

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 712 798**

51 Int. Cl.:

C12Q 1/6886 (2008.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **12.03.2014 PCT/US2014/024589**

87 Fecha y número de publicación internacional: **02.10.2014 WO14159652**

96 Fecha de presentación y número de la solicitud europea: **12.03.2014 E 14776150 (6)**

97 Fecha y número de publicación de la concesión europea: **09.01.2019 EP 2971179**

54 Título: **Detección de neoplasias**

30 Prioridad:

14.03.2013 US 201361784429 P

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

14.05.2019

73 Titular/es:

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

72 Inventor/es:

**AHLQUIST, DAVID A.;
KISIEL, JOHN B.;
TAYLOR, WILLIAM R.;
YAB, TRACY C.;
MAHONEY, DOUGLAS W.;
LIDGARD, GRAHAM, P. y
ALLAWI, HATIM, T.**

74 Agente/Representante:

CARPINTERO LÓPEZ, Mario

ES 2 712 798 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

Campo de la invención

5 Se proporciona en el presente documento una tecnología relacionada con la detección de neoplasias y, en particular, pero no exclusivamente, a procedimientos, composiciones, y usos relacionados para detectar neoplasias malignas y premalignas tales como el cáncer pancreático y el cáncer colorrectal.

Antecedentes

10 En su conjunto, los cánceres gastrointestinales representan más mortalidad por cáncer que cualquier otro sistema orgánico. Aunque en la actualidad existe un cribado de cánceres colorrectales, la mortalidad anual en Estados Unidos producida por cánceres del tracto gastrointestinal superior supera las 90.000, en comparación con las aproximadamente 50.000 para el cáncer colorrectal. Asombrosamente, 43.000 hombres y mujeres se diagnostican cada año con cáncer pancreático (PanC), lo que producirá casi 37.000 muertes anuales (Jemal y col. (2010) "Cancer statistics" CA Cancer J Clin 60: 277-300). Como resultado, PanC es la cuarta causa de muerte por cáncer principal (id). Los pacientes que presentan síntomas habitualmente ya se encuentran en un estado avanzado de la enfermedad y solamente un 15 % cumplen los criterios de una cirugía potencialmente curativa (Ghaneh y col. (2007) "Biology and management of pancreatic cancer" Gut 56: 1134-52). A pesar de la cirugía, el 85 % morirá por enfermedad recurrente (Sohn y col. (2000) "Resected adenocarcinoma of the pancreas-616 patients: results, outcomes, and prognostic indicators" J Gastrointest Surg 4: 567-79). La mortalidad por PanC supera el 95 % y la supervivencia a 5 años es menos del 25 % para en los pacientes que se han sometido a cirugía curativa (Cleary y col (2004) "Prognostic factors in resected pancreatic adenocarcinoma: analysis of actual 5-year survivors" J Am Coll Surg 198: 722-31; Yeo y col (1995) "Pancreaticoduodenectomy for cancer of the head of the pancreas. 201 patients" Ann Surg 221: 721-33).

25 Entre los pacientes con predisposición sindrómica a PanC o antecedentes familiares fuertes, las estrategias de cribado agresivas e invasivas usando exploraciones por tomografía computerizada o ultrasonidos con endoscopia has mostrado un rendimiento del 10 % para neoplasias (Canto y col. (2006) "Screening for early pancreatic neoplasia in high-risk individuals: a prospective controlled study" Clin Gastroenterol Hepatol 4: 766-81). Esta estrategia de cribado no es práctica para la población en general, donde la mayoría del PanC se produce sin una predisposición conocida (Klein y col. (2001) "Familial pancreatic cancer" Cancer J 7: 266-73).

30 La casi uniforme letalidad del PanC ha generado mucho interés en la comprensión de la biología del tumor pancreático. Se han identificado lesiones precursoras, incluidas la neoplasia intraepitelial pancreática (PanIN, grados I-III) y la neoplasia mucinosa papilar intraductal (IPMN) (Fernandez-del Castillo y col. (2010) "Intraductal papillary mucinous neoplasms of the pancreas" Gastroenterology 139: 708-13, 713.e1-2; Haugk (2010) "Pancreatic intraepithelial neoplasia - can we detect early pancreatic cancer?" Histopathology 57: 503-14). El estudio de las lesiones tanto precursoras como malignas ha identificado numerosas características moleculares en niveles genéticos, epigenéticos, y proteómicos que se podrían aprovechar para la terapia o usarse como biomarcadores para la detección precoz y el cribado (Kaiser (2008) "Cancer genetics. A detailed genetic portrait of the deadliest human cancers" Science 321: 1280-1; Omura y col. (2009) "Epigenetics and epigenetic alterations in pancreatic cancer" Int J Clin Exp Pathol 2: 310-26; Tonack y col. (2009) "Pancreatic cancer: proteomic approaches to a challenging disease" Pancreatology 9: 567-76). Recientes estudios de cartografiado de lesiones tumorales y metastásicas han mostrado que puede haber un período latencia significativo entre el desarrollo del PanC maligno y el desarrollo de la enfermedad metastásica, lo que sugiere la existencia de una amplia ventana de oportunidad para la detección y el tratamiento curativo de lesiones presintomáticas en la etapa inicial (Yachida y col. (2010) "Distant metastasis occurs late during the genetic evolution of pancreatic cancer" Nature 467: 1114-7).

45 PanC emite (por ejemplo, exfolia) las células y el ADN a los efluentes locales y finalmente a las heces. Por ejemplo, se puede identificar (por ejemplo, secuenciarse) un ADN que contiene un gen KRAS mutante a partir del jugo pancreático de pacientes con cáncer pancreático, PanIN, e IPMN (Yamaguchi y col. (2005) "Pancreatic juice cytology in IPMN of the pancreas" Pancreatology 5: 416-21). Anteriormente, se han usado ensayos de alta sensibilidad para detectar ADN mutante en heces emparejadas de pacientes con cáncer de páncreas de los que se sabía que sus tumores extirpados contenían las mismas secuencias (Zou y col (2009) "T2036 PanDetection of Gastrointestinal Neoplasms By Stool DNA Testing: Establishment of Feasibility" Gastroenterology 136: A-625). Una limitación en el número de marcadores está relacionada con el procedimiento no desvelado de su detección en ensayos convencionales; de forma típica, cada sitio de mutación para múltiples genes se puede estudiar por separado para conseguir alta sensibilidad.

55 El ADN metilado se ha estudiado como una clase potencial de biomarcadores en los tejidos de la mayoría de los tumores. En muchos casos, las ADN metiltransferasas añaden un grupo metilo al ADN en los sitios de islas citosina-fosfato-guanina (CpG) en forma de un control epigenético o expresión génica. En un mecanismo atractivo en biología, se cree que los eventos de metilación adquirida en regiones promotoras de los genes supresores tumorales silencian la expresión, contribuyendo de esta forma a la oncogénesis. La metilación del ADN puede ser una

herramienta diagnóstica más estable química y biológicamente que la expresión de ARN o de proteínas (Laird (2010) "Principles and challenges of genome-wide DNA methylation analysis" *Nat Rev Genet* 11: 191-203). Además, en otros cánceres, como el cáncer de colon esporádico, los marcadores de la metilación ofrecen una especificidad excelente y son muchos más informativos y sensibles que las mutaciones individuales en el ADN (Zou y col (2007) "Highly methylated genes in colorectal neoplasia: implications for screening" *Cancer Epidemiol Biomarkers Prev* 16: 2686-96).

El análisis de las islas CpG ha proporcionado hallazgos importantes cuando se aplican a modelos animales, y líneas de células humanas. Por ejemplo, Zhang y colaboradores han descubierto que amplicones de diferentes partes de la misma isla CpG pueden tener diferentes niveles de metilación (Zhang y col. (2009) "DNA methylation analysis of chromosome 21 gene promoters at single base pair and single allele resolution" *PLoS Genet* 5: e1000438). Además, los niveles de metilación estaban distribuidos bimodalmente entre secuencias fuertemente metiladas y no metiladas, respaldando además el patrón binario de la actividad de la ADN metiltransferasa análogo a un interruptor (Zhang y col. (2009) "DNA methylation analysis of chromosome 21 gene promoters at single base pair and single allele resolution" *PLoS Genet* 5: e1000438). El análisis de tejidos murino *in vivo* y de líneas de células *in vitro* ha demostrado que solo aproximadamente un 0,3 % de los promotores con alta densidad de CpG (HCP, definido por tener >7 % de secuencia CpG dentro de una región de 300 pares de bases) está metilado, mientras que zonas con baja densidad de CpG (LCP, definido por tener <5 % de secuencias CpG en una región de 300 pares de bases) mostraban tendencia a estar frecuentemente metiladas con un patrón dinámico específico de tejido (Meissner y col. (2008) "Genome-scale DNA methylation maps of pluripotent and differentiated cells" *Nature* 454: 766-70). Los HCP incluyen promotores para genes constitutivos ubicuos y genes con el desarrollo fuertemente regulado. Entre los sitios HCP metilados >50 % se encuentran varios marcadores establecidos tales como Wnt 2, NDRG2, SFRP2, y BMP3 (Meissner y col. (2008) "Genome-scale DNA methylation maps of pluripotent and differentiated cells" *Nature* 454: 766-70).

Para el cáncer pancreático, lesiones PanIN, y IPMN, se ha estudiado la metilación de marcadores a nivel tisular (Omura y col. (2008) "Genome-wide profiling of methylated promoters in pancreatic adenocarcinoma" *Cancer Biol Ther* 7: 1146-56; Sato y col. (2008) "CpG island methylation profile of pancreatic intraepithelial neoplasia" *Mod Pathol* 21: 238-44; Hong y col. (2008) "Multiple genes are hypermethylated in intraductal papillary mucinous neoplasms of the pancreas" *Mod Pathol* 21: 1499-507). Por ejemplo, los marcadores MDFI, ZNF415, CNTNAP2, y ELOVL4 estaban fuertemente metilados en el 96 %, 86 %, 82 %, y 68 % de los cánceres estudiados mientras que, comparativamente, solo el 9 %, 6 %, 3 %, y 7 % de los páncreas de control (no cancerosos), respectivamente, estaban fuertemente metilados en los mismos cuatro loci (Omura y col. (2008) "Genomewide profiling of methylated promoters in pancreatic adenocarcinoma" *Cancer Biol Ther* 7: 1146-56). Se ha descubierto que la medición del estado de metilación tanto de MDFI como de CNTNAP2 proporciona un indicador del cáncer pancreático que tiene al mismo tiempo una elevada sensibilidad (>90 %) y una elevada especificidad (>85 %) (Omura y col. (2008) "Genome-wide profiling of methylated promoters in pancreatic adenocarcinoma" *Cancer Biol Ther* 7: 1146-56).

Además, Sato y colaboradores han descubierto ocho genes que se expresan de forma diferentes en líneas celulares de cáncer pancreático antes y después del tratamiento con un inhibidor de la metiltransferasa (Sato y col. (2003) "Discovery of novel targets for aberrant methylation in pancreatic carcinoma using high-throughput microarrays" *Cancer Res* 63: 3735-42). Posteriormente, estos marcadores se evaluaron mediante PCR específica de metilación (MSP) del ADN de lesiones de Pan-IN. Los resultados mostraron que SARP-2 (proteína 1 relacionada con Frizzled secretada, SFRP1) tenía un 83 % de sensibilidad y podría discriminar entre Pan-IN 2 y Pan-IN 3 de grado superior (Sato y col. (2008) "CpG island methylation profile of pancreatic intraepithelial neoplasia" *Mod Pathol* 21: 238-44). La discriminación de un precursor de alto grado o cáncer en etapa inicial de una lesión de grado menor es importante cuando se considera la morbilidad de la pancreatoduodenectomía o la pancreatocetomía distal, las principales terapias quirúrgicas del PanC. Cuando estudiaron los precursores IPMN tanto del conducto principal como de la rama lateral, Hong y colaboradores mostraron una elevada sensibilidad y especificidad para SFRP1, así como, especialmente en combinación con UCHL1 (Hong y col. (2008) "Multiple genes are hypermethylated in intraductal papillary mucinous neoplasms of the pancreas" *Mod Pathol* 21: 1499-507). El inhibidor de la ruta del factor tisular 2 (TFPI2) tiene un papel supresor tumoral bien establecido en neoplasias malignas del GU y el GI, incluidos los cánceres de próstata, de cuello de útero, colorrectal, gástrico, esofágico y pancreático (Ma y col. (2011) "MicroRNA-616 induces androgen-independent growth of prostate cancer cells by suppressing expression of tissue factor pathway inhibitor TFPI-2" *Cancer Res* 71: 583-92; Lim y col. (2010) "Cervical dysplasia: assessing methylation status (Methylight) of CCNA1, DAPK1, HS3ST2, PAX1 and TFPI2 to improve diagnostic accuracy" *Gynecol Oncol* 119: 225-31; Hibi y col. (2010) "Methylation of TFPI2 gene is frequently detected in advanced well-differentiated colorectal cancer" *Anticancer Res* 30: 1205-7; Glockner y col. (2009) "Methylation of TFPI2 in stool DNA: a potential novel biomarker for the detection of colorectal cancer" *Cancer Res* 69: 4691-9; Hibi y col. (2010) "Methylation of the TFPI2 gene is frequently detected in advanced gastric carcinoma" *Anticancer Res* 30: 4131-3; Tsunoda y col. (2009) "Methylation of CLDN6, FBN2, RBP1, RBP4, TFPI2, and TMEFF2 in esophageal squamous cell carcinoma" *Oncol Rep* 21: 1067-73; Tang y col. (2010) "Prognostic significance of tissue factor pathway inhibitor-2 in pancreatic carcinoma and its effect on tumor invasion and metastasis" *Med Oncol* 27: 867-75; Brune y col. (2008) "Genetic and epigenetic alterations of familial pancreatic cancers" *Cancer Epidemiol Biomarkers Prev* 17: 3536-4). También se ha comprobado que este marcador se envía a la luz GI y tenía una sensibilidad del 73 % cuando se analizó en jugo pancreático de sujetos con cánceres y normales (Matsubayashi y col. (2006) "DNA methylation alterations in the

pancreatic juice of patients with suspected pancreatic disease" Cancer Res 66: 1208-17).

TFPI2 está comprendida entre un elevado número de potenciales mutaciones y marcadores de la metilación estudiados en muestras de tejidos y de heces en candidatos de neoplasia colorrectal. En un análisis de un conjunto de ensayos de formación sobre heces de archivo procedentes de casi 700 sujetos, un panel de multimarcadores de metilación, que incluye TFPI2, BMP3, NDRG4, y vimentina, demostró tener una sensibilidad del 85 % para CRC y una sensibilidad del 64 % en adenomas colorrectales avanzados, ambos con una especificidad del 90 % (Ahlquist D y col. (2010) "Next Generation Stool DNA Testing for Detection of Colorectal Neoplasia - Early Marker Evaluation", presentado en Colorectal Cancer: Biology to Therapy, American Association for Cancer Research).

Una investigación anterior ha sometido a ensayo los marcadores de la metilación del cáncer colorrectal para la detección de PanC. En particular, un estudio control de casos comparó el ADN procedente de casos de tumor PanC con el ADN procedente del epitelio colónico usando marcadores dirigidos a MSP anteriormente notificados en PanC (por ejemplo, MDFI, SFRP2, UCHL1, CNTNAP2, y TFPI2) y marcadores adicionales discriminantes de neoplasias colónicas (por ejemplo, BMP3, EYA4, vimentina, y NDRG4). Un modelo de regresión multimarcador mostró que EYA4, UCHL1, y MDFI fueron muy discriminantes, con un área bajo la curva de las características operativas del receptor de 0,85. Como marcador individual, se ha descubierto que BMP3 tiene un área bajo la curva de las características operativas del receptor de 0,90. Estos cuatro marcadores y un KRAS mutante se analizaron posteriormente en un conjunto más grande de muestras de heces procedentes de sujetos con PanC en una comparación enmascarada con heces correspondientes de individuos con un resultado de colonoscopia normal. Individualmente, BMP3 y KRAS fueron muy específicos pero poco sensibles; en combinación, la sensibilidad mejoró al 65 % manteniendo un 88 % de especificidad (Kisiel, y col. (2011) "Stool DNA screening for colorectal cancer: opportunities to improve value with next generation tests" J Clin Gastroenterol 45: 301-8. Estos resultados sugieren que las diferencias en la metilación de UCHL1, EYA4, y MDFI en el páncreas están ocluidos por la metilación colónica del fondo en una comparación basada en heces. Por tanto, se necesita en el cribado del cáncer un panel marcador para PanC que sea muy informativo y presente una elevada especificidad del PanC a nivel tisular cuando se busca en muestras tomadas de un sujeto (por ejemplo, una muestra de heces).

Sumario

En consecuencia, se proporciona en el presente documento tecnología para marcadores de cáncer pancreático y otros marcadores de cribado del cáncer gastrointestinal que proporcionan una alta relación señal-ruido y un bajo nivel de fondo cuando se detecta en muestras tomadas de un sujeto (por ejemplo muestra de heces). Los marcadores se identificaron en un estudio de control de caso comparando el estado de metilación de marcadores de ADN procedentes de sujetos con PanC en estadio I y II con el estado de metilación de los mismos marcadores de ADN procedentes de sujetos de control (por ejemplo, tejido normal tal como colon normal y/o páncreas no neoplásico (véanse, los Ejemplos 1 y 11).

Los marcadores y/o paneles de marcadores (Por ejemplo, una región cromosómica que tiene una anotación seleccionada entre ABCB1, ADCY1, BHLHE23 (LOC63930), c13orf18, CACNA1C, chr12 133, CLEC11A, ELMO1, EOMES, CLEC 11, SHH, GJC1, IHIF1, IKZF1, KCNK12, KCNN2, PCBP3, PRKCB, RSPO3, SCARF2, SLC38A3, ST8SIA1, TWIST1, VWC2, WT1, y ZNF71) se identificaron en un estudio de control de caso comparando el estado de metilación de marcadores de ADN (por ejemplo, procedentes de sujetos con PanC en estadio I y II con el estado de metilación de los mismos marcadores de ADN procedentes de sujetos de control (por ejemplo, tejido normal tal como colon normal y/o páncreas no neoplásico (véanse, los Ejemplos 2 y 8).

Los marcadores y/o paneles de marcadores (Por ejemplo, una región cromosómica que tiene una anotación seleccionada entre NDRG4, SFRP1, BMP3, HPP1, y/o APC) se identificaron en estudios de control del caso comparando el estado de metilación de marcadores del ADN procedentes de tejido esofágico de sujetos con esófago de Barrett con el estado de metilación de los mismos marcadores de ADN procedentes de sujetos del control (véanse los Ejemplo 4 y 10).

Los marcadores y/o paneles de marcadores (Por ejemplo, una región cromosómica que tiene una anotación seleccionada entre ADCY1, PRKCB, KCNK12, C13ORF18, IKZF1, TWIST1, ELMO, 55957, CD1D, CLEC11A, KCNN2, BMP3, y/o NDRG4) se identificaron en estudios de control del caso comparando el estado de metilación de marcadores del ADN procedentes de una muestra de jugo pancreático de sujetos con cáncer de páncreas con el estado de metilación de los mismos marcadores de ADN procedentes de sujetos del control (véanse los Ejemplo 5 y 6).

