44749205	MAX.chrl 1.44749119-44749205
11	8040551-8040677	MAX.chrl 1.8040551-8040677
12	133484966-133485857	MAX.chrl 2.133484966-133485857
14	100437680-100437767	MAX.chrl 4.100437680-100437767
14	105400087-105400182	MAX.chrl 4.105400087-105400182
15	34806855-34807014	MAX.chrl 5.34806855-34807014
18	77558550-77558609	MAX.chrl 8.77558550-77558609
19	20959229-20959691	MAX.chrl 9.20959229-20959691
19	22034447-22034696	MAX.chrl 9.22034447-22034696
19	22034799-22034887	MAX.chrl 9.22034799-22034887
2	144694517-144695025	MAX.chr2.144694517-144695025
21	47063135-47064177	MAX.chr21.47063135-47064177
3	115231555-115231576	MAX.chr3.115231555-115231576
3	44039952-44040054	M AX.chr3.44039952-44040054
8	30769438-30769680	M AX.chr8.30769438-30769680
6	37664238-37664539	MDGA1
5	88185490-88185589	MEF2C
22	39853199-39853295	MGAT3
17	74864552-74864821	MGAT5B
3	154797723-154797909	MME
6	132722283-132722484	MOXD1
17	8533282-8534168	MYH10
11	112832731-112832815	NCAM1
16	58497979-58498250	NDRG4
17	47573986-47574084	NGFR
8	41503949-41504137	NKX6-3
5	142784971-142785160	NR3C1
7	108095348-108095805	NRCAM
17	8925482-8925838	NTN1
1	107683961-107684314	NTNG1
1	107683064-107683372	NTNG1
1	107684447-107684545	NTNG1
3	13461109-13461191	NUP210
11	79150971-79151076	ODZ4
15	53082447-53083044	ONECUT1
8	145106742-145106921	OPLAH

ES 2 980 778 T3

8	145106349-145106456	OPLAH
3	142682282-142682813	PAQR9
5	140855415-140856027	PCDHGA1
5	76506245-76506578	PDE8B
11	104034769-104034920	PDGFD
4	55099106-55099473	PDGFRA
1	9711854-9711974	PIK3CD
8	79428485-79428684	PKIA
1	150122783-150123157	PLEKHO1
1	38510915-38511213	POU3F1
17	56833684-56833978	PPM1E
14	102248062-102248216	PPP2R5C
14	102247525-102247929	PPP2R5C
5	122425730-122425886	PRDM6
20	47444582-47444776	PREX1
16	23847825-23848168	PRKCB
16	23846951-23847056	PRKCB
8	30890580-30890912	PURG
6	163834534-163834925	QKI
6	163835376-163835472	QKI
20	4803969-4804077	RASSF2
20	4803201-4803703	RASSF2
18	56936593-56936656	RAX
15	93631919-93632242	RGMA
9	77111900-77112005	RORB
6	127440413-127441057	RSPO3
17	26698693-26699117	SARM1
8	97505964-97506676	SDC2
7	3339895-3340903	SDK1
19	40005836-40005892	SELV
17	75368800-75369056	Septin9
22	26565137-26565417	SEZ6L
10	7452746-7452779	SFMBT2
10	7452885-7452956	SFMBT2
10	7452029-7452452	SFMBT2
10	7450242-7450831	SFMBT2
10	7451097-7451185	SFMBT2
6	118228394-118228979	SLC35F1
14	7065516-70655712	SLC8A3
14	70655268-70655368	SLC8A3
11	65601167-65601514	SNX32
13	95363646-95363959	SOX21
12	24715703-24715776	SOX5
12	24715012-24715416	SOX5
12	24716178-24716294	SOX5
8	10587893-10588143	SOX7
7	75896637-75896925	SRRM3
8	134583587-134583963	ST3GAL1
12	65218900-65218994	TBC1D30
17	45810562-45810819	TBX21
8	67874670-67875083	TCF24
18	53255390-53255565	TCF4
21	32931523-32931688	TIAM1
21	32930226-32930576	TIAM1

ES 2 980 778 T3

15	83776196-83776373	TM6SF1
2	135476019-135476390	TMEM163
2	39893089-39893224	TMEM178
2	12857915-12858230	TRIB2
7	19156788-19156858	TWIST1
1	213124472-213124778	VASH2
1	108507609-108507674	VAV3
1	108507074-108507498	VAV3
5	82768837-82769031	VCAN
10	17271896-17271978	VIM
10	17270955-17271052	VIM
13	27131683-27131757	WASF3
2	175547056-175547390	WIPF1
1	228195339-228195413	WNT3A
8	10873760-10874271	XKR6
8	10872819-10873619	XKR6
10	31608798-31608892	ZEB1
10	31608394-31608690	ZEB1
10	31609049-31609227	ZEB1
19	58951402-58951530	ZNF132
4	332064-332199	ZNF141
17	16472295-16472694	ZNF287
19	54024023-54024436	ZNF331
19	53661526-53662618	ZNF347
19	22018746-22019004	ZNF43
16	88496963-88497197	ZNF469
19	37064200-37064435	ZNF529
19	37960066-37960505	ZNF569
19	12267378-12267677	ZNF625
19	20149796-20149923	ZNF682
2	185463105-185463763	ZNF804A

Tabla 4B: Regiones DMR de adenomas grandes clasificadas por área bajo la curva ROC

Anotación de genes	Cromosoma	Coordenada cromosómica	Área bajo la curva ROC
ADD2	2	70994754-70995045	1,0000
AGBL4	1	49242089-49242514	1,0000
AKAP12	6	151561598-151561873	1,0000
ANKRD33B	5	10565042-10565191	1,0000
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
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
GRASP	12	52400919-52401166	1,0000
ITGA4	2	182321830-182321983	1,0000
KCNQ2	20	62103225-62103324	1,0000
LOC100131199	7	140772542-140772873	1,0000
LONRF2	2	100938402-100939005	1,0000
MGAT3	22	39853199-39853295	1,0000

ES 2 980 778 T3

OPLAH	8	145106742-145106921	1,0000
OPLAH	8	145106349-145106456	1,0000
PDE8B	5	76506245-76506578	1,0000
PDGFD	11	104034769-104034920	1,0000
PKIA	8	79428485-79428684	1,0000
POU3F1	1	38510915-38511213	1,0000
QKI	6	163834534-163834925	1,0000
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
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
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
ANKRD33B	5	10564655-10564807	0,9983
MAX.chr1 0.22541502-22541671	10	22541502-22541671	0,9982
MAX.chr1 1.44749119-44749205	11	44749119-44749205	0,9982
GALNT14	2	31360809-31360992	0,9980
AKR1B1	7	134142981-134143723	0,9967
CHST2	3	142839223-142839576	0,9967
EYA4	6	133562229-133562380	0,9967
ZNF625	19	12267378-12267677	0,9967
LMO1	11	8284746-8284871	0,9965
ZNF469	16	88496963-88497197	0,9964
ESR1	6	152129389-152129636	0,9951
KCNK12	2	47797187-47797452	0,9951
MAX.chr1 1.8040551-8040677	11	8040551-8040677	0,9951
MOXD1	6	132722283-132722484	0,9951
PPP2R5C	14	102247525-102247929	0,9951
PPP2R5C	14	102248062-102248216	0,9949
CHST11	12	104850745-104851001	0,9948
LASS4	19	8274584-8274671	0,9945
GSG1L	16	28074472-28074761	0,9935
MAX.chr1 1.120435350-120435981	11	120435350-120435981	0,9935
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
MAX.chr9 22034447-22034696	19	22034447-22034696	0,9926
ZNF347	19	53661526-53662618	0,9921
MAX.chr8 30769438-30769680	8	30769438-30769680	0,9918
SOX5	12	24715012-24715416	0,9918
CHST15	10	125852905-125853007	0,9913
ODZ4	11	79150971-79151076	0,9913
SOX21	13	95363646-95363959	0,9908
SEZ6L	22	26565137-26565417	0,9902