Se identificó un marcador (por ejemplo, una región cromosómica que tiene una anotación CD1D) en un estudio de control del caso por comparación del estado de metilación de un marcador de ADN (por ejemplo, CD1D) procedente de una muestra de heces de sujetos con cáncer de páncreas con el estado de metilación del mismo marcador de ADN procedentes de sujetos del control que no tienen cáncer de páncreas (véase, el Ejemplo 7).

Se identificó un marcador (por ejemplo, miR-1290) en un estudio de control del caso comparando la cantidad cuantificada de un marcador de ADN (por ejemplo, miR-1290) de una muestra de heces de sujetos con cáncer de páncreas con la cantidad cuantificada del mismo marcador de ADN procedentes de sujetos de control que no tienen

cáncer de páncreas (véase, Ejemplo 9).

Un análisis estadístico adicional de los resultados demostró que la tecnología descrita en el presente documento basada en estos marcadores predice de una forma específica y sensible un sitio tumoral.

5 Como se describe en el presente documento, la tecnología proporciona numerosos marcadores de ADN metilados y subconjuntos de los mismos (por ejemplo, conjuntos de 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 o más marcadores) con elevada discriminación para las neoplasias gastrointestinales tanto en su conjuntos como y/o en sitios tumorales adicionales. Los experimentos aplicaron un filtro de selección a los marcadores candidatos para identificar marcadores que proporcionan una alta relación señal-ruido y un bajo nivel de fondo para proporcionar una elevada especificidad, por ejemplo, cuando se analizan medios distantes (por ejemplo, heces, sangre, orina, tejido metastásico, etc.) con los fines de cribado o diagnóstico del cáncer. Además, se realizaron experimentos para demostrar que los marcadores del ADN metilado identificado predicen el sitio del tumor. Por tanto, la tecnología proporciona marcadores específicos, combinaciones de marcadores, y algoritmos para predecir el sitio del tumor.

15 En algunas realizaciones, la tecnología está relacionado con la evaluación de la presencia y del estado de metilación de uno o más de los marcadores identificados en el presente documento en una muestra biológica. Estos marcadores comprenden una o más regiones diferencialmente metiladas (DMR) como se analiza en el presente documento, por ejemplo, como se proporciona en la Tabla 1 y/o la Tabla 10. El estado de la metilación se evalúa en las realizaciones de la tecnología. Por tanto, la tecnología proporcionada en el presente documento no está restringida al procedimiento por el cual se mide el estado de la metilación de un gen. Por ejemplo, en algunas realizaciones, el estado de metilación se mide por un procedimiento de cribado genómico. Por ejemplo, un procedimiento implica la exploración genómica con hitos de restricción (Kawai y col. (1994) Mol. Cell. Biol. 14: 7421-7427) y otro ejemplo implica la PCR cebada arbitrariamente sensible a metilación PCR (Gonzalzo y col. (1997) Cancer Res. 57: 594-599). En algunas realizaciones, los cambios en los patrones de metilación en sitios CpG específicos se controlan mediante la digestión del ADN genómico con enzimas de restricción sensibles a la metilación seguido de análisis Southern de las regiones de interés (procedimiento de digestión-Southern). En algunas realizaciones, en análisis de los cambios en los patrones de metilación implica un procedimiento basado en PCR que implica digestión de ADN genómico con enzimas de restricción sensibles a la metilación antes de la amplificación mediante PCR (Singer-Sam y col. (1990) Nucl. Acids Res. 18: 687). Además, Se han notificado otras técnicas que utilizan el tratamiento del ADN con bisulfito como punto de partida del análisis de metilación. Esto incluye PCR específica de metilación (MSP) (Herman y col. (1992) Proc. Natl. Acad. Sci. USA 93: 9821-9826) y digestión con enzimas de restricción de los productos de la PCR amplificados procedentes del 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 la PCR para la detección de mutaciones génicas (Kuppuswamy y col. (1991) Proc. Natl. Acad. Sci. USA 88: 1143-1147) y la cuantificación de la expresión específica de alelo (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 que terminan inmediatamente en 5' respecto del único nucleótido a analizar. Los procedimientos que utilizan un "ensayo cuantitativo Ms-SNuPE" como se describe en la patente de EE.UU. N.º 7.037.650, se utilizan en algunas realizaciones.

40 Tras evaluar el estado de metilación, el estado de metilación se expresa frecuentemente como la fracción o porcentaje de las hebras individuales del ADN que están metiladas en un sitio en particular (por ejemplo, en un único nucleótido, en una región o locus en particular, en una secuencia de interés más larga, por ejemplo, hasta ~100 pb, 200 pb, 500 pb, subsecuencia de 1000 pb de un ADN o más largo) con respecto a la población total de ADN en la muestra que comprende dicho sitio en particular. Tradicionalmente, la cantidad de ácido nucleico no metilado se determina mediante la PCR usando calibradores. Después, una cantidad conocida de ADN se trata con bisulfito, y la secuencia específica de metilación resultante se determina bien mediante PCR en tiempo real u otra amplificación exponencial, por ejemplo, un ensayo QuARTS (por ejemplo, como se proporciona en la patente de Estados Unidos n.º 8.361.720; y en la solicitud de patente de Estados Unidos publicada con números WO 2012/0122088 y WO 2012/0122106).

50 Por ejemplo, en algunas realizaciones, los procedimientos comprenden generar una curva patrón para la diana sin metilar usando patrones externos. La curva patrón se construye a partir de al menos dos puntos y se refiere al valor Ct en tiempo real para el ADN sin metilar respecto a patrones cuantitativos conocidos. Después, se construye una segunda curva patrón para la diana metilada a partir de al menos dos puntos y patrones externos. Esta segunda curva patrón se refiere al Ct del ADN metilado respecto a patrones cuantitativos conocidos. A continuación, se determinan los valores de Ct de la muestra para las poblaciones metilada y sin metilar, y se calculan los equivalentes genómicos del ADN a partir de las curvas patrón producidas durante las dos primeras etapas. El porcentaje de metilación en el sitio de interés se calcula a partir de la cantidad de ADN metilados con respecto a la cantidad total de ADN en la población, por ejemplo, (número de ADN metilados) / (el número de ADN metilados + número de ADN sin metilar) x 100.

60 En el presente documento se proporcionan también composiciones y kits para llevar a la práctica los procedimientos. Por ejemplo, en algunas realizaciones, los reactivos (por ejemplo, cebadores, sondas) específicos de uno o más marcadores se proporcionar en solitario o en conjuntos (por ejemplo, conjuntos de pares de cebadores para amplificar una pluralidad de marcadores). También se pueden proporcionar reactivos adicionales para realizar un

ensayo de detección (por ejemplo, enzimas, tampones, controles positivos y negativos para realizar QuARTS, PCR, secuenciación, bisulfito, y otros ensayos). En algunas realizaciones, los kits contienen uno o más reactivos necesarios, suficientes, o útiles, para realizar el procedimiento. También se proporcionan mezclas de reacciones que contienen los reactivos. Se proporcionan además conjuntos de reactivos de mezcla maestra que contiene una pluralidad de reactivos que se pueden añadir entre sí y/o a una muestra de ensayo para completar una mezcla de reacción.

En algunas realizaciones, la tecnología descrita en el presente documento está asociada con una máquina programable diseñada para realizar una secuencia de operaciones aritméticas o lógicas como se proporciona en los procedimientos descritos en el presente documento. Por ejemplo, algunas realizaciones de la tecnología están asociadas (por ejemplo, implementadas) con un programa informático y/o un equipo informático. En un aspecto, la tecnología se refiere a un ordenador que comprende una forma de memoria, un elemento para realizar operaciones aritméticas y lógicas, y un elemento de procesamiento (por ejemplo, un microprocesador) para ejecutar una serie de instrucciones (por ejemplo, un procedimiento como se proporciona en el presente documento) para leer, manipular, y almacenar datos. En algunas realizaciones, un microprocesador forma parte de un sistema para determinar un estado de metilación (por ejemplo, de uno o más DMR, por ejemplo, DMR 1-107 como se proporciona en la Tabla 1, por ejemplo, DMR 1-449 en la Tabla 10); que compara estados de metilación (por ejemplo, de uno o más DMR, por ejemplo, DMR 1-107 como se proporciona en la Tabla 1, por ejemplo, DMR 1-449 en la Tabla 10); generando curvas patrón; determina un valor de Ct; calcula una fracción, frecuencia, o porcentaje de metilación (por ejemplo, de uno o más DMR, por ejemplo, DMR 1-107 como se proporciona en la Tabla 1, por ejemplo, DMR 1-449 en la Tabla 10); identificar una isla CpG; determinar una especificidad y/o una sensibilidad para un ensayo o marcador; calcular una curva ROB, y una ABC asociada; análisis de secuencias; todo ellos como se describe en el presente documento, o es conocido en la técnica.

En algunas realizaciones, un microprocesador u ordenador utiliza los datos del estado de metilación en un algoritmo para predecir un sitio de un cáncer.

En algunas realizaciones, un programa o componente informático recibe los resultados de múltiples ensayos y determina un valor único como el resultado a notificar al usuario que indica un riesgo de cáncer en función de los resultados de los múltiples ensayos (por ejemplo, determinación del estado de metilación de múltiples DMR, por ejemplo, como se proporciona en la Tabla 1, por ejemplo, como se proporciona en la Tabla 10). 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, determinación de los estados de metilación de múltiples marcadores (tales como múltiples DMR, por ejemplo, como se proporciona en la Tabla 1, por ejemplo, como se proporciona en la Tabla 10). En algunas realizaciones, el estado de metilación un DMR define una dimensión y puede haber valores en un espacio multidimensional y la coordenada definido por los estados de metilación de los múltiples DMR es un resultado, por ejemplo, a notificar al usuario, por ejemplo, relacionado con el riesgo de cáncer.

Algunas realizaciones comprenden un medio de almacenamiento y componentes de memoria. Los componentes de memoria (por ejemplo, memoria volátil y/o no volátil) es de utilidad para almacenar instrucciones (por ejemplo, una realización de un procedimiento tal como se proporciona en el presente documento) y/o datos (por ejemplo, un resultado de trabajo tales como mediciones de la metilación, secuencias, y descripciones estadísticas asociadas con los anteriores). Algunas realizaciones se refieren a sistemas que también comprenden uno o más de una CPU, una tarjeta gráfica, y una interfaz de usuario (por ejemplo, que comprende un dispositivo de salida tal como una pantalla y un dispositivo de entrada tal como un teclado).

Las máquinas programables asociadas con la tecnología comprenden tecnologías existentes convencionales y tecnologías en desarrollo o que aún no se ha desarrollado (por ejemplo, un ordenador cuántico, un ordenador químico, un ordenador de ADN, un ordenador óptico, un ordenador basado en espintrónica, etc.).

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

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

En algunas realizaciones, la tecnología proporcionada en el presente documento está asociada con una pluralidad de dispositivos programables que operan de forma coordinada para llevar a cabo un procedimiento como se describe en el presente documento. Por ejemplo, en algunas realizaciones, una pluralidad de ordenadores (por ejemplo, conectados a una red) puede funcionar en paralelo para recopilar y procesar los datos, por ejemplo, en una implementación de clústeres de cálculo o redes de computación o cualquier otra arquitectura informática distribuida que se basa en ordenadores completos (con CPU incorporadas, almacenamiento, fuentes de alimentación, interfaces de red, etc.) conectados a una red (privada, pública, o la Internet) mediante una interfaz de red

convencional, como Ethernet, fibra óptica, o mediante tecnología de redes inalámbricas.

Por ejemplo, algunas realizaciones proporcionan un ordenador que incluye un medio legible por el ordenador. La realización incluye una memoria de acceso aleatorio (RAM) acoplada a un procesador. El procesador ejecuta instrucciones del programa ejecutable por ordenador almacenadas en la memoria. Dichos procesadores pueden incluir un microprocesador, un ASIC, una máquina de estado, u otro procesador, y puede ser cualquiera de un número de procesadores de cálculo, tales como 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 las instrucciones que, cuando el procesador las ejecuta, hacen que el procesador lleve a cabo las etapas descritas en el presente documento.

Las realizaciones de medios legibles por ordenador incluyen, aunque no de forma limitativa, un dispositivo electrónico, óptico, magnético, u otro dispositivo de almacenamiento o transmisión que puede proporcionar a un procesador instrucciones legibles por ordenador. Otros ejemplos de medios adecuados incluyen, aunque no de forma limitativa, un disco flexible, CD-ROM, DVD, disco magnético, chip de memoria, ROM, RAM, un ASIC, un procesador configurado, todos los medios ópticos, todas las cintas magnéticas u otros medios magnéticos, o cualquier otro medio desde el que el procesador de un ordenador puede leer las instrucciones. Asimismo, otras diversas formas de medios legibles por ordenador pueden transmitir o transportar instrucciones a un ordenador, incluido un enrutador, red privada o pública, u otro dispositivo o canal de transmisión, tanto cableado como inalámbrico. Las instrucciones pueden comprender código escrito en cualquier lenguaje de programación para ordenadores adecuado, incluidos, por ejemplo, C, C++, C#, Visual Basic, Java, Python, Perl, y JavaScript.

En algunas realizaciones, los ordenadores están conectados a una red. Los ordenadores también puede incluir numerosos dispositivos externos o internos tales como un ratón, un CD-ROM, DVD, un teclado, una pantalla, u otros dispositivos de entrada y salida. Los ejemplos de ordenadores son ordenadores personales, asistentes digitales, asistentes digitales personales, teléfonos celulares, teléfonos móviles, teléfonos inteligentes, paginadores, tabletas digitales manuales, ordenadores portátiles, aplicaciones de Internet, y otros dispositivos basados en procesadores.

En general, los ordenadores relacionados con aspectos de la tecnología proporcionada en el presente documento pueden ser cualquier tipo de plataforma de tipo informático que funcione con cualquier sistema operativo, tal como Microsoft Windows, Linux, UNIX, Mac OS X, etc., capaz de soportar uno o más programas que comprenden la tecnología proporcionada en el presente documento. 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 palabras, una aplicación de hoja de cálculo, una aplicación de correo electrónico, una aplicación de mensajería instantánea, una aplicación de presentaciones, una aplicación de navegador de Internet, una aplicación de calendario/agente, y cualquier otra aplicación capaz de ser ejecutada mediante un dispositivo cliente.

Todos estos componentes, ordenadores y sistemas descritos en el presente documento como asociados con la tecnología pueden ser lógicos o virtuales.

En consecuencia, se proporciona en el presente documento una tecnología relacionada con un procedimiento para el cribado de neoplasias en una muestra obtenida de un sujeto, comprendiendo el procedimiento: analizar el estado de metilación de un marcador en una muestra obtenida de un sujeto; e identificar el sujeto que tiene una neoplasia cuando el estado de metilación del marcador es diferente de un estado de metilación del marcador analizado en un sujeto que no tiene una neoplasia, en el que el marcador comprende una base en una región metilado de manera diferencial (DMR) seleccionada entre un grupo que consiste de DMR 1-107 como se proporciona en la Tabla 1, y/o DMR 1-449 en la Tabla 10. En algunas realizaciones, el procedimiento comprende además localizar el sitio de la neoplasia en el sujeto, en el que el estado de metilación del marcador indica el sitio de la neoplasia en el sujeto. La tecnología se refiere a identificar y discriminar cánceres gastrointestinales, por ejemplo, en algunas realizaciones, la neoplasia es una neoplasia gastrointestinal. En algunas realizaciones, la neoplasia está presente en la región gastrointestinal superior del paciente y, en algunas realizaciones, la neoplasia está presente en la región gastrointestinal inferior del paciente. En realizaciones particulares, la neoplasia es una neoplasia del páncreas, una neoplasia colorrectal, una neoplasia del conducto biliar, o un adenoma. La tecnología también abarca determinar el estado o el estadio de un cáncer, por ejemplo, en algunas realizaciones, la neoplasia es precancerosa. Algunas realizaciones proporcionan procedimientos que comprenden analizar una pluralidad de marcadores, por ejemplo, que comprende analizar de 2 a 11 marcadores.

La tecnología no está limitada a la evaluación del estado de metilación. En algunas realizaciones, la evaluación del estado de metilación del marcador de 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 la extensión de la metilación de una pluralidad de bases. Además, en algunas realizaciones, el estado de metilación del marcador comprende una metilación del marcador aumentada con respecto a un estado de metilación normal del marcador. En algunas realizaciones, el estado de metilación del marcador comprende una metilación del marcador más baja con respecto a un estado de metilación del marcador normal. En algunas realizaciones, el estado de metilación del marcador comprende un patrón de metilación diferente del marcador con respecto a un estado de metilación del marcador normal.

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 es un promotor CpG de alta densidad.

- 5 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 jugo pancreático, una muestra de fluido de un quiste pancreático, una muestra de sangre (por ejemplo, plasma, suero, sangre completa), una excreción, o una muestra de orina.

- 10 Además, la tecnología no está limitada por el procedimiento utilizado para determinar el estado de metilación. En algunas realizaciones, el ensayo comprende usar la reacción en cadena de la polimerasa específica de metilación, secuenciación de ácidos nucleicos, espectrometría de masas, nucleasa específica de metilación, separación por masas, o captura de la diana. En algunas realizaciones, el análisis comprende el uso de un oligonucleótido específico de la metilación. En algunas realizaciones, la tecnología utiliza masivamente la secuenciación en paralelo (por ejemplo, secuenciación avanzada) para determinar el estado de metilación, por ejemplo, secuenciación por síntesis, secuenciación en tiempo real (por ejemplo, de una sola molécula), secuenciación con emulsión en perla, secuenciación en nanoporos, etc.

- 15 La tecnología proporciona reactivos para detectar un DMR, por ejemplo, en algunas realizaciones se proporciona un conjunto de oligonucleótidos que comprende las secuencias proporcionadas en la SEQ ID NO: 1-202. En algunas realizaciones se proporciona un oligonucleótido que comprende una secuencia complementarios a una región cromosómica que tiene una base de una a DMR, por ejemplo, un oligonucleótido sensible al estado de metilación del DMR.

- 20 La tecnología proporciona varios paneles de marcadores, por ejemplo, en algunas realizaciones, el marcador comprende una región cromosómica que tiene una anotación que es ABCB1, ADCY1, BHLHE23 (LOC63930), c13orf18, CACNA1C, chr12.133, CLEC11A, ELMO1, EOMES, GJC1, IHIF1, IKZF1, KCNK12, KCNN2, NDRG4, PCBP3, PRKCB, RSPO3, SCARF2, SLC38A3, ST8SIA1, TWIST1, VWC2, WT1, o ZNF71, y que comprende el marcador (véase, Tablas 1 y 9). Además, las realizaciones proporcionan un procedimiento para analizar las DMR de la Tabla 1 que es DMR N.º 11, 14, 15, 65, 21, 22, 23, 5, 29, 30, 38, 39, 41, 50, 51, 55, 57, 60, 61, 8, 75, 81, 82, 84, 87, 93, 94, 98, 99, 103, 104, o 107, y/o una DMR correspondiente a Chr16:58497395-58497458. Algunas realizaciones proporcionan la determinación del estado de metilación de un marcador, en el que una región cromosómica que tiene una anotación que es CLEC11A, C13ORF18, KCNN2, ABCB1, SLC38A3, CD1D, IKZF1, ADCY1, CHR12133, RSPO3, MBP3, PRKCB, NDRG4, ELMO, o TWIST1 comprende el marcador. En algunas realizaciones, los procedimientos comprenden el estado de metilación de dos marcadores, por ejemplo, un par de marcadores proporcionados en una fila de la Tabla 5.