ES 2 980 778 T3

GAS7	17	10102237-10102576	0,9899
MAX.chrl 1.123301853-123301941	11	123301853-123301941	0,9889
ELMO1	7	37487755-37488565	0,9886
VIM	10	17270955-17271052	0,9886
WNT3A	1	228195339-228195413	0,9886
SLC8A3	14	70655516-70655712	0,9879
PLEKHO1	1	150122783-150123157	0,9869
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
SOX5	12	24715703-24715776	0,9847
EHD3	2	31456804-31457263	0,9837
TMEM163	2	135476019-135476390	0,9837
MAX.chrl 4.105400087-105400182	14	105400087-105400182	0,9835
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
WIPF1	2	175547056-175547390	0,9820
GRID1	10	88125930-88126495	0,9804
MAX.chrl 1.123301366-123301506	11	123301366-123301506	0,9804
RASSF2	20	4803969-4804077	0,9804
TCF4	18	53255390-53255565	0,9804
TRIB2	2	12857915-12858230	0,9804
ZNF331	19	54024023-54024436	0,9804
SLC35F1	6	118228394-118228979	0,9794
COL23A1	5	178016833-178017456	0,9788
FAM159A	1	53098973-53099237	0,9788
GABBR2	9	101471421-101471519	0,9788
CHST2	3	142838645-142839023	0,9775
MAX.chrl 9.22034799-22034887	19	22034799-22034887	0,9775
CHST10	2	101033758-101034005	0,9771
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
CHST1	11	45686306-45686534	0,9755
MAX.chrl 8.77558550-77558609	18	77558550-77558609	0,9753
LINGO3	19	2290273-2290393	0,9740
LASS4	19	8274360-8274430	0,9739
ZEB1	10	31609049-31609227	0,9722
ZNF43	19	22018746-22019004	0,9722
C12orf39	12	21680721-21680828	0,9706
MAX.chrl 0.22541884-22542001	10	22541884-22542001	0,9706
MAX.chrl 1.123301058-123301255	11	123301058-123301255	0,9706
VIM	10	17271896-17271978	0,9699
AKAP12	6	151561236-151561473	0,9690
C16orf5	16	4588091-4588817	0,9690
RORB	9	77111900-77112005	0,9690
NRCAM	7	108095348-108095805	0,9689

ES 2 980 778 T3

ELOVL4	6	80656845-80657306	0,9683
CHST2	3	142838025-142838494	0,9673
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
C9orf50	9	132382813-132382909	0,9659
MAX.chr1.2.133484966-133485857	12	133484966-133485857	0,9657
PURG	8	30890580-30890912	0,9643
AFF3	2	100721643-100721967	0,9641
PDGFRA	4	55099106-55099473	0,9641
M AX.chr3.44039952-44040054	3	44039952-44040054	0,9637
TWIST1	7	19156788-19156858	0,9632
MEF2C	5	88185490-88185589	0,9608
VCAN	5	82768837-82769031	0,9602
NCAM1	11	112832731-112832815	0,9600
CAMK4	5	110559508-110560719	0,9592
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
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
SDK1	7	3339895-3340903	0,9510
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
TM6SF1	15	83776196-83776373	0,9462
MAX.chr1.244013190-244013393	1	244013190-244013393	0,9461
MAX.chr1.39269813-39270150	1	39269813-39270150	0,9461
MAX.chr4.100437680-100437767	14	100437680-100437767	0,9446
EGR2	10	64575060-64575283	0,9444
SOX7	8	10587893-10588143	0,9428
LRRC4	7	127671918-127672318	0,9395
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
JAKMIP1	4	6201350-6201560	0,9360
ANK1	8	41754327-41754726	0,9346
FAM126A	7	23053043-23053438	0,9346
SFMBT2	10	7452885-7452956	0,9338
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

ES 2 980 778 T3

ARHGAP20	11	110582796-110583345	0,9248
MAX.chr1 0.22765155-22765223	10	22765155-22765223	0,9242
ZNF141	4	332064-332199	0,9221
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
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
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
PRKCB	16	23846951-23847056	0,9053
GLI3	7	42276418-42277414	0,9036
Septin9	17	75368800-75369056	0,9031
FAM72A	1	206137408-206137473	0,9016
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
MDGA1	6	37664238-37664539	0,8987
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
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
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
DAB2IP	9	124461296-124461420	0,8787
CHST15	10	125852012-125852098	0,8780
C6orf147	6	74019826-74019955	0,8775
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
MAX.chr1 5.34806855-34807014	15	34806855-34807014	0,8722
CAND2	3	12838197-12838303	0,8719
TBX21	17	45810562-45810819	0,8702
MGAT5B	17	74864552-74864821	0,8685

ES 2 980 778 T3

ZNF569	19	37960066-37960505	0,8676
NTNG1	1	107684447-107684545	0,8672
WASF3	13	27131683-27131757	0,8647
LIPE	19	42905798-42906349	0,8644
DOCK10	2	225906664-225906922	0,8578
LTBP4	19	41119795-41119907	0,8546
MAX.chr1 9.20959229-20959691	19	20959229-20959691	0,8521
FGR	1	27960931-27961018	0,8513

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

Cromosoma	Coordenadas cromosómicas	Anotación de genes
5	10564655-10564710	ANKRD33B
12	103351885-103351983	ASCL1
1	203598574-203598800	ATP2B4
6	91005003-91005091	BACH2
6	7727026-7727129	BMP6
6	105584890-105584983	BVES
10	21784521-21784567	C10orf114
1	226737152-226737231	C1orf95
5	110559571-110559638	CAMK4
2	56411545-56411640	CCDC85A
1	158150837-158150885	CD1D
1	158151102-158151205	CD1D
10	90967004-90967028	CH25H
2	101033768-101033858	CHST10
12	104850745-104850879	CHST11
3	142838194-142838411	CHST2
9	34590231-34590344	CNTFR
12	49484149-49484231	DHH
15	30484732-30484813	DKFZP434L187
1	65731412-65731530	DNAJC6
2	225906664-225906763	DOCK10
2	225907515-225907632	DOCK10
2	73520956-73520964	EGR4
6	80656889-80656974	ELOVL4
6	80657208-80657306	ELOVL4
5	111754713-111754810	EPB41L4A
3	96532270-96532344	EPHA6
7	23053937-23054066	FAM126A
11	125365196-125365327	FEZ1
13	22243643-22243727	FGF9
7	90894876-90894960	FZD1
17	42907793-42907827	GJC1
3	179169408-179169505	GNB4
1	101005577-101005661	GPR88
1	53068071-53068182	GPX7
7	6570755-6570845	GRID2IP
16	10275378-10275472	GRIN2A
17	14205388-14205498	HS3ST3B1
7	23509037-23509225	IGF2BP3
6	39281409-39281488	KCNK17
15	79724426-79724525	KIAA1024
2	208031024-208031104	KLF7

ES 2 980 778 T3

2	208031731-208031826	KLF7
2	30454421-30454492	LBH
2	30454871-30454977	LBH
2	74726179-74726265	LBX2
5	87970308-87970374	LOC645323
5	87970772-87970894	LOC645323
2	170220089-170220148	LRP2
12	40618617-40618655	LRRK2
1	25944147-25944152	MAN1C1
11	123301366-123301387	MAX.chr1.123301366-123301387
17	45867397-45867662	MAX.chr17.45867397-45867662
19	55963254-55963329	MAX.chr19.55963254-55963329
2	127783352-127783403	MAX.chr2.127783352-127783403
2	96192422-96192521	M AX.chr2.96192422-96192521
20	1783778-1783841	M AX.chr20.1783778-1783841
22	42310340-42310438	MAX.chr22.42310340-42310438
3	43935668-43935753	M AX.chr3.43935668-43935753
4	186049639-186049660	MAX.chr4.186049639-186049660
6	114664537-114664631	M AX.chr6.114664537-114664631
7	127807622-127807693	MAX.chr7.127807622-127807693
7	149745500-149745592	MAX.chr7.149745500-149745592
9	114074-114160	MAX.chr9.114074-114160
9	114354-114435	MAX.chr9.114354-114435
9	99983903-99984118	M AX.chr9.99983903-99984118
6	37664654-37664664	MDGA1
15	66546092-66546108	MEGF11
15	82339716-82339790	MEX3B
11	30607877-30607973	MPPED2
4	113437673-113437953	NEUROG2
1	153651840-153651933	NPR1
5	142785023-142785050	NR3C1
8	32406662-32406739	NRG1
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

ES 2 980 778 T3

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

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

Anotación de genes	Cromosoma	Coordenada cromosómica	Área bajo la curva ROC
CAMK4	5	110559571-110559638	1,0000
FGF9	13	22243643-22243727	1,0000
GJC1	17	42907793-42907827	1,0000
GPX7	1	53068071-53068182	1,0000
GRIN2A	16	10275378-10275472	1,0000
IGF2BP3	7	23509037-23509225	1,0000
MAX.chr1.9.55963254-55963329	19	55963254-55963329	1,0000
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
PIK3CD	1	9711931-9711974	1,0000
PTENP1	9	33677215-33677313	1,0000
RAB31	18	9708795-9708891	1,0000
RNF220	1	44872395-44872487	1,0000
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
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
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
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
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
CD1D	1	158150837-158150885	0,9618