- 25 Se proporcionan realizaciones en kit, por ejemplo, un kit que comprende un reactivo de bisulfito; y un ácido nucleico del control que comprende una secuencia derivada de una DMR seleccionada de un grupo que consiste en DMR 1-107 (de la Tabla 1) y/o una DMR seleccionada de un grupo que consiste en DMR 1-449 (de la Tabla 10) y que tiene un estado de metilación asociado con un sujeto que no tiene un cáncer. En algunas realizaciones, los kits comprenden un reactivo de bisulfito y un oligonucleótido como se describe en el presente documento. En algunas realizaciones, los kits comprenden un reactivo de bisulfito; y un ácido nucleico del control que comprende una secuencia derivada de una DMR seleccionada de un grupo que consiste en DMR 1-107 (de la Tabla 1) y/o una DMR 1-449 (de la Tabla 10) y que tiene un estado de metilación asociado con un sujeto que tiene un cáncer. Algunas realizaciones de kit comprenden un sistema de recogida de muestra para obtener una muestra de un sujeto (por ejemplo, una muestra de heces); reactivos para aislar un ácido nucleico de la muestra; un reactivo de bisulfito; y un oligonucleótido como se describe en el presente documento.

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

- 40 Se proporcionan realizaciones adicionales relacionadas con el procedimiento para cribar una neoplasia en una muestra obtenida de un sujeto, por ejemplo, un procedimiento que comprende determinar un estado de metilación de un marcador en la muestra que comprende una base en una DMR que es una o más de DMR 1-107 (de la Tabla 1) y/o una o más de DMR 1-449 (de la Tabla 10); que compara el estado de metilación del marcador de la muestra del sujeto a un estado de metilación del marcador de una muestra de control normal de un sujeto que no tiene un cáncer; y determinar un intervalo de confianza y/o un valor de p de la diferencia en el estado de metilación de la muestra del sujeto y la muestra normal del control. En algunas realizaciones, el intervalo de confianza es 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 procedimientos 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 un cáncer para identificar diferencias entre las dos secuencias; e identificar el sujeto que tiene una neoplasia cuando hay diferencias presentes.

Los sistemas para cribar una neoplasia en una muestra obtenida de un sujeto se proporcionan mediante la tecnología. Las realizaciones ilustrativas de sistemas incluyen, por ejemplo, una sistema para cribar una neoplasia en una muestra obtenida de un sujeto, comprendiendo el sistema un análisis de componentes configurado para determinar el estado de metilación de una muestra, un componente de programa informático configurado para comparar el estado de metilación de la muestra con el estado de metilación de una muestra del control o una muestra de referencia registrado en una base de datos, y un componente de alerta configurado para alertar al usuario de un el estado de metilación asociado con cáncer. En algunas realizaciones, se determina una alerta mediante un componente de programa informático que recibe los resultados de múltiples ensayos (por ejemplo, determinación de los estados de metilación de múltiples marcadores, por ejemplo, DMR, por ejemplo, como se proporciona en la Tabla 1, por ejemplo, como se proporciona en la Tabla 10) y calcular un valor o resultado a notificar basándose en los múltiples resultados. Algunas realizaciones proporcionan una base de datos de parámetros ponderados asociados con cada una de las DMR proporcionadas en el presente documento para su uso en el cálculo de un valor o resultado y/o una alerta a notificar a un usuario (por ejemplo, tal como un médico, enfermera, especialista médico, etc.). En algunas realizaciones, todos los resultados de múltiples ensayos se notifican y, en algunas realizaciones, uno o más resultados se utilizan para proporcionar una puntuación, valor, o resultado, basándose en un valor compuesto de uno o más resultados a partir de múltiples ensayos que es indicativo de un riesgo de cáncer 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 asilar un ácido nucleico, un componente para recoger una muestra tal como un componente para recoger una muestra de heces. En algunas realizaciones, el sistema comprende secuencias de ácidos nucleicos que comprenden una DMR. En algunas realizaciones, la base de datos comprende secuencias de ácidos nucleicos procedentes de sujetos que no tienen un cáncer. También se proporcionan ácidos nucleicos, por ejemplo, un conjunto de ácidos nucleicos, teniendo cada ácido nucleico una secuencia que comprende una DMR. En algunas realizaciones, el conjunto de ácidos nucleicos en los que cada ácido nucleico tiene una secuencia de un sujeto que no tiene un cáncer. Las realizaciones relacionadas con el sistema comprenden un conjunto de ácidos nucleicos como se ha descrito y una base de datos de secuencias de ácidos nucleicos asociados con el conjunto de ácidos nucleicos. Algunas realizaciones comprenden además un reactivo de bisulfito. Y, algunas realizaciones comprenden además un secuenciador de ácidos nucleicos.

Otras realizaciones serán evidentes para las personas expertas en la técnica relevante basándose en las enseñanzas contenidas en el presente documento.

Breve descripción de los dibujos

Estas y otras características, aspectos y ventajas de la presente tecnología se entenderán mejor con respecto a los siguientes dibujos:

La Figura 1 es una representación gráfica que muestra la importancia de los marcadores en un conjunto de marcadores de metilación según se determina por la disminución media en la precisión de la predicción del sitio. La Figura 2 muestra los niveles de marcador de BMP3 y NDRG4 en cepillados (cardias + esófago completo) en casos de Barrett y controles como se describe en el Ejemplo 8. La Figura 3 muestra una ABC de miR-1290 en heces como se describe en el Ejemplo 9.

Se entiende que las figuras no están necesariamente dibujadas a escala, ni tampoco los objetos de las figuras están dibujadas a escala respecto a otras. Las figuras son representaciones gráficas que se proporcionan para aportar claridad y comprensión a las diversas realizaciones de aparatos, sistemas, composiciones y procedimientos desvelados en el presente documento. Cuando sea posible, se usan los mismos números de referencia en todos los dibujos para referirse a las mismas partes o a partes análogas. Además, se deberá apreciar que los dibujos no están previstos para limitar el ámbito de las presentes enseñanzas en cualquier forma.

Descripción detallada

Se proporciona en el presente documento una tecnología relacionada con la detección de neoplasias y, en particular, pero no exclusivamente, a procedimientos, composiciones, y usos relacionados para detectar neoplasias malignas y premalignas tales como el cáncer pancreático y el cáncer colorrectal. Respecto a la tecnología descrita en el presente documento, los encabezados de sección utilizados tienen fines organizativos solamente, y no deben considerarse como limitantes de la materia sujeto de cualquier forma.

En esta descripción detallada de las diferentes realizaciones, con fines de explicación, se definen numerosos detalles específicos para proporcionar una comprensión completa de las realizaciones desveladas. Un experto en la

materia apreciará, sin embargo, que estas diversas realizaciones se pueden llevar a la práctica 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 los procedimientos se presentan y se realizan son ilustrativas, y se contempla que las secuencias se pueden variar y seguir estando incluidas en el espíritu y el ámbito de las diferentes realizaciones desveladas en el presente documento.

A menos que se defina de otra manera, todos los términos técnicos y científicos utilizados en el presente documento tienen el mismo significado que el que entiende una persona experta en la técnica a la que pertenecen las distintas realizaciones descritas en el presente documento. Cuando las definiciones de los términos en las referencias parezcan diferir de las definiciones proporcionadas en las presentes enseñanzas, la definición proporcionada en las presentes enseñanzas tendrán preeminencia.

Definiciones

Para facilitar la comprensión de la presente tecnología, se definen a continuación varios términos y expresiones. Se proporcionan 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 explícitamente asociados en el presente documento, salvo que el contexto indique claramente otra cosa. La expresión "en una realización" como se utiliza en el presente documento no se refiere necesariamente a la misma realización, aunque podría. Además, la expresión "en otra realización" tal como se usa en el presente documento, no se refiere necesariamente a una realización diferente, aunque podría. Por lo tanto, como se describe a continuación, varias realizaciones de la invención se pueden combinar fácilmente, sin apartarse del ámbito o el espíritu de la invención.

Además, como se usa en el presente documento, el término "o" es un operador "o" inclusivo y es equivalente al término "y/o" salvo que el contexto indique claramente otra cosa. El término "basado en" no es exclusivo y permite que esté basado en factores adicionales no descritos, salvo que el contexto indique claramente otra cosa. Además, a lo largo de la memoria descriptiva, el significado de "un", "uno/a" y "el/la" incluye las referencias plurales. El significado de "en" incluye "en" y "sobre".

Como se usa en el presente documento, un "ácido nucleico" o "molécula de ácido nucleico" se refiere de forma general a cualquier ácido ribonucleico o ácido desoxirribonucleico, que puede ser ARN o ADN no modificado o modificado. "Ácidos nucleicos" incluyen, sin limitación, ácidos nucleicos monocatenarios y bicatenarios. Como se usa en el presente documento, el término "ácido nucleico" también incluye el ADN como se ha descrito 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". El término "ácido nucleico" como se usa en el presente documento abarca dichas formas de ácidos nucleicos modificadas química, enzimática, o metabólicamente, así como las formas químicas de ADN características de virus y células, que incluyen, por ejemplo, células simples y complejas.

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

Como se usa en el presente documento, los términos "locus" o "región" de un ácido nucleico se refieren a una subregión de un ácido nucleico, por ejemplo, un gen o un cromosoma, un único nucleótido, una isla CpG, etc.

Los términos "complementario" y "complementariedad" se refieren a nucleótidos (por ejemplo, 1 nucleótido) o polinucleótidos (por ejemplo, una secuencia de nucleótidos) relacionados por las reglas de emparejamiento de bases. Por ejemplo, la secuencia 5'-A-G-T-3' es complementaria de la secuencia 3'-T-C-A-5'. La complementariedad puede ser "parcial", en la que solamente parte de las bases de los ácidos nucleicos están emparejadas según las reglas de emparejamiento de bases. O, puede existir una complementariedad "completa" o "total" entre los ácidos nucleicos. El grado de complementariedad entre las cadenas de ácido nucleico altera la eficacia y la resistencia de la hibridación entre las cadenas de ácidos nucleicos. Esto es de especial importancia en las reacciones de amplificación y en los procedimientos 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 las secuencias de codificación necesarias para la producción de un ARN o de un polipéptido o de sus precursores. Un polipéptido funcional se puede codificar mediante una secuencia de codificación: de longitud completa o mediante cualquier parte de la secuencia de codificación, siempre que la actividad o las propiedades funcionales deseadas (por ejemplo, actividad enzimática, unión al ligando, transducción de la señal, etc.) del polipéptido se retenga. El término "porción" cuando se usa en referencia a un gen, se refiere a fragmentos de dicho gen. Los fragmentos pueden variar en tamaño desde unos pocos nucleótidos a toda la secuencia génica menos un nucleótido. Por lo

tanto, "un nucleótido que comprende al menos una porción el gen" puede comprender fragmentos del gen o el gen completo.

El término "gen" también abarca las regiones de codificación de un gen estructural e incluye secuencias situadas adyacentes a la región de codificación en ambos extremos 5' y 3', por ejemplo, en una distancia de aproximadamente 1 kb de cada extremo, de manera que el gen corresponde a la longitud del ARNm de longitud completa (por ejemplo, que comprende secuencias de codificación, regulación, estructurales, y otras secuencias). Las secuencias situadas en 5' de la región de codificación y presentes en el ARNm se denominan como secuencias 5' no traducidas o secuencias no traducidas. Las secuencias situadas en 3' o después de la región de codificación y presentes en el ARNm se denominan secuencias 3' no traducidas o secuencias no traducidas en 3'. El término "gen" abarca tanto ADNc como las formas genómicas de un gen. En algunos microorganismos (por ejemplo, eucariotas), una forma genómica o clon de un gen contiene la región de codificación interrumpida por secuencias no codificantes denominadas "intrones" o "regiones intervinientes" o "secuencias intervinientes". Los intrones son segmentos de un gen que se transcriben en ARN nuclear (hnARN); los intrones pueden contener elementos reguladores tales como potenciadores. Los intrones se eliminan o "retiran por corte y empalme" del transcrito nuclear o primario; por tanto, los intrones están ausentes en el transcrito de ARN mensajero (ARNm). El ARNm actúa durante la traducción para especificar la secuencia o el orden de los aminoácidos en el polipéptido nascente.

Además de contener intrones, las formas genómicas de un gen también pueden incluir secuencias situadas en ambos extremos 5' y 3' de las secuencias que están presentes en el transcrito de ARN. Estas secuencias se denominan como secuencias o regiones "flanqueantes" (estas secuencias flanqueantes están situadas en 5' o 3' respecto a las secuencias no traducidas presentes en el transcrito de ARN). La región flanqueante en 5' puede contener secuencias reguladoras tales como promotores y potenciadores que controlan o alteran la transcripción del gen. La región flanqueante 3' puede contener las secuencias que dirigir la terminación o la transcripción, la escisión posterior a la transcripción, y la poliadenilación.

El término "natural" cuando se aplica en referencia a un gen se refiere a un gen que tiene las características de un gen o aislado de su fuente de origen natural. El término "natural" cuando se aplica en referencia a un producto génico se refiere a un producto génico que tiene las características de un producto génico aislado de su fuente de origen natural. La expresión "origen natural" cuando se aplica a un objeto se refiere al hecho de que un objeto se puede encontrar en la naturaleza. Por ejemplo, un polipéptido o secuencia de polinucleótidos que está presente en un organismo (incluidos los virus) que se puede aislar de una fuente en la naturaleza y que no haya sido intencionadamente modificado por la mano de una persona en el laboratorio es de origen natural. Un gen natural es frecuentemente el gen o alelo observado con mayor frecuencia en una población y, por tanto, es el que recibe la designación arbitraria de forma "normal" o "natural" del gen. En cambio, el término "modificado" o "mutante" cuando se hace referencia a un gen o a un producto génico se refiere, respectivamente, a un gen o producto génico que presenta modificaciones en la secuencia y/o en las propiedades funcionales (por ejemplo, alteración en sus características) cuando se compara con el gen o producto génico natural. Se resalta que los mutantes de origen natural se pueden aislar; estos se identifican por el hecho de que tienen características alteradas cuando se comparan con el gen o producto génico silvestre.

El término "alelo" se refiere a la variación de un gen; las variaciones incluyen, aunque no de forma limitativa, variantes y mutantes, loci polimórficos, y loci polimórficos de nucleótidos únicos, cambios de marco, y mutaciones de corte y empalme. Un alelo se puede producir de forma natural en una población o podría aparecer durante la vida de cualquier individuo en particular de la población.

Por lo tanto, los términos "variante" y "mutante" cuando se utilizan en referencia a una secuencia de nucleótidos hacen referencia a una secuencia de nucleótidos que se diferencia de otra, habitualmente relacionada, secuencia de nucleótidos ácidos. Una "variación" es una diferencia entre dos secuencia de nucleótidos diferentes; de forma típica, una secuencia es una secuencia de referencia.

"Amplificación" es un caso especial de replicación de ácido nucleico que implica especificidad de molde. Esto contrasta con la replicación no específica de molde (por ejemplo, replicación que es dependiente de molde pero no dependiente de un molde específico). La especificidad de molde se distingue en este punto de fidelidad de replicación (por ejemplo, síntesis de la secuencia polinucleotídica correcta) y de especificidad de nucleótidos (ribonucleótido o desoxinucleótido). La especificidad de molde se describe frecuentemente en términos de especificidad de "diana". Las secuencias diana son "dianas" en el sentido que deberán clasificarse de otro ácido nucleico. Las técnicas de amplificación se han diseñado fundamentalmente para esta clasificación.

La amplificación de ácidos nucleicos se refiere de forma general a la producción de múltiples copias de un polinucleótido, o de una porción del polinucleótido, partiendo de forma típica de una pequeña cantidad del polinucleótido (por ejemplo, una única molécula de polinucleótido, 10 a 100 copias de una molécula de polinucleótido, que puede ser o no exactamente el mismo), donde los productos de la amplificación, o amplicones, se pueden detectar de forma general. La amplificación de los polinucleótidos abarca una variedad de procedimientos químicos y enzimáticos. La generación de múltiples copias de ADN a partir de una o unas pocas copias de una molécula de ADN diana o molde durante una reacción en cadena de la polimerasa (PCR) o una reacción en cadena de la ligasa (LCR; véase, por ejemplo, la patente de Estados Unidos n.º 5.494.810) son formas de amplificación. Los

tipos adicionales de amplificación incluyen, aunque no de forma limitativa, PCR específica de alelo (véase, por ejemplo, la patente de Estados Unidos n.º 5.639.611), la PCR con ensamblaje (véase, por ejemplo, la patente de Estados Unidos n.º 5.965.408), amplificación dependiente de helicasa (véase, por ejemplo, la patente de los Estados Unidos n.º 7.662.594), PCR de inicio caliente (véanse, por ejemplo, las patentes de Estados Unidos con números 5.773.258 y 5.338.671), PCR específica de intersecuencia, PCR inversa (véanse, por ejemplo, Triglia, y col. (1988) *Nucleic Acids Res.*, 16:8186), PCR mediada por ligadura (véase, por ejemplo, Guilfoyle, R. y col., *Nucleic Acids Research*, 25:1854-1858 (1997); la patente de Estados Unidos n.º 5.508.169), PCR específica de metilación (véase, por ejemplo, Herman, y col., (1996) *PNAS* 93(13) 9821-9826), PCR con minicebador, amplificación de sonda multiplexada dependiente de la ligadura (véase, por ejemplo, Schouten, y col., (2002) *Nucleic Acids Research* 30(12): e57), PCR multiplexada (véase, por ejemplo, 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 con extensión por solapamiento (véase, por ejemplo, Higuchi, y col., (1988) *Nucleic Acids Research* 16(15) 7351-7367), PCR en tiempo real (véase, por ejemplo, Higuchi, y col., (1992) *Biotechnology* 10:413-417; Higuchi, y col., (1993) *Biotechnology* 11:1026-1030), PCR con transcripción inversa (véase, por ejemplo, Bustin, S.A. (2000) *J. Molecular Endocrinology* 25:169-193), PCR en fase sólida, PCR entrelazada térmica asimétrica, y la PCR Touchdown (véase, por ejemplo, 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 los polinucleótidos también se puede llevar a cabo usando PCR digital (véanse, por ejemplo, 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 de EE.UU. n.º 20070202525).

La expresión "reacción en cadena de la polimerasa" ("PCR") se refiere al procedimiento de K.B. Mullis, patentes de Estados Unidos de N.º 4.683.195, 4.683.202 y 4.965.188, que describen un procedimiento para aumentar la concentración de un segmento de una secuencia diana en una mezcla de ADN genómico sin clonación ni purificación. Este procedimiento para amplificar la secuencia diana consiste en introducir un importante exceso de dos oligonucleótidos cebadores en la mezcla de ADN que contiene la secuencia diana deseada, seguido por una secuencia precisa de ciclación térmica en presencia de una ADN polimerasa. Los dos cebadores son complementarios de sus respectivas hebras de la secuencia diana bicatenaria. Para realizar la amplificación, la mezcla se desnaturaliza, y los cebadores se hibridan con sus secuencias complementarias dentro de la molécula diana. Tras la hibridación, los cebadores se extienden con una polimerasa para formar un nuevo par de hebras complementarias. Las etapas de desnaturalización, hibridación de cebadores, y extensión con polimerasa se pueden repetir varias veces (es decir, la desnaturalización, la hibridación y la extensión constituyen un "ciclo"; puede haber numerosos "ciclos") para obtener una elevada concentración de un segmento amplificado de la secuencia diana deseada. La longitud del segmento amplificado de la secuencia diana deseada se determina mediante las posiciones relativas de los cebadores entre sí y, por tanto, su longitud es un parámetro controlable. En virtud de los aspectos repetitivos del procedimiento, el procedimiento se denomina como la "reacción en cadena de la polimerasa" ("PCR"). Puesto que los segmentos amplificados deseados de la secuencia diana se convierten en las secuencias predominantes (en términos de concentración) de la mezcla, se dice que están "amplificadas mediante PCR" y son "productos de PCR" o "amplicones".