ES 2 980 778 T3

ATP2B4	1	203598574-203598800	0,9559
C10orf114	10	21784521-21784567	0,9496
KCNK17	6	39281409-39281488	0,9485
FEZ1	11	125365196-125365327	0,9429
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
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
EPB41L4A	5	111754713-111754810	0,9034
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
RORB	9	77111911-77112005	0,8905
ZEB1	10	31608625-31608690	0,8897
NEUROG2	4	113437673-113437953	0,8856
TCF4	18	52989026-52989191	0,8807
M AX.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
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
LRRK2	12	40618617-40618655	0,8697
BACH2	6	91005003-91005091	0,8687
CHST10	2	101033768-101033858	0,8676
CHST2	3	142838194-142838411	0,8676
M AX.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
ASCL1	12	103351885-103351983	0,8603
KLF7	2	208031024-208031104	0,8600
CNTFR	9	34590231-34590344	0,8592
M AX.chr6.114664537-114664631	6	114664537-114664631	0,8592
SFMBT2	10	7450571-7450659	0,8585
TNFSF11	13	43148769-43148861	0,8571
DNAJC6	1	65731412-65731530	0,8550
GRID2IP	7	6570755-6570845	0,8550
BMP6	6	7727026-7727129	0,8474
MAX.chr1.123301366-123301387	11	123301366-123301387	0,8471
PPP2R5C	14	102247811-102247881	0,8466
SERHL2	22	42949849-42949919	0,8464
RCSD1	1	167599730-167599772	0,8456

ES 2 980 778 T3

	M AX.chr2.96192422-96192521	2	96192422-96192521	0,8431
	ANKRD33B	5	10564655-10564710	0,8407
	NPR1	1	153651840-153651933	0,8384
5	EPHA6	3	96532270-96532344	0,8267
	FAM126A	7	23053937-23054066	0,8262
	CH25H	10	90967004-90967028	0,8235
	KIAA1024	15	79724426-79724525	0,8235
10	LOC645323	5	87970308-87970374	0,8235
	ZNF329	19	58666209-58666308	0,8224
	UST	6	149069140-149069222	0,8199
	CCDC85A	2	56411545-56411640	0,8193
15	PLEKHOL	1	150122951-150122989	0,8176
	SLC35F1	6	118228394-118228493	0,8082
	LBX2	2	74726179-74726265	0,8059
	DHH	12	49484149-49484231	0,8046
20	MAX.chr7.127807622-127807693	7	127807622-127807693	0,8007
	C1orf95	1	226737152-226737231	0,8000

Ejemplo 3.

Este ejemplo demuestra que NDRG4, BMP3, OPLAH, FLI1, PDGFD, CHST_7889, SFMBT2_895, SFMBT2_896, SFMBT2_897, CHST2_7890, VAV3 y DTX1 son marcadores eficaces para detectar el cáncer colorrectal en las muestras de heces.

Las secuencias de sondas y cebadores directos e inversos 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 la región DMR proporcionadas en la tabla 6A.

Las sondas de captura para cada marcador se diseñaron para que tuvieran una temperatura de fusión de 75-80 °C y longitudes entre 25-35 bases (véase la tabla 6C). Además, la región de hibridación de la sonda de captura se seleccionó para que estuviera dentro de la huella de los QuARTS posteriores al bisulfito. La tabla 6C proporciona el marcador de metilación y las secuencias de sonda de captura respectivas.