La especificidad de molde se consigue en la mayoría de las técnicas de amplificación mediante la selección de la enzima. Las enzimas de amplificación son enzimas que, en las condiciones que se usan, procesarán solamente secuencias de ácido nucleico específicas en una mezcla heterogénea de ácido nucleico. Por ejemplo, en el caso de la Q-beta replicasa, el ARN de -MDV-1 es el molde específico de la replicasa (Kacian y col., *Proc. Natl. Acad. Sci. USA*, 69:3038 [1972]). Otro ácido nucleico no se replicará con esta enzima de amplificación. De manera similar, en el caso de la ADN polimerasa T7, esta enzima de amplificación tiene una especificidad rigurosa por sus promotores (Chamberlin y col, *Nature*, 228:227 [1970]). En el caso de la ADN polimerasa T4, la enzima no ligará los dos oligonucleótidos o polinucleótidos, donde haya un emparejamiento incorrecto entre el sustrato oligonucleótido o polinucleótido y el molde en la unión de ligadura (Wu y Wallace (1989) *Genomics* 4:560). Finalmente, las ADN polimerasas termoestables dependientes del molde (por ejemplo, las ADN polimerasas Taq y Pfu), debido a su capacidad de actuar a elevada temperatura, parece que presentan elevada especificidad por las secuencias unidades y de esta forma definidas por los cebadores; la temperatura elevada da como resultado condiciones termodinámicas que favorecen la hibridación del cebador con las secuencias diana y la no hibridación con secuencias no diana (H. A. Erlich (ed.), *PCR Technology*, Stockton Press [1989]).

Como se usa en el presente documento, la expresión "ensayo de detección de ácidos nucleicos" se refiere a cualquier procedimiento para determinar la composición de nucleótidos de un ácido nucleico de interés. El ensayo de detección de ácidos nucleicos incluye, aunque no de forma limitativa, procedimientos de secuenciación de ADN, procedimientos de hibridación de sondas, ensayos de escisión específicos de la estructura (por ejemplo, el ensayo INVADER, Hologic, Inc.) y se describen, por ejemplo, en las patentes de Estados Unidos N.º 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*, EE.UU., 97:8272 (2000), y en el documento US 2009/0253142; procedimientos de escisión por emparejamiento enzimático incorrecto (por ejemplo, Variagenics, patentes de Estados Unidos n.º 6.110.684, 5.958.692 y 5.851.770); reacción en cadena de la polimerasa; procedimientos de hibridación con ramificación (por ejemplo, Chiron, patentes de Estados Unidos N.º 5.849.481, 5.710.264, 5.124.246 y 5.624.802); replicación con círculo rodante (por ejemplo,

- patentes de Estados Unidos N.º 6.210.884, 6.183.960 y 6.235.502); NASBA (por ejemplo, patente de Estados Unidos N.º 5.409.818); tecnología de baliza molecular (por ejemplo, patente de Estados Unidos N.º 6.150.097); tecnología E-sensor (Motorola, patentes de Estados Unidos N.º 6.248.229, 6.221.583, 6.013.170 y 6.063.573); tecnología de sonda cíclica (por ejemplo, patentes de Estados Unidos N.º 5.403.711, 5.011.769, y 5.660.988); procedimientos de amplificación de la señal Dade Behring (por ejemplo, patentes de Estados Unidos N.º 6.121.001, 6.110.677, 5.914.230, 5.882.867 y 5.792.614); reacción en cadena de la ligasa (por ejemplo, Barnay Proc. Natl. Acad. Sci USA 88, 189-93 (1991)); y procedimientos de hibridación en sándwich (por ejemplo, patente de Estados Unidos N.º 5.288.609).
- 5 La expresión "ácido nucleico amplificable" se refiere a un ácido nucleico que se puede amplificar por cualquier procedimiento de amplificación. Se contempla que "ácido nucleico amplificable" comprenderá habitualmente "molde de muestra".
- 10 La expresión "molde de ADN" se refiere a un ácido nucleico procedente de una muestra que se ha analizado para determinar la presencia de una "diana" (definida a continuación). En cambio, "molde de fondo" se usa en referencia a un ácido nucleico diferente al molde de muestra que puede estar presente o no en una muestra. Lo más habitual es que el molde de fondo sea involuntario. Puede ser el resultado del arrastre o se puede deber a la presencia de ácidos nucleicos contaminantes que se deben purificar de la muestra. Por ejemplo, los ácidos nucleicos procedentes de organismos diferentes a los que se deben detectar pueden estar presente como fondo en una muestra de ensayo.
- 15 El término "cebador" se refiere a un oligonucleótido, tanto de origen natural como purificado mediante una reacción de digestión o producido sintéticamente, que es capaz de actuar como un punto de inicio de la síntesis cuando se coloca en condiciones en las que se induce la síntesis de un producto de extensión del cebador que es complementario de una hebra de ácido nucleico, (por ejemplo, en presencia de nucleótidos y de un agente inductor tal como una ADN polimerasa y a una temperatura y pH adecuados). Preferentemente, el cebador es monocatenario para obtener la mayor eficacia de amplificación, aunque de forma alternativa, puede ser bicatenario. Si es bicatenario, el cebador se trata en primer lugar para separar su hebras antes de usarse para preparar productos de extensión. Preferentemente, el cebador es un oligodesoxirribonucleótido. El cebador debe ser suficientemente largo para cebar la síntesis de los productos de extensión en presencia del agente inductor. Las longitudes exactas de los cebadores dependerán de muchos factores, entre los que se incluyen la temperatura, origen del cebador y el uso del procedimiento.
- 20 El término "sonda" se refiere a un oligonucleótido (por ejemplo, una secuencia de nucleótidos), tanto de origen natural como purificado mediante una reacción de digestión o producido sintéticamente, de forma recombinante, o mediante amplificación por PCR, que se puede hibridar con otro oligonucleótido de interés. Una sonda puede ser monocatenaria o bicatenaria. Las sondas son útiles en la detección, identificación y aislamiento de secuencias génicas específicas (por ejemplo, una "sonda de captura"). Se contempla que cualquier sonda utilizada en la presente invención puede, en algunas realizaciones, estar marcada con cualquier "molécula indicadora" de forma que se pueda detectar en cualquier sistema de detección, incluidos, aunque no de forma limitativa, enzimas (por ejemplo, ELISA, así como un ensayo histoquímico basado en enzimas), sistemas fluorescentes, radioactivos y luminiscentes. No se pretende que la presente invención esté limitada a ningún sistema o marca de detección concretos.
- 25 El término "sonda" se refiere a un oligonucleótido (por ejemplo, una secuencia de nucleótidos), tanto de origen natural como purificado mediante una reacción de digestión o producido sintéticamente, de forma recombinante, o mediante amplificación por PCR, que se puede hibridar con otro oligonucleótido de interés. Una sonda puede ser monocatenaria o bicatenaria. Las sondas son útiles en la detección, identificación y aislamiento de secuencias génicas específicas (por ejemplo, una "sonda de captura"). Se contempla que cualquier sonda utilizada en la presente invención puede, en algunas realizaciones, estar marcada con cualquier "molécula indicadora" de forma que se pueda detectar en cualquier sistema de detección, incluidos, aunque no de forma limitativa, enzimas (por ejemplo, ELISA, así como un ensayo histoquímico basado en enzimas), sistemas fluorescentes, radioactivos y luminiscentes. No se pretende que la presente invención esté limitada a ningún sistema o marca de detección concretos.
- 30 Como se usa en el presente documento, "metilación" se refiere a la metilación de citosina en las posiciones C5 o N4 de citosina, la posición N6 de adenina, y otros tipos de metilación de ácido nucleico. El ADN amplificado *in vitro* no suele estar metilado porque los procedimientos de amplificación de ADN *in vitro* típicos no retienen el patrón de metilación del molde de amplificación. Sin embargo, "ADN no metilado" o "ADN metilado" también se puede referir a un ADN amplificado cuyo molde original estaba no metilado o metilado, respectivamente.
- 35 En consecuencia, tal como se usa en el presente documento, un "nucleótido metilado" o una "base nucleotídica metilada" se refiere a la presencia de un resto metilo en una base nucleotídica, donde el resto metilo no está presente en una base nucleotídica típica reconocida. Por ejemplo, la citosina no contiene un resto metilo en su anillo de pirimidina, pero el la 5-metilcitosina contiene un resto metilo en la posición 5 de su anillo de pirimidina. Por lo tanto, la citosina no es un nucleótido metilado y la 5-metilcitosina es un nucleótido metilado. En otro ejemplo, timina contiene un resto metilo en la posición 5 de su anillo de pirimidina; sin embargo, a efectos del presente documento, 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.
- 40 Como se usa en el presente documento, una "molécula de ácido nucleico metilada" se refiere a una molécula de ácido nucleico que contiene uno o más nucleótidos metilados.
- 45 Como se usa en el presente documento, un "estado de metilación", "perfil de metilación", y "estado de metilación" de una molécula de ácido nucleico se refiere a la presencia o ausencia de una o más bases nucleotídicas metiladas en la molécula de ácido nucleico. Por ejemplo, una molécula de ácido nucleico que contiene una citosina metilada se considera metilado (por ejemplo, el estado de metilación de la molécula de ácido nucleico es metilado). Una molécula de ácido nucleico que no contiene ningún nucleótido metilado se considera no metilada.
- 50 Como se usa en el presente documento, un "estado de metilación", "perfil de metilación", y "estado de metilación" de una molécula de ácido nucleico se refiere a la presencia o ausencia de una o más bases nucleotídicas metiladas en la molécula de ácido nucleico. Por ejemplo, una molécula de ácido nucleico que contiene una citosina metilada se considera metilado (por ejemplo, el estado de metilación de la molécula de ácido nucleico es metilado). Una molécula de ácido nucleico que no contiene ningún nucleótido metilado se considera no metilada.
- 55 Como se usa en el presente documento, un "estado de metilación", "perfil de metilación", y "estado de metilación" de una molécula de ácido nucleico se refiere a la presencia o ausencia de una o más bases nucleotídicas metiladas en la molécula de ácido nucleico. Por ejemplo, una molécula de ácido nucleico que contiene una citosina metilada se considera metilado (por ejemplo, el estado de metilación de la molécula de ácido nucleico es metilado). Una molécula de ácido nucleico que no contiene ningún nucleótido metilado se considera no metilada.

El estado de metilación de una secuencia de ácido nucleico concreta (por ejemplo, un gen marcador o región de ADN como se describe en el presente documento) puede indicar el estado de metilación de cada base en la secuencia o puede indicar el estado de metilación de un subconjunto de las bases (por ejemplo, de una o más citosinas) dentro de la secuencia, o puede indicar información relativa a la densidad de metilación regional dentro de la secuencia, proporcionando o no información precisa sobre las ubicaciones de la secuencia donde se produce la metilación.

El estado de metilación de un locus nucleotídico en una molécula de ácido nucleico se refiere a la presencia o ausencia de un nucleótido metilado en un locus concreto de la molécula de ácido nucleico. Por ejemplo, el estado de metilación de una citosina en el 7º nucleótido de una molécula de ácido nucleico es 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 es no 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 opcionalmente puede representarse o indicarse mediante un "valor de metilación" (por ejemplo, que representa una frecuencia, fracción, relación, porcentaje de metilación, etc.). Se puede generar un valor de metilación, por ejemplo, cuantificando la cantidad de ácido nucleico intacto presente después de la digestión con restricción con una enzima de restricción dependiente de la metilación o comparando los perfiles de amplificación después de la reacción con bisulfito o mediante la comparación entre secuencias de ácidos nucleicos tratadas y no tratadas con bisulfito. En consecuencia, un valor, por ejemplo, un valor de metilación, representa el estado de metilación y por tanto se puede usar como indicador cuantitativo del estado de metilación en múltiples copias de un locus. Esto es de especial interés cuando es deseable comparar el estado de metilación de una secuencia de una muestra con un valor umbral o de referencia.

Como se usa en el presente documento, "frecuencia de metilación" o "porcentaje de metilación (%)" se refiere al número de casos en los que una molécula o locus están metilados con respecto al número de casos en los que la molécula o el locus están sin metilar.

Por tanto, 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 concreto relevante para la metilación. Dichas características incluyen, aunque no de forma limitativa, si cualquiera de los restos de citosina (C) dentro de esta secuencia de ADN está metilado, la situación del uno o varios restos de C metilados, la frecuencia o porcentaje de C metilada en la totalidad de una región concreta de un ácido nucleico, y las diferencias alélicas en la metilación debidas a, por ejemplo, diferencias en el origen de los alelos. Las expresiones "estatuto de metilación", "perfil de metilación", y "estado de metilación" también se refieren a la concentración relativa, concentración absoluta o patrón de C metilada o C no metilada en la totalidad de cualquier región concreta de un ácido nucleico en una muestra biológica. Por ejemplo, si el uno o varios restos de citosina (C) dentro de una secuencia de ácido nucleico están metilados, esto se puede denominar como "hipermetilado" o tener "aumento en la metilación", mientras que si el uno o varios restos de citosina (C) dentro de una secuencia de ADN no están metilados, esto se puede denominar como "hipometilado" o tener "disminución en la metilación". Análogamente, si el uno o varios restos de citosina (C) dentro de una secuencia de ácido nucleico están metilados en comparación con otra secuencia de ácido nucleico (por ejemplo, de una región diferente o de un individuo diferente, etc.) dicha secuencia se considera hipermetilada o que tiene un aumento en la metilación en comparación con la otra secuencia de ácido nucleico. Como alternativa, si el uno o varios restos de citosina (C) dentro de una secuencia de ácido nucleico no están metilados en comparación con otra secuencia de ácido nucleico (por ejemplo, de una región diferente o de un individuo diferente, etc.) dicha secuencia se considera hipometilada o que tiene menor metilación en comparación con la otra secuencia de ácido nucleico. Además, la expresión "patrón de metilación" tal como se usa en el presente documento se refiere al conjunto de sitios de nucleótidos metilados y no metilados en una región de ácido nucleico. Dos ácidos nucleicos pueden tener una frecuencia de metilación o porcentaje de metilación igual o similar pero diferentes patrones de metilación cuando el número de nucleótidos metilados y no metilados son iguales o similares en toda la región pero las ubicaciones de los nucleótidos metilados y no metilados es diferentes. Se dice que las secuencias están "metiladas de forma diferencial" o que tienen una "diferencia en la metilación" o que tienen un "estado de metilación diferente" cuando difieren en la extensión (por ejemplo, uno tiene una metilación aumentada o disminuida respecto al otro), frecuencia, o patrón de metilación. La expresión "metilación diferencial" se refiere a una diferencia en el nivel o patrón de metilación del ácido nucleico en la muestra positiva para el cáncer en comparación con el nivel o patrón de metilación del ácido nucleico en la muestra negativa para el cáncer. También se puede referir a la diferencia en niveles o patrones entre pacientes que tienen recidiva de cáncer después de la cirugía en comparación con pacientes que no presentan recidiva. La metilación diferencial y los niveles específicos de patrones de metilación del ADN son biomarcadores de pronóstico y predictivos, por ejemplo, una vez se ha definido el corte correcto o las características predictivas.

La frecuencia del estado de metilación se puede usar para describir una población de individuos o una muestra de un único individuo. Por ejemplo, un locus nucleotídico que tiene una frecuencia del 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 usar, por ejemplo, para describir el grado en el que un locus nucleotídico o una región de ácido nucleico están metilados en una población de individuos o en una colección de ácidos nucleicos. Por lo tanto, cuando la metilación en una primera población o combinación de moléculas de ácido nucleico es diferente de la metilación en una segunda

población o combinación de moléculas de ácido nucleico, la frecuencia del estado de metilación de la primera población o combinación será diferente de la frecuencia del estado de metilación de la segunda población o combinación. Dicha frecuencia se puede usar, también, por ejemplo, para describir el grado en el que un locus nucleotídico o una región de ácido nucleico están metilados en un único individuo. Por ejemplo, dicha frecuencia se puede usar para describir el grado en el que un grupo de células de una muestra de tejido están metiladas o no metiladas en un locus nucleotídico o región de ácido nucleico.

Tal como se usa en el presente documento, un "locus nucleotídico" se refiere a la ubicación de un nucleótido en una molécula de ácido nucleico. Un locus nucleotídico de un nucleótido metilado se refiere a la situación de un nucleótido metilado en una molécula de ácido nucleico.

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

Como se usa en el presente documento, una "isla CpG" se refiere a una región rica en G:C del ADN genómico que contiene un número mayor de dinucleótidos CpG con respecto al ADN genómico total. Una isla CpG puede tener al menos 100, 200, o más pares de bases de longitud, donde el contenido G:C de la región es al menos un 50 % y la relación entre la frecuencia de CpG observada y la frecuencia esperada es 0,6; en algunos casos, una isla CpG puede tener al menos 500 bases de longitud, donde el contenido G:C de la región es al menos un 55 %) y la relación entre la frecuencia de CpG observada y la frecuencia esperada es 0,65. La frecuencia de CpG observada respecto de la frecuencia esperada se puede calcular de acuerdo con el procedimiento proporcionado en Gardiner-Garden y col (1987) *J. Mol. Biol.* 196: 261-281. Por ejemplo, la frecuencia de CpG observada respecto de la frecuencia esperada se puede calcular según la fórmula $R = (A \times B) / (C \times D)$, donde R es el cociente de la frecuencia de CpG observada respecto de la frecuencia esperada, A es el número de dinucleótidos CpG en la secuencia analizada, B es el número de nucleótidos en la secuencia analizada, C es el número total de nucleótidos C en la secuencia analizada, y D es el número total de nucleótidos G en la secuencia analizada. El estado de metilación se determina de forma típica en las islas CpG, por ejemplo, en las regiones promotoras. Se apreciará que aunque otras secuencias del genoma humano sean propensas a la metilación del ADN, tales como CpA y CpT (véanse 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).