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

Marcador de metilación	Cebador directo	Secuencia de la sonda	Cebador inverso
VAV3	TCGGAGTCGA GTTTAGCGC (Id. de sec. n.º: 54)	CCACGGACG- CGGCGTTCGCGA/3C6/ (Id. de sec. n.º: 55)	CGAAATCGAAAAACAAA AACCGC (Id. de sec. n.º: 56)
CHST2_7889	CGAGTTCGGT AGTTGTACGT AGA (Id. de sec. n.º: 57)	CGCCGAGG- TCGTCGATACCG/3C6/ (Id. de sec. n.º: 58)	CGAAATACGAACGCGAA ATCTAAAAC (Id. de sec. n.º: 59)
SFMBT2_897	GTcGTcGTTcG AGAGGGTA (Id. de sec. n.º: 60)	CCACGGACG-CGCCGAGG-CTACG AACCGAA/3C6/ (Versión 1) (Id. de sec. n.º: 64)	CGAACAAAAACGAACGA ACGAA (Id. de sec. n.º: 62)
SFMBT2_896	GACGCGGAGATAGTGTGG (Versión 1) (Id. de sec. n.º: 63)	ATCGGTTTCGTT/3C6/ (Id. de sec. n.º: 61)	CCTAACCAACGCACTCA ACC (Versión 1) (Id. de sec. n.º: 65)
	GCGTTTAGGT TGGTCGGAG (Versión 2) (Id. de sec. n.º: 90)	CGCCGAGGCCGAA AAACACTAC/3C6/ (Versión 2) (Id. de sec. n.º: 91)	ACGCACTCAACCTACGA AC (Versión 2) (Id. de sec. n.º: 92)
SFMBT2_895	TTAGCGAcGTA GTcGTcGTTG (Versión 1) (Id. de sec. n.º: 66)	CCACGGACG- CGAAAACGCGAA/3C6/ (Versión 1) (Id. de sec. n.º: 67)	CCCAACGCGAAAAAAC GC (Versión 1) (Id. de sec. n.º: 68)
	GCGACGTAGT CGTCGTTGT (Versión 2) (Id. de sec. n.º: 93)	CCACGGACGAAAA CGCGAAA/3C6/ (Versión 2) (Id. de sec. n.º: 94)	CG (Versión 2) (Id. de sec. n.º: 95)
CHST2_7890	GTATAGCGCG ATTTCTAGcG (Id. de sec. n.º: 69)	CGCCGAGG- CGAACATCCTCC/3C6/ (Id. de sec. n.º: 70)	AATTACCTACGCTATCCG CCC (Id. de sec. n.º: 71)
OPLAH	cGTcGcGTTTT TcGGTTATACG (Id. de sec. n.º: 72)	CCACGGACG- GCACCGTAAAC/3C6/ (Id. de sec. n.º: 73)	CGCGAAAACTAAAAAC CGCG (Id. de sec. n.º: 74)
PDGFD	AAACGTTAATT TGTTGTTTGT	AC I I CCGAACGCG TATAAATACC	GCGAATAATAAACGTTA

ES 2 980 778 T3

	TCGTT (Versión 1) (Id. de sec. n.º: 75)	(Versión 1) (Id. de sec. n.º: 76)	ATTTGTTGTTTGTTCG (Versión 1) (Id. de sec. n.º: 77)
5	GCGAATAAAT AACGTTAATT TGTTGTTTGTTCG (Versión 2) (Id. de sec. n.º: 96)	CCACGGACGCGCA CTTCCTTA/3C6/ (Versión 2) (Id. de sec. n.º: 97)	ACTTTCCGAACGCGTATA AATACC (Versión 2) (Id. de sec. n.º: 98)
	FLI1	GTTGcGAGGT TAGGTTGTAAT CG (Id. de sec. n.º: 78)	CGCCGAGG- CGTCCATTTAAC/3C6/(Id. de sec. n.º: 79)
	DTX1	GAGTCGCGGT TTCGTTTTTC (Id. de sec. n.º: 81)	CGCCGAGG- CGCGTTCGTTTT /3C6/(Id. de sec. n.º: 82)
10	NDRG4	CGGTTTTTCG TCGTTTTTTCG (Id. de sec. n.º: 84)	CGCCGCTTACGTTAATAA TCCC (Id. de sec. n.º: 80)
	BMP3	GTTTAATTTTC GGTTCGTCG TC (Id. de sec. n.º: 87)	GACGCGACGACCGAAAA AC (Id. de sec. n.º: 83)
		CGCCGAGG GTTC GTTTATCG/3C6/(Id. de sec. n.º: 85)	CCGCCTTCTACGCGACT A (Id. de sec. n.º: 86)
		CGCCGAGG CGGTTTTTTCG/3C6/(Id. de sec. n.º: 88)	CGCTACGAAACTCCG A (Id. de sec. n.º: 89)

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

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

35 Tabla 6C. Marcador de metilación y secuencias de sondas de captura respectivas.

Marcador de metilación	Secuencia de sonda de captura
VAV3	/5AmMC6/GATCGAGGGAGCAGGAGCCGCGGCTGACGGGTC GCG (Id. de sec. n.º: 99)
40 CHST2_7889	/5AmMC6/CGGTGCCGAGAGCTGCCAGAGAGTTGGATTCTGC G (Id. de sec. n.º: 100)
SFMBT2_897	/5AmMC6/GCGAGCGGGCAAGGGCGGGCGAGC (Id. de sec. n.º: 101)
SFMBT2_896	/5AmMC6/ACCTGCGGGCCGAAGGGCTGCTCTCCGG (Id. de sec. n.º: 102)
SFMBT2_895	/5AmMC6/AGGAGACGCGGGAGCGCGGGGTAGGTAGC (Id. de sec. n.º: 103)
45 CHST2_7890	/5AmMC6/GGCATCCTCCCGGTGATGGAAGCAGCCGCCG G (Id. de sec. n.º: 104)
OPLAH	/5AmMC6/GGAAGGCGCGGCTCGGTGAGCACTGACAGCA G (Id. de sec. n.º: 105)
PDGFD	/5AmMC6/TCGCCGAGCTCTCCCAAATTCCTGCATGCTGAA CTTT (Id. de sec. n.º: 106)
FLI1	/5AmMC6/CCGTCCATTTGGCCAAGTCTGCAGCCGAGCC (Id. de sec. n.º: 107)
50 DTX1	/5AmMC6/CTGCGTCCGTCCGTGCGCCGGCAGTCTGTCCA (Id. de sec. n.º: 108)
NDRG4	/5AmMC6/TCCCTCGCGGTGGCTTCCGCCTTCTGCGCGGCT GGGGTGCCCGGTGG (Id. de sec. n.º: 109)
BMP3	/5AmMC6/GCGGGACACTCCGAAGGCGCAAGGAG (Id. de sec. n.º: 110)