Como se usa en el presente documento, un reactivo que modifica un nucleótido de la moléculas 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 o a otro agente que pueda cambiar la secuencia de ácido nucleico de una molécula de ácido nucleico de forma que refleje el estado de metilación de la molécula de ácido nucleico. Los procedimientos para tratar una molécula de ácido nucleico con dicho reactivo pueden incluir la puesta en contacto de la molécula de ácido nucleico con el reactivo, acoplado con etapas adicionales, si se desea, para realizar el cambio deseado de secuencia de nucleótidos. Dicho cambio en la secuencia de nucleótidos de una molécula de ácido nucleico puede dar como resultado una molécula de ácido nucleico en la que cada nucleótido metilado se modifique a un nucleótido diferente. Dicho cambio en la secuencia de nucleótidos puede dar como resultado una molécula de ácido nucleico en la que cada nucleótido no metilado se modifique a un nucleótido diferente. Dicho cambio en la secuencia de nucleótidos puede dar como resultado una molécula de ácido nucleico en la que cada nucleótido seleccionado que está sin metilar (por ejemplo, cada citosina sin metilar) se modifique a 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 es un nucleótido que sea un nucleótido metilado (por ejemplo, cada citosina metilada) se modifique a un nucleótido diferente. Como se usa en el presente documento, el uso de un reactivo que modifica un nucleótido seleccionado se refiere a un reactivo que modifica un nucleótido de los cuatro nucleótidos que aparecen de forma típica en una molécula de ácido nucleico (C, G, T, y A para el ADN y C, G, U, y A para el ARN), de tal forma que el reactivo modifica el un nucleótido sin modificar los otros tres nucleótidos. En una realización ilustrativa, dicho reactivo modifica un nucleótido sin metilar seleccionado para producir un nucleótido diferente. En otra realización ilustrativa, dicho reactivo puede desaminar los nucleótidos de citosina sin metilar. Un reactivo ilustrativo es bisulfito.

Como se usa en el presente documento, la expresión "reactivo de bisulfito" se refiere a un reactivo que comprende, en algunas realizaciones, bisulfito, disulfito, hidrogenosulfito, o combinaciones de los mismos, para distinguir entre citidinas metiladas y no metiladas, por ejemplo, en secuencias de dinucleótidos CpG.

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

La expresión "MS AP-PCR" (reacción en cadena de la polimerasa con cebado arbitrario dependiente de la metilación) se refiere a la tecnología reconocida en la técnica que permite una exploración global del genoma usando cebadores ricos en CG para centrarse en las regiones que contienen con mayor probabilidad dinucleótidos

CpG, y descrita por Gonzalzo y col. (1997) Cancer Research 57: 594-599.

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

5 El término "HeavyMethyl™" se refiere a un ensayo en el que sondas de bloqueo específicas de la metilación (también denominadas en el presente documento como bloqueantes) cubren las posiciones CpG comprendidas, o cubiertas mediante, los cebadores de amplificación permiten la amplificación selectiva específica de metilación de una muestra de ácido nucleico.

10 La expresión ensayo "HeavyMethyl™ MethyLight™" se refiere al ensayo HeavyMethyl™ MethyLight™, que es una variación del ensayo MethyLight™, en el que el ensayo MethyLight™ se combina con sondas de bloqueo específicas de metilación que cubren las posiciones CpG comprendidas entre los cebadores de amplificación.

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

15 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 de Estados Unidos N.º 5.786.146.

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

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

20 Como se usa en el presente documento, un "nucleótido seleccionado" se refiere a un nucleótido de los cuatro nucleótidos que aparecen de forma típica en una molécula de ácido nucleico (C, G, T, y A para el ADN y C, G, U, y A para el ARN), y puede incluir derivados metilados de los nucleótidos habituales (por ejemplo, cuando C es el nucleótido seleccionado, tanto C metilada como C no metilada se incluyen en el significado de un nucleótido seleccionado), mientras que un nucleótido seleccionado metilado se refiere específicamente a un nucleótido
25 metilado que aparece de forma típica y nucleótidos seleccionados no metilados se refiere específicamente a un nucleótido no metilado que aparece de forma típica.

Las expresiones "enzima de restricción específica de metilación" o "enzima de restricción sensible a la metilación" se refiere 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
30 no está metilado o está semimetilado, el corte no se producirá o se producirá con una eficacia significativamente reducida si el sitio de reconocimiento está metilado. En el caso de una enzima de restricción que corta específicamente si el sitio de reconocimiento está metilado, el corte no se producirá o se producirá con una eficacia significativamente reducida si el sitio de reconocimiento no está metilado. Son preferidas las enzimas de restricción específicas de metilación, cuyo sitio de reconocimiento contiene un dinucleótido CG (por ejemplo, una secuencia de
35 reconocimiento tal como CGCG o CCCGGG). En algunas realizaciones se prefieren además enzimas de restricción que no cortan si la citosina de este dinucleótido está metilada en el átomo de carbono C5.

Como se usa en el presente documento, un "nucleótido diferente" se refiere a un nucleótido que es químicamente diferente de un nucleótido seleccionado, de forma típica, de tal forma que el nucleótido diferente tiene propiedades de emparejamiento de bases según Watson-Crick que difieren del nucleótido seleccionado, por lo cual, el nucleótido
40 que aparece de forma típica que es complementario del nucleótido seleccionado no es el mismo que nucleótido que aparece de forma típica que es complementario de 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. Tal como se usa en el presente documento, un nucleótido que es complementario del nucleótido seleccionado o que es complementario de los nucleótidos diferentes se refiere a un
45 nucleótido que se empareja por bases, en condiciones de rigurosidad elevada, con el nucleótido seleccionado o el nucleótido diferente con mayor afinidad que el emparejamiento de bases con el nucleótido complementario con tres de los cuatro nucleótidos que se producen de forma natural. Un ejemplo de complementariedad es el emparejamiento de bases según Watson-Crick del ADN (por ejemplo, A-T y C-G) y del ARN (por ejemplo, A-U y C-G). Por lo tanto, por ejemplo, los pares de bases G, en condiciones de rigurosidad elevada, con mayor afinidad por C que por el emparejamiento de bases con G, A, o T y, por lo tanto, cuando C es el nucleótido seleccionado, G es un
50 nucleótido complementario del nucleótido seleccionado.

Como se usa en el presente documento, la "sensibilidad" de un marcador dado se refiere al porcentaje de muestras que notifican un valor de metilación del ADN por encima de un valor de umbral que distingue entre muestras neoplásicas y no neoplásicas. En algunas realizaciones, se define un positivo como la neoplasia confirmada por
55 histología que notifica un valor de metilación del ADN por encima de un valor de umbral (por ejemplo, el intervalo asociado con la enfermedad), y se define un falso negativo como una neoplasia confirmada por histología que notifica un valor de metilación del ADN por debajo del valor de umbral (por ejemplo, el intervalo asociado con

ausencia de enfermedad). El valor de la sensibilidad, por lo tanto, refleja la probabilidad de que una medición de la metilación del ADN para un marcador dado obtenido de una muestra conocida de la enfermedad esté en el intervalo de las mediciones asociadas con la enfermedad. Como se define en el presente documento, la relevancia clínica del valor de sensibilidad calculado representa una estimación de la probabilidad de que un determinado marcador detecte la presencia de una dolencia clínica cuando se aplica a un sujeto con dicha dolencia.

Como se usa en el presente documento, la "especificidad" de un marcador dado se refiere al porcentaje de muestras no neoplásicas que notifican un valor de metilación del ADN por debajo de un valor de umbral que distingue entre muestras neoplásicas y no neoplásicas. En algunas realizaciones, se define un negativo como la muestra no neoplásica confirmada por histología que notifica un valor de metilación del ADN por debajo de un valor de umbral (por ejemplo, el intervalo asociado con ausencia de enfermedad) y se define un falso positivo como una muestra no neoplásica confirmada por histología que notifica un valor de metilación del ADN por encima del valor de umbral (por ejemplo, el intervalo asociado con la enfermedad). El valor de la especificidad, por lo tanto, refleja la probabilidad de que una medición de la metilación del ADN para un marcador dado obtenido de una muestra no neoplásica conocida esté en el intervalo de las mediciones asociadas con la ausencia de enfermedad. Como se define en el presente documento, la relevancia clínica del valor de especificidad calculado representa una estimación de la probabilidad de que un determinado marcador detecte la ausencia de una dolencia clínica cuando se aplica a un paciente sin dicha dolencia.

El término "ABC" tal como se usa en el presente documento es la abreviatura de "área bajo la curva". Especialmente, se refiere al área bajo una curva de características operativas del receptor (COR). La curva COR es una representación gráfica del índice de positivos verdaderos frente al índice de positivos falsos para los diferentes puntos de corte posibles de una prueba diagnóstica. Muestra la compensación entre la sensibilidad y especificidad, dependiendo del punto de corte seleccionado (cualquier aumento de sensibilidad irá acompañado de una disminución de especificidad). El área bajo una curva COR (ABC) es una medida de la precisión de una prueba diagnóstica (cuanto mayor es el área, mejor; el óptimo es 1; una prueba aleatoria tendría una curva COR representada en la diagonal con un área de 0,5; como referencia: J. P. Egan. (1975) Signal Detection Theory and ROC Analysis, Academic Press, Nueva York).

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

Como se usa en el presente documento, el término "adenoma" se refiere a un tumor benigno de origen glandular. Aunque estos crecimientos son benignos, con el tiempo, pueden evolucionar a malignos.

El término "precanceroso" o "preneoplásico" y equivalentes de los mismos se refieren a cualquier trastorno proliferativo celular que experimenta una transformación maligna.

Un "sitio" de una neoplasia, adenoma, cáncer, etc. es el tejido, órgano, tipo de célula, zona anatómica, parte del cuerpo, etc. del cuerpo de un sujeto donde la neoplasia, adenoma, cáncer, etc. se encuentra.

Como se usa en el presente documento, la aplicación de una prueba "diagnóstica" incluye la detección o identificación de una patología o dolencia en un sujeto, determinar la probabilidad que un sujeto contraiga dicha enfermedad o dolencia, determinar la probabilidad de que un sujeto con una enfermedad o dolencia responda a la terapia, determinar el pronóstico de un sujeto con una enfermedad o dolencia (o su probabilidad de progresión o regresión), y determinar el efecto de un tratamiento en un sujeto con una enfermedad o dolencia. Por ejemplo, se puede usar un diagnóstico para detectar la presencia o probabilidad de que un sujeto contraiga una neoplasia o la probabilidad de que dicho sujeto responda favorablemente a un compuesto (por ejemplo, un producto farmacéutico, por ejemplo, un fármaco) u otro tratamiento.

El término "marcador", como se usa en el presente documento, se refiere a una sustancia (por ejemplo, un ácido nucleico, o una región de un ácido nucleico) que es capaz de diagnosticar un cáncer por distinguir células cancerosas de células normales, por ejemplo, según su estado de metilación.

El término "aislado", cuando se usa en referencia a un ácido nucleico, como en "un oligonucleótido aislado" se refiere a una secuencia de ácido nucleico identificada y separada de al menos un ácido nucleico contaminante con el que habitualmente está asociada en su fuente natural. El ácido nucleico aislado está presente en una forma o escenario que es diferente del que se encuentra en la naturaleza. En cambio, ácidos nucleicos no aislados, tales como ADN y ARN, se encuentran en el estado que existen en la naturaleza. Los ejemplos de ácidos nucleicos no aislados incluyen: una secuencia de ADN dada (por ejemplo, un gen) que aparece en el cromosoma de la célula hospedadora cerca de genes vecinos; secuencias de ARN, tales como una secuencia específica de ARNm que codifica una proteína específica, que se encuentra en la célula como mezcla con otros muchos ARNm que codifican una multitud de proteínas. Sin embargo, el ácido nucleico aislado que codifica una proteína en particular incluye, a modo de ejemplo, dichos ácidos nucleicos en células que normalmente expresan la proteína, donde el ácido nucleico está en una ubicación cromosómica diferente de la de las células naturales, o bien está flanqueada por una secuencia de ácido nucleico diferente de la que se encuentra en la naturaleza. El ácido nucleico u oligonucleótido aislado puede

estar presente de forma monocatenaria o bicatenaria. Cuando un ácido nucleico u oligonucleótido aislado se va a utilizar para expresar una proteína, el oligonucleótido contendrá como mínimo la hebra de sentido directo o codificante (es decir, el oligonucleótido puede ser monocatenario), pero puede contener ambas hebras de sentido directo y de sentido contrario (es decir, el oligonucleótido puede ser bicatenario). Un ácido nucleico aislado puede, tras aislarse de su entorno natural o típico, puede combinarse con otros ácidos nucleicos o moléculas. Por ejemplo, un ácido nucleico aislado puede estar presente en una célula hospedadora en la que se haya introducido, por ejemplo, para la expresión heteróloga.

El término "purificado" se refiere a moléculas, tanto de ácido nucleico como secuencias de ácidos nucleicos que se extraen de su entorno natural, aisladas o separadas. Una "secuencia de ácido nucleico aislada" puede ser, por tanto, una secuencia de ácido nucleico purificada. Las moléculas "prácticamente purificadas" están exentas en al menos un 60 %, preferentemente exentas en al menos un 75 %, y más preferentemente exentas en al menos un 90 % de otros componentes con los que están asociadas. Como se usa en el presente documento, los términos "purificado" o "purificar" también se refieren a la eliminación de los contaminantes de una muestra. La eliminación de las 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 vegetales, bacterianas, de levadura, o de mamífero, y los polipéptidos se purifican por eliminación de las proteínas de la célula hospedadora; por tanto, el porcentaje de polipéptidos recombinantes aumenta de esta forma en la muestra.

La expresión "composición que comprende" una secuencia polinucleotídica o polipeptídica dada se refiere ampliamente a cualquier composición que comprende la secuencia polinucleotídica o polipeptídica dada. La composición puede comprender una solución acuosa que contiene sales (por ejemplo, NaCl), detergentes (por ejemplo, SDS), y otros componentes (por ejemplo, solución de Denhardt, leche en polvo, ADN de esperma de salmón, etc.).

El término "muestra", se usa en su sentido más amplio. En un sentido, puede referirse a una célula o tejido animal. En otro sentido, se entiende que muestra incluye un espécimen o cultivo obtenido de cualquier origen, así como muestras biológicas y ambientales. Las muestras biológicas se pueden obtener de plantas o animales (incluidos los seres humanos) y abarcan fluidos, sólidos, tejidos y gases. Las muestras ambientales incluyen material ambiental tal como material de la superficie, suelo, agua, y muestras industriales. Estos ejemplos no deben tomarse como una limitación de los tipos de muestra aplicables en la presente invención.

Como se usa en el presente documento, una "muestra remota", como se utiliza en algunos contextos, se refiere a una muestra recogida indirectamente de un sitio que no es la célula, tejido, u órgano fuente de la muestra. Por ejemplo, cuando el material de la muestra originario del páncreas se evalúa en una muestra de heces (por ejemplo, no de una muestra tomada directamente del páncreas), la muestra es una muestra remota.

Como se usa en el presente documento, los términos "paciente" o "sujeto" se refieren a organismos que serán objeto de varios ensayos proporcionados por la tecnología. El término "sujeto" incluye animales, preferentemente mamíferos, incluidos los seres humanos. En una realización preferida, el sujeto es un primate. En una realización incluso más preferida, el sujeto es un ser humano.

Como se usa en el presente documento, el término "kit" se refiere a cualquier sistema de administración para administrar materiales. En el contexto de ensayos de reacción, dichos sistemas de administración incluyen sistemas que permiten el almacenamiento, transporte o administración de los reactivos de reacción (por ejemplo, oligonucleótidos, enzimas, etc. en los recipientes adecuados) y/o materiales de soporte (por ejemplo, tampones, instrucciones por escrito para realizar el ensayo, etc.) de un sitio a otro. Por ejemplo, los kits incluyen uno o más recipientes (por ejemplo, cajas) que contienen los reactivos de reacción y/o materiales de soporte más importantes. Como se usa en el presente documento, el término "kit fragmentado" se refiere a sistemas de administración que comprenden dos o más recipientes separados, donde cada uno contiene una subparte de la totalidad de los componentes del kit. Los recipientes se pueden suministrar al recipiente previsto juntos o por separado. Por ejemplo, un primer recipiente puede contener una enzima para su uso en un ensayo, mientras que un segundo recipiente contiene oligonucleótidos. La expresión "kit fragmentado" pretende abarcar kits que contienen reactivos específicos de análisis (ASR, por sus siglas en inglés) regulados de conformidad al artículo 520(e) de la ley federal estadounidense de alimentos, fármacos y productos cosméticos [Federal Food, Drug, and Cosmetic Act, pero no está limitada a estos. De hecho, cualquier sistema de administración que comprenda dos o más recipientes separados, donde cada uno contiene una subparte de la totalidad de los componentes del kit, está incluida en la expresión "kit fragmentado". En cambio, un "kit combinado" se refiere a un sistema de administración que contiene todos los componentes de un ensayo de reacción en un solo recipiente (por ejemplo, en una sola caja que aloja cada uno de los componentes deseados). El término "kit" incluye tanto kits fragmentados como kits combinados.

Realizaciones de la tecnología

Se proporciona en el presente documento tecnología para marcadores de cáncer pancreático y otros marcadores de cribado del cáncer gastrointestinal que proporcionan una alta relación señal-ruido y un bajo nivel de fondo cuando se detecta en muestras tomadas de un sujeto (por ejemplo muestra de heces). Los marcadores se identificaron en un estudio de control de caso comparando el estado de metilación de marcadores de ADN procedentes de sujetos con

PanC en estadio I y II con el estado de metilación de los mismos marcadores de ADN procedentes de sujetos de control (por ejemplo, tejido normal tal como colon normal y/o páncreas no neoplásico (véanse, los Ejemplos 1 y 11).

5 Los marcadores y/o paneles de marcadores (Por ejemplo, una región cromosómica que tiene una anotación seleccionada entre ABCB1, ADCY1, BHLHE23 (LOC63930), c13orf18, CACNA1C, chr12 133, CLEC11A, ELMO1, EOMES, CLEC 11, SHH, GJC1, IHIF1, IKZF1, KCNK12, KCNN2, PCBP3, PRKCB, RSPO3, SCARF2, SLC38A3, ST8SIA1, TWIST1, VWC2, WT1, y ZNF71) se identificaron en un estudio de control de caso comparando el estado de metilación de marcadores de ADN (por ejemplo, procedentes de sujetos con PanC en estadio I y II con el estado de metilación de los mismos marcadores de ADN procedentes de sujetos de control (por ejemplo, tejido normal tal como colon normal y/o páncreas no neoplásico (véanse, los Ejemplos 2 y 8).

10 Los marcadores y/o paneles de marcadores (Por ejemplo, una región cromosómica que tiene una anotación seleccionada entre NDRG4, SFRP1, BMP3, HPP1, y/o APC) se identificaron en estudios de control del caso comparando el estado de metilación de marcadores del ADN procedentes de tejido esofágico de sujetos con esófago de Barrett con el estado de metilación de los mismos marcadores de ADN procedentes de sujetos del control (véanse los Ejemplo 4 y 10).

15 Los marcadores y/o paneles de marcadores (Por ejemplo, una región cromosómica que tiene una anotación seleccionada entre ADCY1, PRKCB, KCNK12, C13ORF18, IKZF1, TWIST1, ELMO, 55957, CD1D, CLEC11A, KCNN2, BMP3, y/o NDRG4) se identificaron en estudios de control del caso comparando el estado de metilación de marcadores del ADN procedentes de una muestra de jugo pancreático de sujetos con cáncer de páncreas con el estado de metilación de los mismos marcadores de ADN procedentes de sujetos del control (véanse los Ejemplo 5 y 6).

Se identificó un marcador (por ejemplo, una región cromosómica que tiene una anotación CD1D) en un estudio de control del caso por comparación del estado de metilación de un marcador de ADN (por ejemplo, CD1D) procedente de una muestra de heces de sujetos con cáncer de páncreas con el estado de metilación del mismo marcador de ADN procedentes de sujetos del control que no tienen cáncer de páncreas (véase, el Ejemplo 7).