55 Cada sonda de captura se sintetizó con una modificación de 5'-NH₂ para permitir el acoplamiento a partículas magnéticas modificadas con -COOH mediante una química de acoplamiento de carbodiimida estándar. Además, se sintetizó un oligonucleótido complementario a la sonda de captura para contener un marcador 5'-Cy₃. Esta sonda complementaria se usó para confirmar el acoplamiento de la sonda de captura a partículas magnéticas.

60 Para probar la eficacia de captura de cada sonda y evaluar el rendimiento de los marcadores, se hicieron dos grupos de heces de pacientes normales y con cáncer. El grupo de heces cancerosas provino de 6 pacientes (3 son de CCR, 1 es AA y 2 desconocidas). Del mismo modo, las heces normales procedían de 6 pacientes normales sin CCR. Las heces se prepararon mezclando el sobrenadante después de la centrifugación homogeneizada. El sobrenadante combinado se dividió luego en alícuotas en muestras de captura individuales que contenían 14 ml de sobrenadantes.

65

ES 2 980 778 T3

Las sondas de captura se diseñaron 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 que estuviera dentro de la huella de los QUARTS posteriores al bisulfito.

5 Para realizar la captura, las perlas de captura (partículas magnéticas con sondas de captura unidas covalentemente) para dos marcadores más las perlas de captura ACTB se combinaron para formar un conjunto de perlas de captura triplex.

10 La captura se realizó por triplicado de 14 ml de sobrenadantes fecales positivos y normales (la excepción fue que el VAV3 y el DTX1 se realizaron por duplicado, ya que el grupo se estaba agotando).

15 Una vez completada la captura, el ADN fecal se eluyó de las perlas de captura con NaOH 0,1 N a 42 °C durante 20 minutos, seguido de la conversión con bisulfito a 56 °C durante 1 hora usando bisulfito de amonio. A continuación, el ADN de las heces se desulfonó y se purificó usando perlas magnéticas recubiertas de sílice y se eluyó en 70 ul de Tris-HCl 10 mM, pH 8, EDTA 0,1 mM. A continuación, se probaron 10 ul de eluyente y se cuantificaron en ensayos de QuARTS.

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

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

25 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 depósito de heces positivo y normal, lo que indica que estos marcadores son candidatos para la detección del CCR en las heces.

Tabla 6D.

Marcador de metilación	Cadenas promedio del grupo positivo de cadenas	Cadenas promedio del grupo normal	Factor de diferencia (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.

50 Procedimiento ilustrativo para la detección de un cáncer colorrectal en un sujeto humano.

55 Poner en contacto un ácido nucleico (por ejemplo, ADN genómico, por ejemplo, aislado de un fluido corporal tal como una muestra de heces o tejido colorrectal) obtenido de un sujeto humano con al menos un reactivo o una serie de reactivos que distingue entre dinucleótidos de CpG metilados y no metilados dentro de al menos un marcador DMR seleccionado entre:

- 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);
- 60 • 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, CHST2_7890, VAV3 y DTX1 (como se indica en la tabla 6).

65

Identificar que el sujeto 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 ilustrativo para la detección de un adenoma grande de base colorrectal en un sujeto humano.

5 Poner en contacto un ácido nucleico (por ejemplo, ADN genómico, por ejemplo, aislado de un fluido corporal tal como una muestra de heces o tejido colorrectal) obtenido de un sujeto humano con al menos un reactivo o una serie de reactivos que distingue entre dinucleótidos de CpG metilados y no metilados dentro de al menos un marcador DMR seleccionado entre:

- 10
- ADD2, AGBL4, AKAP12, ANKRD33B, ASCL1, Clorf70, CHST11, CHST15, DTX1, ECEL1, EYA4, FLI1, FLJ43390, FLT3, GRASP, ITGA4, KCNQ2, LOC100131199, LONFR2, MGAT3, OPLAH (145106742-145106921), OPLAH (145106349-145106456), PDE8B, PDGFD, PKIA, POU3F1, OKI, RASSF2, RSPO3, SDC2, SFMBT2 (7452746-7452779), SFMBT2 (7452029-7452452), SFMBT2 (7450242-7450831), SOX5, VAV3 (108507609-108507674), VAV3 (108507074-108507498) y ZNF132 (como se indica en la tabla 4).
- 15

Identificar que el sujeto tiene un adenoma grande de base 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.

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

Poner en contacto un ácido nucleico (por ejemplo, ADN genómico, por ejemplo, aislado de un fluido corporal tal como una muestra de heces o tejido colorrectal) obtenido de un sujeto humano con al menos un reactivo o una serie de reactivos que distingue entre dinucleótidos de CpG metilados y no metilados dentro de al menos un marcador DMR seleccionado entre:

- 25
- CAMK4, FGF9, GJC1, GPX7, GRIN2A, IGF2BP3, MAX.chr1 9.55963254-55963329, MAX.chr2.127783352-127783403, MAX.chr4.186049639-186049660, NRG1, NTNG1, PIK3CD, PTENP1, RAB31, RNF220, SLC30A10 y ZNF853 (como se indica en la tabla 5).
- 30

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

35 Procedimiento ilustrativo para la detección de un cáncer colorrectal en un sujeto humano que tiene una enfermedad inflamatoria intestinal.

Poner en contacto un ácido nucleico (por ejemplo, ADN genómico, por ejemplo, aislado de un fluido corporal, tal como una muestra de heces o tejido colorrectal) obtenido de un sujeto humano que tiene una enfermedad inflamatoria intestinal con al menos un reactivo o una serie de reactivos que distingan entre dinucleótidos de CpG metilados y no metilados dentro de al menos un marcador DMR seleccionado entre:

- 40
- BMP3, NDRG4, PDGFG, CHST2 y SFMBT2 (como se indica en la tabla 2).

45 Identificar que el sujeto 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.

REIVINDICACIONES

- 5 1. Un método de detección de una neoplasia colorrectal en una muestra obtenida de un sujeto, el método comprende:
- 1) analizar un estado de metilación de un marcador en una muestra obtenida de un sujeto; e
- 10 2) identificar que el sujeto tiene una neoplasia cuando el estado de metilación del marcador comprende un aumento de la metilación en relación con el estado de metilación normal de ese marcador,
- en donde el marcador comprende una base en una región diferencialmente metilada (DMR) y en donde la región DMR es PPP2R5C.
- 15 2. El método de la reivindicación 1, en donde la neoplasia es precancerosa, un cáncer colorrectal, un adenoma colorrectal grande y/o un pólipo serrado sésil.
3. El método de la reivindicación 1, en donde analizar el estado de metilación de un marcador comprende tratar la muestra obtenida con un reactivo que modifica selectivamente los residuos de citosina no metilados de la muestra para producir residuos modificados, pero que no modifica los residuos de citosina metilados; y determinar el nivel de metilación del marcador en la muestra que se ha sometido a la etapa de tratamiento.
- 20 4. El método de la reivindicación 1, que comprende analizar una pluralidad de marcadores.
- 25 5. El método de la reivindicación 1, que comprende analizar de 2 a 11 marcadores.
6. El método de la reivindicación 1, que comprende analizar de 12 a 107 marcadores.
- 30 7. El método de la reivindicación 1, en donde analizar el estado de metilación del marcador en la muestra comprende determinar el estado de metilación de una base.
8. El método de la reivindicación 1, en donde analizar el estado de metilación del marcador en la muestra comprende determinar el grado de metilación en una pluralidad de bases.
- 35 9. El método de la reivindicación 1, que comprende analizar el estado de metilación de una cadena directa o analizar el estado de metilación de una cadena inversa.
10. El método de la reivindicación 1, en donde el marcador es una región de 100 o menos bases.
- 40 11. El método de la reivindicación 1, en donde el marcador es una región de 500 o menos bases.
12. El método de la reivindicación 1, en donde el marcador es una región de 1000 o menos bases.
- 45 13. El método de la reivindicación 1, en donde el marcador es una región de 5000 o menos bases.