25 Se identificó un marcador (por ejemplo, miR-1290) en un estudio de control del caso comparando la cantidad cuantificada de un marcador de ADN (por ejemplo, miR-1290) de una muestra de heces de sujetos con cáncer de páncreas con la cantidad cuantificada del mismo marcador de ADN procedentes de sujetos de control que no tienen cáncer de páncreas (véase, Ejemplo 9).

30 Además, la tecnología proporciona varios paneles de marcadores, por ejemplo, en algunas realizaciones, el marcador comprende una región cromosómica que tiene una anotación que es ABCB1, ADCY1, BHLHE23 (LOC63930), c13orf18, CACNA1C, chr12.133, CLEC11A, ELMO1, EOMES, GJC1, IHIF1, IKZF1, KCNK12, KCNN2, NDRG4, PCBP3, PRKCB, RSPO3, SCARF2, SLC38A3, ST8SIA1, TWIST1, VWC2, WT1, o ZNF71, y que comprende el marcador (véase, Tablas 1 y 9). Además, las realizaciones proporcionan un procedimiento para analizar las DMR de la Tabla 1 que es la DMR N.º 11, 14, 15, 65, 21,22, 23, 5, 29, 30, 38, 39, 41, 50, 51, 55, 57, 60, 61, 8, 75, 81, 82, 84, 87, 93, 94, 98, 99, 103, 104, o 107, y/o una DMR correspondiente a Chr16:58497395-58497458. Algunas realizaciones proporcionan la determinación del estado de metilación de un marcador, en el que una región cromosómica que tiene una anotación que es CLEC11A, C13ORF18, KCNN2, ABCB1, SLC38A3, CD1D, IKZF1, ADCY1, CHR12133, RSPO3, MBP3, PRKCB, NDRG4, ELMO, o TWIST1 comprende el marcador. En algunas realizaciones, los procedimientos comprenden el estado de metilación de dos marcadores, por ejemplo, un par de marcadores proporcionados en una fila de la Tabla 5.

Aunque la divulgación del presente documento se refiere a determinadas realizaciones ilustradas, se debe entender que dichas realizaciones se presentan a modo de ejemplo y no como limitación.

45 En aspectos particulares, la presente tecnología proporciona composiciones y procedimientos para identificar, determinar, y/o clasificar un cáncer tal como un cáncer gastrointestinal superior (por ejemplo, cáncer del esófago, páncreas, estómago) o cáncer gastrointestinal inferior (por ejemplo, adenoma, cáncer colorrectal). En aspectos relacionados, la tecnología proporciona composiciones y procedimientos para identificar, predecir, y/o detectar, el sitio de un cáncer. Los procedimientos comprenden determinar el estado de metilación de al menos un marcador de la metilación en una muestra biológica aislada de un sujeto, en los que un cambio en el estado de metilación del marcador es indicativo de la presencia, clase, o sitio de un cáncer. Las realizaciones particulares se refieren a marcadores que comprenden una región diferencialmente metilada (DMR, por ejemplo, DMR 1-107, véase la Tabla 50 1, por ejemplo, DMR 1-449, véase la Tabla 10) que se utilizan para el diagnóstico (por ejemplo, cribado) de trastornos celulares proliferativos neoplásicos (por ejemplo, cáncer), incluida la detección precoz durante las etapas precancerosas y en la predicción el sitio de una neoplasia (por ejemplo, discriminando entre tipos de cáncer, por ejemplo, cánceres gastrointestinales superiores y cánceres gastrointestinales inferiores). Además, los marcadores se utilizan para la diferenciación entre neoplasias y trastornos proliferativos benignos. En aspectos particulares, la presente tecnología desvela un procedimiento en el que un trastorno celular proliferativo neoplásico se distingue de un trastorno celular proliferativo benigno.

Los marcadores de la presente tecnología son especialmente eficaces para detectar o distinguir entre los trastornos

proliferativos colorrectales y pancreáticos, proporcionando de esta manera medios mejorados para la detección precoz, clasificación y tratamiento de dichos trastornos.

Además de las realizaciones en las que el análisis de metilación de al menos un marcador, se analiza una región de un marcador, o una base de un marcador que comprende una DMR (por ejemplo, DMR 1-107 de la Tabla 1) (por ejemplo, DMR 1-449 de la Tabla 10) proporcionada en el presente documento y relacionada en la Tabla 1 o 10, la tecnología proporciona también paneles de marcadores que comprenden al menos un marcador, región de un marcador, o base de un marcador que comprende una DMR con utilidad en la detección de cánceres, especialmente el cáncer colorrectal, cáncer pancreático, y otros cánceres del tracto gastrointestinal superior e inferior.

Algunas realizaciones de la tecnología se basan en el análisis del estado de metilación de CpG en 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 la técnica del bisulfito junto con uno o más ensayos de metilación para determinar el estado de metilación de secuencias de dinucleótidos CpG dentro de al menos un marcador que comprenden una DMR (por ejemplo, como se proporciona en la Tabla 1 (por ejemplo, DMR 1-107)) (por ejemplo, como se proporciona en la Tabla 10 (por ejemplo, DMR 1-449)). Los dinucleótidos CpG genómicos pueden estar metilados o no metilados (conocido de manera alternativa como sobremetilados o con defecto de metilación, respectivamente). Sin embargo, los procedimientos de la presente invención son adecuados para el análisis de muestras biológicas de naturaleza heterogénea, por ejemplo, una baja concentración de células tumorales, o materiales biológicos derivados de las mismas, dentro de un fondo de una muestra remota (por ejemplo, sangre, efluente de órgano, o heces). En consecuencia, cuando se analiza el estado de metilación de una posición 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 CpG concreta.

De acuerdo con la presente tecnología, la determinación del estado de metilación de las secuencias del dinucleótido CpG en marcadores que comprenden una DMR tiene utilidad tanto en el diagnóstico como en la caracterización de cánceres tales como el cáncer gastrointestinal superior (por ejemplo, cáncer de esófago, páncreas, estómago) o cáncer gastrointestinal inferior (por ejemplo, adenoma, cáncer colorrectal).

Combinaciones de marcadores

En algunas realizaciones, la tecnología se refiere a la evaluación del estado de metilación de combinaciones de marcadores que comprenden una DMR de la Tabla 1 (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 la Tabla 10 (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 DMR. En algunas realizaciones, la evaluación del estado de metilación de más de un marcador aumenta la especificidad y/o sensibilidad de un cribado o diagnóstico para identificar una neoplasia en un sujeto, por ejemplo, un cáncer gastrointestinal superior (por ejemplo, esófago, páncreas, estómago) o un cáncer gastrointestinal inferior (por ejemplo, adenoma, colorrectal). En algunas realizaciones, un marcador o combinación de marcadores discrimina entre tipos y/o ubicaciones de una neoplasia. Por ejemplo, combinaciones de marcadores discriminan neoplasia esofágica, neoplasia del estómago, neoplasia pancreática, neoplasia colorrectal, y adenomas entre sí, de otras neoplasias, y/o entre tejido normal (por ejemplo, no canceroso, no precanceroso) tissue.

Se predicen varios cánceres mediante varias combinaciones de marcadores, por ejemplo, según se identifican por técnicas estadísticas relacionadas con la especificidad y la sensibilidad de la predicción. La tecnología proporciona procedimientos para identificar combinaciones predictivas y para validar combinaciones predictivas para algunos cánceres.

En algunas realizaciones, combinaciones de marcadores (por ejemplo, que comprenden una DMR) predicen el sitio de una neoplasia. Por ejemplo, durante el desarrollo de la tecnología descrita en el presente documento, se realizaron análisis estadísticos para validar los resultados de sensibilidad y especificidad de combinaciones de marcadores. Por ejemplo, las parejas de marcadores predijeron con precisión el sitio del tumor en >90 % de las muestras, las 17 principales parejas de marcadores predijeron con precisión el sitio del tumor en >80 % de las muestras, y las 49 principales parejas de marcadores predijeron con precisión el sitio del tumor en el 70 % de las muestras.

Procedimientos para evaluar el estado de metilación

El procedimiento usado con más frecuencia para analizar un ácido nucleico para determinar la presencia de 5-metilcitosina se basa en el procedimiento 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 procedimiento del bisulfito para cartografiar 5-metilcitosinas se basa en la observación de que citosina, pero no 5-metilcitosina, reacciona con el ion hidrogenosulfito (también denominado como bisulfito). La reacción se realiza habitualmente de acuerdo con las siguientes etapas: en primer lugar, la citosina reacciona con hidrogenosulfito para formar una citosina sulfonada. A continuación, la desaminación espontánea del intermedio de reacción sulfonado da como resultado un uracilo sulfonado. Finalmente, el uracilo sulfonado se desulfona en condiciones alcalinas para formar uracilo. La detección es posible porque el uracilo forma pares de bases con adenina (de ahí que comporte

como la timina), mientras que la 5-metilcitosina se empareja con la guanina (de ahí que comporte como citosina). Esto hace posible la discriminación entre las citosinas metiladas y las citosinas no metiladas, por ejemplo, secuencia genómica con bisulfito (Grigg G, & Clark S, *Bioessays* (1994) 16: 431-36; Grigg G, *DNA Seq.* (1996) 6: 189-98) o PCR específica de metilación PCR (MSP) como se desvela, por ejemplo, en la Patente de los Estados Unidos n.º 5.786.146.

Algunas tecnologías convencionales están relacionadas con procedimientos que comprenden encerrar el ADN a analizar en una matriz de agarosa, evitando de esta forma la difusión y la renaturalización del ADN (el bisulfito solamente reacciona con ADN monocatenario), y sustituir 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). Es por tanto posible analizar células individuales para determinar su estado de metilación, ilustrado la utilidad y la sensibilidad del procedimiento. Una descripción general de procedimientos convencionales para detectar 5-metilcitosina se proporciona en Rein, T., y col. (1998) *Nucleic Acids Res.* 26: 2255.

La técnica del bisulfito implica, de forma típica, una amplificación de fragmentos cortos específicos de un ácido nucleico conocido después de un tratamiento con bisulfito, a continuación analizar el producto por secuenciación (Olek & Walter (1997) *Nat. Genet.* 17: 275-6) o una reacciones de extensión del cebador (Gonzalzo & Jones (1997) *Nucleic Acids Res.* 25: 2529-31; documento WO 95/00669; patente de Estados Unidos N.º 6.251.594) para analizar las posiciones de citosina individuales. Algunos procedimientos utilizan digestión enzimática (Xiong & Laird (1997) *Nucleic Acids Res.* 25: 2532-4). La detección mediante hibridación también se ha descrito en la técnica (Olek y col., WO 99/28498). Además, el uso de la técnica del bisulfito para la detección de la metilación con respecto a los genes individuales se ha descrito (Grigg & 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; documento WO 9746705; documento WO 9515373).

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

Por ejemplo, la secuenciación genómica se ha simplificado mediante el análisis de los patrones de metilación y las distribuciones de 5-metilcitosina mediante el uso del tratamiento con bisulfito (Frommer y col. (1992) *Proc. Natl. Acad. Sci. USA* 89: 1827-1831). Además, la digestión de los productos de la PCR con enzimas de restricción amplificados a partir del ADN convertido con bisulfito es de utilidad para evaluar el estado de la metilación, por ejemplo, como se describe en Sadri & Hornsby (1997) *Nucl. Acids Res.* 24: 5058-5059 o como se realiza en el procedimiento conocido como COBRA (análisis combinado con restricción de bisulfito) (Xiong & Laird (1997) *Nucleic Acids Res.* 25: 2532-2534).

El análisis COBRA™ es un ensayo de metilación cuantitativo útil para determinar los niveles de metilación del ADN en loci específicos en pequeñas cantidades de ADN genómico (Xiong & Laird, *Nucleic Acids Res.* 25:2532-2534, 1997). En resumen, la digestión con enzimas de restricción se utiliza para revelar las diferencias en la secuencia dependientes de metilación en productos de la PCR de ADN tratado con bisulfito. Las diferencias en la secuencia dependientes de metilación se introducen por vez primera en el ADN genómico mediante tratamiento con bisulfito convencional de acuerdo con el procedimiento descrito por Frommer y col. (*Proc. Natl. Acad. Sci. USA* 89:1827-1831, 1992). La amplificación mediante la PCR del ADN transformado con bisulfito se realiza a continuación usando cebadores específicos de las islas CpG de interés, seguido por la digestión con endonucleasa de restricción, electroforesis en gel, y detección usando sondas de hibridación específicas marcadas. Los niveles de metilación en la muestra de ADN original se representan por las cantidades relativas de producto de la PCR digerido y sin digerir de una forma linealmente cuantitativa para una amplia gama de niveles de metilación del ADN. Además, esta técnica se puede aplicar con fiabilidad al ADN obtenido de muestras de tejidos microdesecadas incluidas en parafina.

Los reactivos típicos (por ejemplo, los que se encontrarían en un kit típico basado en COBRA™) para el análisis COBRA™ pueden incluir, aunque no de forma limitativa: cebadores de la PCR para loci específicos (por ejemplo, genes específicos, marcadores, DMR, regiones de genes, regiones de marcadores, secuencia de ADN tratada con bisulfito, isla CpG, etc.); enzimas de restricción y el tampón adecuado; oligonucleótido para hibridación del gen; oligonucleótido para hibridación del control; kit de marcado con quinasa para la sonda de oligonucleótido; y nucleótidos marcados. Además, los reactivos para la conversión con bisulfito pueden incluir: tampón de desnaturalización de ADN; tampón de sulfonación; reactivos o kits de recuperación del ADN (por ejemplo, precipitación, ultrafiltración, columna de afinidad); tampón de desulfonación; y componentes de recuperación del ADN.

Preferentemente, los ensayos como "MethyLight™" (una técnica de la PCR en tiempo real basada en fluorescencia) (Eads y col., *Cancer Res.* 59:2302-2306, 1999), reacciones Ms-SNuPE™ (extensión del cebador de nucleótido único sensible a la metilación) (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; patente de Estados Unidos N.º 5.786.146) y

amplificación de isla CpG metilada ("MCA"; Toyota y col., Cancer Res. 59:2307-12, 1999) se usan en solitario o combinados con uno o más de estos procedimientos.

La técnica del ensayo "HeavyMethyl™" es un procedimiento cuantitativo para evaluar las diferencias en la metilación basándose en la amplificación específica de la metilación del ADN tratado con bisulfito. Las sondas de bloqueo específicas de la metilación ("bloqueante") que cubren las posiciones CpG comprendidas entre, cubiertas entre, los cebadores de amplificación permiten la amplificación selectiva específica de metilación de una muestra de ácido nucleico.

La expresión ensayo "HeavyMethyl™ MethyLight™" se refiere al ensayo HeavyMethyl™ MethyLight™, que es una variación del ensayo MethyLight™, en el que el ensayo MethyLight™ se combina con sondas de bloqueo específicas de metilación que cubren las posiciones CpG comprendidas entre los cebadores de amplificación. El ensayo HeavyMethyl™ también se puede usar junto con cebadores de amplificación específicos de la metilación.

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

MSP (PCR específica de metilación) permite evaluar el estado de metilación de prácticamente cualquier grupo de sitios CpG dentro de una isla CpG, independiente del uso de enzimas de restricción sensibles a la metilación (Herman y col. Proc. Natl. Acad. Sci. USA 93:9821-9826, 1996; patente de Estados Unidos n.º 5.786.146). En resumen, el ADN se modifica con bisulfito de sodio, que convierte las citosinas no metiladas, pero no citosinas metiladas, en uracilo, y posteriormente los productos se amplifican con cebadores específicos de ADN metilado en comparación con no metilado. MSP solo necesita pequeñas cantidades de ADN, es sensible al 0,1 % de alelos metilados de un locus de isla CpG dado, y se puede realizar sobre ADN extraído de muestras incluidas en parafina. Los reactivos típicos (por ejemplo, como se encontraría en un kit basado en MSP típico) para el análisis MSP puede incluir, aunque no de forma limitativa: cebadores de la PCR para loci específicos metilados y no metilados (por ejemplo, genes específicos, marcadores, DMR, regiones de genes, regiones de marcadores, secuencia de ADN tratada con bisulfito, isla CpG, etc.); tampones y desoxinucleótidos optimizados para la PCR, y sondas específicas.

El ensayo MethyLight™ 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 después de la etapa de la PCR (Eads y col., Cancer Res. 59:2302-2306, 1999). En resumen, el procedimiento MethyLight™ se inicia con una muestra mixta de ADN genómico que se convierte, en una reacción con bisulfito de sodio, en una combinación mixta de diferencias de secuencia dependientes de la metilación de acuerdo con procedimientos convencionales (el procedimiento del bisulfito convierte los restos de citosina no metilados en uracilo). A continuación se realiza la PCR basada en fluorescencia de forma "sesgada", por ejemplo, con cebadores de la PCR que solapan dinucleótidos CpG conocidos. La discriminación de secuencia se produce tanto durante el procedimiento de amplificación como durante el procedimiento de detección de fluorescencia.

El ensayo MethyLight™ se utiliza como prueba cuantitativa para determinar los patrones de metilación en un ácido nucleico, por ejemplo, una muestra de ADN genómico, en la que la discriminación de secuencia se produce durante la hibridación de la sonda. En una versión cuantitativa, la reacción de la PCR proporciona una amplificación específica de la metilación en presencia de una sonda fluorescente que se superpone a un sitio de metilación concreto posible. Se proporciona un control no sesgado de la cantidad de ADN inicial mediante una reacción en la que ninguno de los cebadores, ni la sonda, solapa con ninguno de los dinucleótidos CpG. Como alternativa, se consigue una prueba cualitativa de la metilación genómica sondeando la combinación de la PCR sesgada bien con oligonucleótidos del control que no cubren sitios de metilación conocidos (por ejemplo, una versión fluorescente de las técnicas HeavyMethyl™ y MSP) o con oligonucleótidos que cubren sitios de metilación potenciales.

El procedimiento MethyLight™ se utiliza con cualquier sonda adecuada (por ejemplo, una sonda "TaqMan®", una sonda Lightcycler®, etc. Por ejemplo, en algunas aplicaciones, el ADN bicatenario se trata con bisulfito de sodio y se somete a uno de dos grupos de reacciones de PCR utilizando sondas TaqMan®, por ejemplo, con cebadores MSP y/o oligonucleótidos bloqueantes HeavyMethyl y una sonda TaqMan®. La sonda TaqMan® está doblemente marcada con moléculas fluorescentes "indicadoras" y moléculas "inactivadoras", y está diseñado para ser específica de un contenido relativamente elevado en la región GC, por lo que se hibrida a una temperatura a una temperatura aproximadamente 10 °C mayor en el ciclo de la PCR que los cebadores directos e inversos. Esto permite que la sonda TaqMan® permanezca completamente hibridada durante la etapa de hibridación/extensión de la PCR. Puesto que la polimerasa Taq sintetiza enzimáticamente una nueva hebra durante la PCR, eventualmente alcanzará la sonda TaqMan® hibridada. La actividad de la polimerasa Taq 5' a 3' endonucleasa posteriormente desplazará la sonda TaqMan® digiriéndola para liberar la molécula indicadora fluorescente para detección cuantitativa de su señal ahora no inactivada usando un sistema de detección de fluorescencia en tiempo real.

Los reactivos típicos (por ejemplo, como se encontraría en un kit basado en MethyLight™ típico) para el análisis MethyLight™ pueden incluir, aunque no de forma limitativa: cebadores de la PCR para loci específicos (por ejemplo,

genes específicos, marcadores, DMR, regiones de genes, regiones de marcadores, secuencia de ADN tratada con bisulfito, isla CpG, etc.); sondas TaqMan® o Lightcycler®; tampones y desoxinucleótidos optimizados para la PCR; y polimerasa Taq.

5 El ensayo QM™ (metilación cuantitativa) es una prueba cuantitativa alternativa para determinar patrones de metilación en muestras de ADN genómico, en la que la discriminación de secuencia se produce durante la hibridación de la sonda. En esta versión cuantitativa, la reacción de la PCR proporciona una amplificación no sesgada en presencia de una sonda fluorescente que se superpone a un sitio de metilación concreto posible. Se proporciona un control no sesgado de la cantidad de ADN inicial mediante una reacción en la que ninguno de los cebadores, ni la sonda, solapa con ninguno de los dinucleótidos CpG. Como alternativa, se consigue una prueba
10 cualitativa de la metilación genómica sondeando la combinación de la PCR sesgada bien con oligonucleótidos del control que no cubren sitios de metilación conocidos (una versión fluorescente de las técnicas HeavyMethyl™ y MSP) o con oligonucleótidos que cubren sitios de metilación potenciales.

El procedimiento QM™ se puede usar con cualquier sonda adecuada, por ejemplo, sondas "TaqMan®", sondas Lightcycler®, en el procedimiento de amplificación. Por ejemplo, el ADN genómico bicatenario se trata con bisulfito de sodio y se somete a uno cebadores no sesgados y a la sonda TaqMan®. La sonda TaqMan® está doblemente
15 marcada con moléculas fluorescentes "indicadoras" y moléculas "inactivadoras", y está diseñado para ser específica de un contenido relativamente elevado en la región GC, por lo que se hibrida a una temperatura a una temperatura aproximadamente 10 °C mayor en el ciclo de la PCR que los cebadores directos e inversos. Esto permite que la sonda TaqMan® permanezca completamente hibridada durante la etapa de hibridación/extensión de la PCR. Puesto
20 que la polimerasa Taq sintetiza enzimáticamente una nueva hebra durante la PCR, eventualmente alcanzará la sonda TaqMan® hibridada. La actividad de la polimerasa Taq 5' a 3' endonucleasa posteriormente desplazará la sonda TaqMan® digiriéndola para liberar la molécula indicadora fluorescente para detección cuantitativa de su señal ahora no inactivada usando un sistema de detección de fluorescencia en tiempo real. Los reactivos típicos (por ejemplo, como se encontraría en un kit basado en QM™ típico) para el análisis QM™ pueden incluir, aunque no de forma limitativa: cebadores de la PCR para loci específicos (por ejemplo, genes específicos, marcadores, DMR, regiones de genes, regiones de marcadores, secuencia de ADN tratada con bisulfito, isla CpG, etc.); sondas TaqMan® o Lightcycler®; tampones y desoxinucleótidos optimizados para la PCR; y polimerasa Taq.

La técnica Ms-SNuPE™ es un procedimiento cuantitativo para evaluar las diferencias en la metilación en sitios CpG específicos basándose en el tratamiento del ADN con bisulfito, seguido por extensión del cebador de nucleótido único (Gonzalzo & Jones, Nucleic Acids Res. 25:2529-2531,1997). En resumen, el ADN genómico se hace
30 reaccionar con bisulfito de sodio para convertir la citosina no metilada en uracilo, dejando inalterada la 5-metilcitosina. La amplificación de la secuencia diana deseada se realiza a continuación usando cebadores de la PCR específicos del ADN transformado con bisulfito, y el producto de reacción resultante se aísla y se utiliza como molde para el análisis de metilación en el sitio CpG de interés. Se pueden analizar pequeñas cantidades de ADN (por ejemplo, secciones microseccionadas de patología) y esto evita el uso de enzimas de restricción para determinar el estado de metilación en sitios CpG.

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

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

Un protocolo típico para RRBS comprende las etapas de digerir una muestra de ácido nucleico con una enzima de restricción tal como MspI, rellenar con salientes y colas A, ligar adaptadores, conversión con bisulfito, y PCR. Véase, por ejemplo, y col. (2005) "Genome-scale DNA methylation mapping of clinical samples at single-nucleotide

resolution" Nat Methods 7: 133-6; Meissner y col. (2005) "Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis" Nucleic Acids Res. 33: 5868-77.

En algunas realizaciones, se utiliza un ensayo cuantitativo para la amplificación de la diana y la señal en tiempo real específica de alelo (QuARTS) para evaluar el estado de metilación. Se producen secuencialmente tres reacciones en cada ensayo QuARTS, incluidas amplificación (reacción 1) y escisión de la sonda de la diana (reacción 2) en la reacción primaria; y escisión FRET y generación de la 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 aleta se une de forma suelta al amplicón. La presencia del oligonucleótido invasivo específico en el sitio de unión a la diana produce la escisión para liberar la secuencia de aleta cortando entre la sonda de detección y la secuencia de aleta. La secuencia de aleta es complementaria a una porción no de horquilla del correspondiente casete FRET. En consecuencia, la secuencia de aleta funciona como un oligonucleótido invasivo del casete FRET y realiza una escisión entre el fluoróforo del casete FRET y un inactivador, que produce una señal de fluorescencia. La reacción de escisión puede cortar múltiples sondas por diana y liberar, por tanto, múltiples fluoróforos por aleta, proporciona una amplificación de señal exponencial. QuARTS puede detectar múltiples dianas en un único pocillo de reacción usando casetes 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; solicitud de patente de Estados Unidos con números de serie 12/946.737, 12/946.745, 12/946.752, y 61/548.639.

La expresión "reactivo de bisulfito" se refiere a un reactivo que comprende bisulfito, disulfito, hidrogenosulfito, o combinaciones de los mismos, útiles como se desvela en el presente documento para distinguir entre secuencias de dinucleótido CpG metiladas y no metiladas. Los procedimientos de dicho tratamiento se conocen en la técnica (por ejemplo, documento PCT/EP2004/011715). Se prefiere que el tratamiento con bisulfito se realice en presencia de disolventes desnaturizantes tales como, aunque no de forma limitativa, n-alquilenglicol o dietilenglicol dimetil éter (DME), o en presencia de dioxano o derivados de dioxano. En algunas realizaciones, los disolventes desnaturizantes se utilizan en concentraciones comprendidas entre 1 % y el 35 % (v/v). En algunas realizaciones, la reacción con bisulfito se lleva a cabo en presencia de sequestrantes tales como, aunque no de forma limitativa, derivados de cromano, por ejemplo, 6-hidroxi-2,5,7,8,-tetrametilcromano ácido 2-carboxílico o trihidroxibenzona ácido y derivados de los mismos, por ejemplo, ácido gálico (véase: documento PCT/EP2004/011715). La conversión con bisulfito se lleva a cabo preferentemente a una temperatura de reacción comprendida entre 30 °C y 70 °C, por lo cual, la temperatura se aumenta a más de 85 °C durante cortos periodos de tiempo durante la reacción (véase: documento PCT/EP2004/011715). El ADN tratado con bisulfito preferentemente se purifica antes de la cuantificación. Esto se puede llevar a cabo mediante cualquier medio conocido en la técnica, tales como, aunque no de forma limitativa, ultrafiltración, por ejemplo, mediante columnas Microcon™ (fabricadas por Millipore™). La purificación se lleva a cabo de acuerdo con un protocolo del fabricante modificado (véase, por ejemplo, documento PCT/EP2004/011715).

En algunas realizaciones, los fragmentos del ADN tratado se amplifican usando conjuntos de oligonucleótidos cebadores de acuerdo con la presente invención (por ejemplo, véase la Tabla 2) y una enzima de amplificación. La amplificación de varios segmentos de ADN se puede llevar a cabo simultáneamente en un mismo recipiente de reacción. Normalmente, la amplificación se lleva a cabo usando una reacción en cadena de la polimerasa (PCR). Los amplicones tienen de forma típica de 100 a 2000 pares de bases de longitud.

En otra realización del procedimiento, el estado de metilación de las posiciones CpG dentro o cerca de un marcador que comprenden una DMR (por ejemplo, DMR 1-107 como se proporciona en la Tabla 1) (por ejemplo, DMR 1-449 como se proporciona en la Tabla 10) se puede detectar mediante el uso de oligonucleótidos cebadores específicos de la metilación. Esta técnica (MSP) se ha descrito en la patente de Estados Unidos n.º 6.265.171 de Herman. El uso de cebadores específicos del estado de la metilación para la amplificación del ADN tratado con bisulfito permite la diferenciación entre los ácidos nucleicos metilados y no metilados. Los pares de cebadores MSP contienen al menos un cebador que se hibrida con un dinucleótido CpG tratado con bisulfito. Por lo tanto, la secuencia de dichos cebadores comprende al menos un dinucleótido CpG. Los cebadores MSP específicos del ADN no metilado contienen una "T" en la posición de la posición de C en el CpG.

Los fragmentos obtenidos mediante la amplificación pueden llevar una marca detectable directa o indirectamente. En algunas realizaciones, las marcas son marcadores fluorescentes, radionucleidos, o fragmentos de moléculas que se pueden desprender que tienen una masa típica que se puede detectar en un espectrómetro de masas. Cuando dichas marcas son marcas de masa, algunas realizaciones proporcionan que los amplicones marcados tengan una sola carga neta, positiva o negativa, lo que permite una mejor capacidad de detección en el espectrómetro de masas. La detección se puede llevar a cabo y visualizarse mediante, por ejemplo, espectrometría de masas mediante desorción/ionización láser asistida por matriz (MALDI) o mediante el uso de espectrometría de masas con electropulverización (IEN).

Los procedimientos para aislar ADN adecuados para su uso en estas tecnologías de ensayo son conocidos en la técnica. En particular, algunas realizaciones comprenden aislamiento de los ácidos nucleicos como se describe en la solicitud de patente de Estados Unidos con número de serie 13/470,251 ("Isolation of Nucleic Acids").

60 Procedimientos

En algunas realizaciones de la tecnología, se proporcionan procedimientos que comprenden las siguientes etapas:

- 5 1) poner en contacto un ácido nucleico (por ejemplo, ADN genómico, por ejemplo, aislado de fluidos corporales tales como una muestra de heces o tejido pancreático) obtenido del sujeto con al menos un reactivo o una serie de reactivos que distinguen entre dinucleótidos CpG metilados y no metilados dentro de un marcador que comprende una DMR (por ejemplo, DMR 1-107, por ejemplo, como se proporciona en la Tabla 1) (por ejemplo, DMR 1-449, por ejemplo, como se proporciona en la Tabla 10) y
- 2) detectar una neoplasia o trastorno proliferativo (por ejemplo, conseguido con una sensibilidad mayor o igual al 80 %, y una especificidad mayor o igual al 80 %).

En algunas realizaciones de la tecnología, se proporcionan procedimientos que comprenden las siguientes etapas:

- 10 1) poner en contacto un ácido nucleico (por ejemplo, ADN genómico, por ejemplo, aislado de fluidos corporales tales como una muestra de heces o tejido pancreático) obtenido del sujeto con al menos un reactivo o una serie de reactivos que distinguen entre dinucleótidos CpG metilados y no metilados dentro de al menos un marcador seleccionado entre una región cromosómica que tiene una anotación seleccionada entre el grupo que consiste de ABCB1, ADCY1, BHLHE23 (LOC63930), c13orf18, CACNA1C, chr12 133, CLEC11A, ELMO1, EOMES,
- 15 CLEC 11, SHH, GJC1, IHIF1, IKZF1, KCNK12, KCNN2, PCBP3, PRKCB, RSPO3, SCARF2, SLC38A3, ST8SIA1, TWIST1, VWC2, WT1, y ZNF71, y
- 2) detectar cáncer pancreático (por ejemplo, conseguido con una sensibilidad mayor o igual al 80 %, y una especificidad mayor o igual al 80 %).

En algunas realizaciones de la tecnología, se proporcionan procedimientos que comprenden las siguientes etapas:

- 20 1) poner en contacto un ácido nucleico (por ejemplo, ADN genómico, por ejemplo, aislado de fluidos corporales tales como una muestra de heces o tejido esofágico) obtenido del sujeto con al menos un reactivo o una serie de reactivos que distinguen entre dinucleótidos CpG metilados y no metilados dentro de al menos un marcador seleccionado entre una región cromosómica que tiene una anotación seleccionada entre el grupo que consiste de NDRG4, SFRP1, BMP3, HPP1, y APC, y
- 25 2) detectar un esófago de Barrett (por ejemplo, conseguido con una sensibilidad mayor o igual al 80 %, y una especificidad mayor o igual al 80 %).

En algunas realizaciones de la tecnología, se proporcionan procedimientos que comprenden las siguientes etapas:

- 30 1) poner en contacto un ácido nucleico (por ejemplo, ADN genómico, por ejemplo, aislado de fluidos corporales tales como una muestra de heces o tejido pancreático) obtenido del sujeto con al menos un reactivo o una serie de reactivos que distinguen entre dinucleótidos CpG metilados y no metilados dentro de al menos un marcador seleccionado entre una región cromosómica que tiene una anotación seleccionada entre el grupo que consiste de ADCY1, PRKCB, KCNK12, C13ORF18, IKZF1, TWIST1, ELMO, 55957, CD1D, CLEC11A, KCNN2, BMP3, y NDRG4, y
- 35 2) detectar cáncer pancreático (por ejemplo, conseguido con una sensibilidad mayor o igual al 80 %, y una especificidad mayor o igual al 80 %).

En algunas realizaciones de la tecnología, se proporcionan procedimientos que comprenden las siguientes etapas:

- 40 1) poner en contacto un ácido nucleico (por ejemplo, ADN genómico, por ejemplo, aislado de una muestra de heces) obtenido del sujeto con al menos un reactivo o una serie de reactivos que distinguen entre dinucleótidos CpG metilados y no metilados dentro de una región cromosómica que tiene una CD1D, y
- 2) detectar cáncer pancreático (por ejemplo, conseguido con una sensibilidad mayor o igual al 80 %, y una especificidad mayor o igual al 80 %).

Preferentemente, la sensibilidad es de aproximadamente 70 % a aproximadamente 100 % en peso o de aproximadamente 80 % a aproximadamente 90 % en peso o de aproximadamente 80 % a aproximadamente 85 %. Preferentemente, la especificidad es de aproximadamente 70 % a aproximadamente 100 % en peso o de aproximadamente 80 % a aproximadamente 90 % en peso o de aproximadamente 80 % a aproximadamente 85 %.

El ADN genómico se puede aislar por cualquier medio, incluido el uso de kits comercialmente disponibles. En resumen, cuando el ADN de interés está encapsulado mediante una membrana celular, la muestra biológica debe romperse y lisarse por medios enzimáticos, químicos o mecánicos. La solución de ADN se puede limpiar a continuación de proteínas y otros contaminantes, por ejemplo, mediante digestión con proteinasa K. A continuación, el ADN genómico se recupera de la solución. Esto se puede llevar a cabo por una serie de procedimientos entre los que se incluyen lavado, Extracción con disolventes orgánicos, o unión del ADN a un soporte en fase sólida. La selección del procedimiento estará afectada por varios factores que incluyen el tiempo, coste, y cantidad de ADN necesaria. Todos los tipos de muestras clínicas que comprenden material neoplásico o preneoplásico son adecuados para su uso en el presente procedimiento, por ejemplo, líneas de células, portas histológicas, biopsias, tejido incluido en parafina, fluidos corporales, heces, efluente colónico, 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 procedimientos utilizados para preparar las muestras y proporcionar un ácido nucleico para su análisis. Por ejemplo, en algunas realizaciones, un ADN se aísla de una muestra de heces o de la sangre o del plasma usando captura génica directa, por ejemplo, como se detalla en la solicitud de patente de EE.UU. con número de serie 61/485386 o por un procedimiento relacionado.

5 La muestra de ADN genómico se trata a continuación con al menos un reactivo, o serie de reactivos, que distinguen entre dinucleótidos CpG metilados y no metilados dentro de al menos un marcador que comprende una DMR (por ejemplo, DMR 1-107, por ejemplo, como se proporciona en la Tabla 1) (por ejemplo, DMR 1-449, por ejemplo, como se proporciona en la Tabla 10).

10 En algunas realizaciones, el reactivo convierte las bases de citosina que no están metiladas en la posición 5' en uracilo, timina, u otra base, que sea 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.

15 En algunas realizaciones, la muestra de ADN genómico se trata de tal forma 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 (hidrogenosulfito, disulfito) seguido por hidrólisis alcalina.

20 El ácido nucleico tratado se analiza a continuación 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 DMR, por ejemplo, al menos una DMR seleccionada entre DMR 1-107, por ejemplo, como se proporciona en la Tabla 1) (al menos un gen, secuencia genómica, o nucleótido de un marcador que comprende una DMR, por ejemplo, al menos una DMR seleccionada entre DMR 1-449, por ejemplo, como se proporciona en la Tabla 10). El procedimiento de análisis puede seleccionarse entre los conocidos en la técnica, incluidos los relacionados en el presente documento, por ejemplo, QuARTS y MSP como se describe en el presente documento.

25 Una metilación anómala, más específicamente la hipermetilación de un marcador que comprende una DMR (por ejemplo, DMR 1-107, por ejemplo, como se proporciona en la Tabla 1) (por ejemplo, DMR 1-449, por ejemplo, como se proporciona en la Tabla 10) está asociada con un cáncer y, en algunas realizaciones, predice el sitio del tumor.

30 La tecnología se refiere al análisis de cualquier muestra asociada a un cáncer del sistema gastrointestinal. Por ejemplo, en algunas realizaciones, la muestra comprende un tejido y/o fluido biológico obtenida de un paciente. En algunas realizaciones, la muestra comprende una secreción. En algunas realizaciones, la muestra comprende sangre, suero, plasma, secreciones gástricas, jugo pancreático, una muestra de biopsia gastrointestinal, células microdisecadas de una biopsia gastrointestinal, células gastrointestinales enviadas al interior de la luz gastrointestinal, y/o células gastrointestinales recuperadas de las heces. En algunas realizaciones, el sujeto es un ser humano. Estas muestras pueden tener su origen en el tracto gastrointestinal superior, el tracto gastrointestinal inferior, o comprender células, tejidos y/o secreciones tanto del tracto gastrointestinal superior como del tracto gastrointestinal inferior. La muestra puede incluir células, secreciones, o tejidos del hígado, conductos biliares, páncreas, estómago, colon, recto, esófago, intestino delgado, apéndice, duodeno, pólipos, vesícula biliar, ano, y/o peritoneo. En algunas realizaciones, la muestra comprende fluido celular, ascitis, orina, heces, fluido pancreático, fluido obtenido durante la endoscopia, sangre, moco, o saliva. En algunas realizaciones, la muestra es una muestra de sangre.

35 Dichas muestras se pueden obtener por cualquier número de procedimientos conocidos en la materia, tal como será evidente para el experto en la materia. Por ejemplo, las muestras de orina y fecales son fácilmente accesibles, mientras que las muestras de sangre, ascites, suero o fluido pancreático se pueden obtener por vía parenteral usando una aguja y jeringa, por ejemplo. Las muestras exentas o sustancialmente exentas de células se pueden obtener sometiendo la muestra a diferentes técnicas conocidas del experto en la materia y que incluyen, aunque no de forma limitativa, centrifugación y filtración. Aunque por lo general se prefiere no usar técnicas invasivas para obtener la muestra, puede seguir siendo preferible obtener muestras tales como homogenados de tejido, secciones de tejido y especímenes de biopsias.

40 En algunas realizaciones, la tecnología se refiere a un procedimiento para tratar un paciente, (por ejemplo, un paciente con cáncer gastrointestinal, con un cáncer gastrointestinal en las primeras etapas, o que puede desarrollar cáncer gastrointestinal), comprendiendo el procedimiento determinar el estado de metilación de una o más DMR como se proporciona en el presente documento y administrar un tratamiento al paciente en función de los resultados de la determinación del estado de metilación. El tratamiento puede ser la administración de un compuesto farmacéutico, una vacuna, realizar una intervención quirúrgica, realizar una adquisición de imagen del paciente, realizar otra prueba. Preferentemente, dicho uso es en un procedimiento de cribado clínico, un procedimiento de evaluación del pronóstico, un procedimiento para seguimiento de los resultados de una terapia, un procedimiento para identificar los pacientes con más probabilidad de responder a un determinado tratamiento terapéutico, un procedimiento para obtener imágenes de un paciente o sujeto, y un procedimiento para el cribado y desarrollo de fármacos.

En algunas realizaciones de la tecnología, se proporciona un procedimiento para diagnosticar un cáncer

gastrointestinal en un sujeto. Los términos "diagnosticar" y "diagnóstico" tal como se usa en el presente documento se refieren a procedimientos mediante los que el experto en la técnica puede estimar e incluso determinar si un paciente padece o no una enfermedad o dolencia dada o puede desarrollar una enfermedad o dolencia dada en el futuro. El experto en la materia frecuentemente realiza un diagnóstico sobre la base de uno o más indicadores diagnósticos, tales como, por ejemplo, un biomarcador (por ejemplo, una DMR como se desvela en el presente documento), cuyo estado de metilación es indicativo de la presencia, gravedad, o ausencia de la dolencia.

Junto con el diagnóstico, el pronóstico clínico del cáncer se relaciona con determinar la agresividad del cáncer y la probabilidad de recidiva del tumor para planificar la terapia más eficaz. Si se puede hacer un pronóstico más preciso o incluso se puede evaluar el riesgo potencial de desarrollar cáncer, se puede seleccionar la terapia adecuada y, en algunos casos, una terapia menos severa para el paciente. La valoración (por ejemplo, la determinación del estado de metilación) de los biomarcadores del cáncer es de utilidad para separar sujetos con buen pronóstico y/o bajo riesgo de desarrollar cáncer que no necesitan terapia o bien una terapia limitada de aquellos con mayor probabilidad de desarrollar cáncer o de tener recidiva de un cáncer que se beneficiarían de un tratamiento más intensivo.

Por tanto, "realizar un diagnóstico", o "diagnosticar", como se usa en el presente documento, es inclusivo además de realizar una determinación del riesgo de desarrollar cáncer o determinar un pronóstico, que puede proporcionar la predicción de un resultado clínico (con o sin tratamiento médico), seleccionar un tratamiento adecuado (o si un tratamiento sería eficaz), o realizar un seguimiento de un tratamiento actual y potencialmente cambiar el tratamiento, en función de la medición de los biomarcadores de diagnóstico (por ejemplo, DMR) desvelados en el presente documento. Además, en algunas realizaciones de la presente materia sujeto, se puede realizar una determinación de múltiples biomarcadores con el tiempo para facilitar el diagnóstico y/o el pronóstico. Se puede usar un cambio temporal en el biomarcador para predecir un resultado clínico, seguir la evolución del cáncer gastrointestinal y/o seguir la eficacia de las terapias adecuadas dirigidas contra el cáncer. En una realización de ese tipo, por ejemplo, se podría esperar ver un cambio en el estado de metilación de uno o más biomarcadores (por ejemplo, DMR) desvelados en el presente documento (y potencialmente en uno o más biomarcadores adicionales, en caso de realizar seguimiento) en una muestra biológica a lo largo del tiempo durante el curso de una terapia eficaz.

La materia sujeto actualmente desvelada proporciona además, en algunas realizaciones, un procedimiento para determinar si iniciar o continuar la profilaxia o el tratamiento del cáncer en un sujeto. En algunas realizaciones, el procedimiento comprende proporcionar una serie de muestras biológicas durante un periodo de tiempo precedentes del sujeto; analizar la serie de muestras biológicas para determinar el estado de metilación de al menos un biomarcador desvelado en el presente documento para cada una de las muestras biológicas; y comparar cualquier cambio mensurable en los estados de metilación de uno o más de los biomarcadores en cada una de las muestras biológicas. Los posibles cambios en los estados de metilación de los marcadores a lo largo del periodo de tiempo se pueden usar para predecir el riesgo de desarrollar cáncer, predecir el resultado clínico, determinar si iniciar o continuar la profilaxia o la terapia contra el cáncer, y si una terapia actual es eficaz para tratar el cáncer. Por ejemplo, se puede seleccionar un primer punto antes de iniciar el tratamiento y se puede seleccionar un segundo punto en algún momento después de iniciar el tratamiento. Los estados de metilación se pueden medir en cada una de las muestras tomadas de diferentes puntos temporales y anotar las diferencias cuantitativas y/o cualitativas observadas. Un cambio en los estados de metilación de los niveles de biomarcadores de las diferentes muestras se puede correlacionar con el riesgo de cáncer gastrointestinal, pronóstico, determinación de la eficacia del tratamiento, y/o evolución del cáncer en el sujeto.

En realizaciones preferidas, los procedimientos y composiciones de la invención son para el tratamiento o diagnóstico de la enfermedad en una etapa inicial, por ejemplo, antes de que se presenten los síntomas de la enfermedad. En algunas realizaciones, los procedimientos y composiciones de la invención son para el tratamiento o diagnóstico de la enfermedad en una etapa clínica.

Como se ha indicado, en algunas realizaciones, se pueden realizar múltiples determinaciones de uno o varios biomarcadores de diagnóstico o pronóstico, y se puede usar un cambio temporal en el marcador para determinar el diagnóstico o pronóstico. Por ejemplo, un marcador diagnóstico se puede determinar en un momento inicial, y de nuevo en un segundo momento. En dichas realizaciones, un aumento en el valor del marcador del primer al segundo momento puede ser diagnóstico de un tipo particular o de la gravedad de un cáncer, o un pronóstico dado. Análogamente, una disminución en el valor del marcador del primer al segundo momento puede ser diagnóstico de un tipo particular o de la gravedad del cáncer, o un pronóstico dado. Además, el grado de cambio en uno o más marcadores se puede relacionar con la gravedad del cáncer y futuros acontecimientos adversos. El experto en la materia entenderá que, aunque en algunas realizaciones se pueden hacer mediciones comparativas del mismo biomarcador en múltiples puntos temporales, también se puede medir un biomarcador dado en un punto temporal, y un segunda biomarcador en un segundo punto temporal, y una comparación entre ambos marcadores puede proporcionar información para el diagnóstico.

Como se usa en el presente documento, la expresión "determinación del pronóstico" se refiere a los procedimientos en los que el experto en la técnica puede predecir la evolución o el resultado de una dolencia en un sujeto. El término "pronóstico" no se refiere a la capacidad de predecir la evolución o el resultado de una dolencia con el 100 % de precisión, o incluso que una evolución o resultado dado sea predeciblemente más o menos probable que

ocurra en función del estado de metilación de un biomarcador (por ejemplo, una DMR). En cambio, el experto en la materia entenderá que el término "pronóstico" se refiere a una mayor probabilidad de que se produzca una determinada evolución o resultado; es decir, que es más probable que se produzca una evolución o resultado en un sujeto que presenta una dolencia dada, en comparación con otros individuos que no padecen la dolencia. Por ejemplo, en individuos que no presentan la dolencia (por ejemplo, que tienen un estado de metilación normal de una o más DMR), al posibilidad de un resultado dado (por ejemplo, padecer un cáncer gastrointestinal) puede ser muy baja.

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 que aparece en una muestra de control normal obtenida de un paciente que no tiene cáncer puede señalar que es más probable que un sujeto padezca un cáncer que los sujetos con un nivel que es más similar al el estado de metilación de la muestra del control, como se determina por el nivel de significación estadística. Además, un cambio en el estado de metilación desde el valor inicial (por ejemplo, "normal") puede ser reflejo del pronóstico de un sujeto, y el grado de cambio en el estado de metilación se puede relacionar con la gravedad de los acontecimientos adversos. La significación estadística frecuentemente se determina por comparación de dos o más poblaciones y determinando un intervalo de confianza y/o un valor p. Véase, por ejemplo, Dowdy y Wearden, *Statistics for Research*, John Wiley & Sons, Nueva York, 1983. Los intervalos de confianza ilustrativos de la presente materia sujeto son 90 %, 95 %, 97,5 %, 98 %, 99 %, 99,5 %, 99,9 % y 99,99 %, mientras que los valores de p ilustrativos son 0,1, 0,05, 0,025, 0,02, 0,01, 0,005, 0,001, y 0,0001.

En otras realizaciones, puede establecerse un grado de cambio umbral en el estado de metilación de un marcador de pronóstico o diagnóstico descrito en el presente documento (por ejemplo, una DMR), y el grado de cambio en el estado de metilación del biomarcador en la muestra biológica es fácil de comparar con el grado de cambio umbral en el estado de metilación. Un umbral de cambio preferido en el estado de metilación de los biomarcadores proporcionados en el presente documento es de aproximadamente un 5 %, aproximadamente un 10 %, aproximadamente un 15 %, aproximadamente un 20 %, aproximadamente un 25 %, aproximadamente un 30 %, aproximadamente un 50 %, aproximadamente un 75 %, aproximadamente un 100 %, y aproximadamente un 150 %. En otras realizaciones más, se puede establecer un "nomograma", en el que el estado de metilación de un indicador de pronóstico o diagnóstico (biomarcador o combinación de biomarcadores) se relaciona directamente con una disposición asociada hacia un determinado resultado. El experto en la materia está familiarizado con el uso de estos nomogramas para relacionar dos valores numéricos con la comprensión de que la incertidumbre de esta medición es la misma que la incertidumbre en la concentración del marcador porque se hace referencia a mediciones individuales de la muestra, no medias de población.

En algunas realizaciones, una muestra del control se analiza simultáneamente con la muestra biológica, de forma que los resultados obtenidos de la muestra biológica se pueden comparar con los resultados obtenidos en la muestras del control. Además, se contempla que se puedan proporcionar curvas patrón, con las que se pueden comparar los resultados de la muestra biológica. Dichas curvas patrón presentan los estados de metilación de un biomarcador en función de unidades de ensayo, por ejemplo, Intensidad de la señal de fluorescencia, si se usa una marca fluorescente. Cuando se usan muestras tomadas de múltiples donantes, las curvas pueden proporcionarse para controlar los estados de metilación del uno o más biomarcadores en tejido normal, así como para niveles "de riesgo" de uno o más biomarcadores del tejido tomado de donantes o metaplasia o de donantes con un cáncer gastrointestinal. En algunas realizaciones del procedimiento, se identifica que un sujeto tiene metaplasia tras identificar un estado de metilación anómalo en una o más DMR proporcionadas en el presente documento en una muestra biológica procedente de un sujeto. En otras realizaciones del procedimiento, la detección de un estado de metilación anómalo de uno o más de dichos biomarcadores en una muestra biológica procedente del sujeto da como resultado que el sujeto se identifica por tener cáncer.

El análisis de los marcadores se puede llevar a cabo de forma separada o simultánea con marcadores adicionales dentro de una muestra de ensayo. Por ejemplo, varios marcadores se pueden combinar en una prueba para el procesamiento eficaz de una multiplicidad de muestras y para proporcionar potencialmente mayor eficacia diagnóstica y/o pronóstica. Además, el experto en la técnica reconocerá el valor de someter a ensayo múltiples muestras (por ejemplo, en sucesivos puntos temporales) procedentes del mismo sujeto. Dicho análisis de muestras en serie puede permitir la identificación de los cambios en los estados de metilación del marcador con el tiempo. Los cambios en el estado de metilación, así como la ausencia de cambio en el estado de metilación, puede proporcionar información útil acerca del estado de la enfermedad que incluye, aunque no de forma limitativa, identificar el tiempo transcurrido aproximado desde el inicio del evento, la presencia y cantidad de tejido salvable, las adecuación de una terapias farmacológica, la eficacia de diferentes terapias, y la identificación del resultado del sujeto, incluido el riesgo de futuros eventos.

El análisis de los biomarcadores se puede llevarse a cabo en una gran variedad de formatos físicos. Por ejemplo, el uso de placas de microtitulación o se puede usar automatización para facilitar el procesamiento de grandes cantidades de muestras de ensayo. Como alternativa, se podrían desarrollar formatos de una sola muestra para facilitar el tratamiento y diagnóstico inmediato de una forma rápida, por ejemplo, durante el transporte en ambulancia o en unidades de atención de emergencia.

En algunas realizaciones, se diagnostica al sujeto que tiene un cáncer gastrointestinal si, cuando se compara con un estado de metilación del control, existe una diferencia mensurable en el estado de metilación de al menos un biomarcador de la muestra. Por el contrario, cuando no se identifica ningún cambio en el estado de metilación de la muestra biológica, se puede identificar que el sujeto no tiene cáncer gastrointestinal, ni está en riesgo de cáncer, o que tiene un riesgo de cáncer bajo. En este sentido, los sujetos que tienen cáncer o riesgo del mismo se pueden diferenciar de sujetos que tienen bajo a sustancialmente nada de cáncer o riesgo del mismo. Dichos sujetos que tienen riesgo de desarrollar un cáncer gastrointestinal se pueden llevar a un calendario de cribado más intensivo y/o regular, incluida la vigilancia endoscópica. Por otro lado, los sujetos que tienen de bajo a prácticamente ningún riesgo pueden evitar someterse a una endoscopia, hasta el momento en que un futuro cribado, por ejemplo, por ejemplo, un cribado realizado de acuerdo con la presente tecnología, indique que ha aparecido un riesgo de cáncer gastrointestinal en estos sujetos.

Tal como se ha mencionado anteriormente, dependiendo de la realización del procedimiento de la presente tecnología, la detección del cambio en el estado de metilación del uno o más biomarcadores puede ser una determinación cualitativa o puede ser una determinación cuantitativa. Por tanto, la etapa de diagnosticar un sujeto que tiene, o está en riesgo de desarrollar, un cáncer gastrointestinal indica que se han realizado determinadas mediciones umbral, por ejemplo, el estado de metilación del uno o más biomarcadores de la muestra biológica varía desde un estado de metilación del control predeterminado. En algunas realizaciones del procedimiento, el estado de metilación del control es cualquier estado de metilación detectable en el biomarcador. En otras realizaciones del procedimiento donde una muestra del control se analiza en paralelo con la muestra biológica, el estado de metilación predeterminado es el estado de metilación en la muestra del control. En otras realizaciones del procedimiento, el estado de metilación predeterminado se basa en y/o está identificado por una curva patrón. En otras realizaciones del procedimiento, el estado de metilación predeterminado es un estado específico o un intervalo de estado. Por tanto, el estado de metilación predeterminado se puede seleccionar, entre los límites aceptables que serán evidentes para los expertos en la materia, basándose en parte en la realización del procedimiento que se lleva a la práctica y la especificidad deseada, etc.

Adicionalmente con respecto a los procedimientos de diagnóstico, un sujeto preferido es un sujeto vertebrado. Un vertebrado preferido es de sangre caliente; un vertebrado de sangre caliente preferido es un mamífero. Un mamífero preferido es más preferentemente un ser humano. Como se usa en el presente documento, el término "sujeto" incluye sujetos tanto humanos como animales. Por lo tanto, se proporcionan en el presente documento usos terapéuticos veterinarios. Por tanto, la presente tecnología proporciona el diagnóstico de mamíferos tales como seres humanos, así como aquellos mamíferos de importancia porque están en peligro de extinción, tales como los tigres de Siberia; de importancia económica, como los animales criados en granjas para consumo humano; y/o animales de importancia social para los seres humanos, como los animales que se tienen como mascotas o en los zoológicos. Los ejemplos de dichos animales incluyen, aunque no de forma limitativa: carnívoros como gatos y perros; suínidos, incluidos cerdos, lechones y jabalíes; rumiantes y/o ungulados tales como ganado vacuno, bueyes, ovejas, jirafas, ciervos, cabras, bisontes, y camellos; y caballos. Por lo tanto, también se proporciona el diagnóstico y el tratamiento de los animales de granja, incluidos, aunque no de forma limitativa, suínidos domesticados, rumiantes, ungulados, caballos (incluidos los caballos de carreras), y similares. La materia sujeto actualmente desvelada incluye además un sistema para diagnosticar un cáncer gastrointestinal en un sujeto. El sistema se puede proporcionar, por ejemplo, en forma de un kit comercial que se puede utilizar para cribar un riesgo de cáncer gastrointestinal o diagnosticar un cáncer gastrointestinal en un sujeto del que se ha tomado una muestra biológica. Un sistema ilustrativo proporcionado de acuerdo con la presente tecnología incluye evaluar el estado de metilación de una DMR como se proporciona en la Tabla 1.

Ejemplos

Ejemplo 1 - Identificación de marcadores usando RRBS

En su conjunto, los cánceres gastrointestinales representan más muertes que los de cualquier otro sistema orgánico, y la incidencia agregada del cáncer gastrointestinal superior y el del cáncer colorrectal (CCR) son comparables. Para maximizar la eficacia del cribado y el diagnóstico, se necesitan marcadores moleculares del cáncer gastrointestinal que sean específicos del sitio cuando se analizan en medios distantes como la sangre o heces. Aunque ampliamente informativos, los marcadores de ácidos nucleicos metilados de forma anómala son bastante habituales en los cánceres gastrointestinales superiores y en el CRC.

Durante el desarrollo de la tecnología proporcionada en el presente documento, se recogieron datos de un estudio de control del caso para demostrar que una estrategia de búsqueda en todo el genoma identifica novedosos marcadores candidatos e informativos. Los experimentos preliminares demostraron que el análisis de heces para un marcador génico metilado (BMP3) detecta PanC. Después, se ha comprobado que un ensayo combinado de BMP3 metilado y KRAS mutante KRAS aumentó la detección sobre cualquiera de los marcadores en solitario. Sin embargo, los marcadores discriminantes en tejidos demostraron ser malos marcadores en heces debido a un elevado fondo de metilación, por ejemplo, como se detectó en los especímenes del control.

Población del estudio, adquisición de especímenes, y muestras

La población diana eran pacientes con cáncer de páncreas tratados en la clínica Mayo. La población accesible incluye los que se han sometido a una pancreatectomía distal, una pancreatoduodenectomía, o una colectomía con un espécimen de resección archivado, y un diagnóstico patológico confirmado. El ADN del epitelio colónico se había extraído previamente de especímenes microdisecados en el laboratorio de Biospecimens Accessioning Processing (BAP) usando un protocolo de fenol-cloroformo. El personal de Pancreas SPORE usó los datos de las variables emparejadas de estas muestras para seleccionar las muestras de registro de tejidos. Estas fueron revisadas por un patólogo experto para confirmar el caso y controlar el estado y excluir casos de neoplasias producidas a partir de IPMN, que puede tener una biología subyacente distinta. El personal de SPORE dispuso la microdisección en el laboratorio BAP y la extracción del ADN en los casos pancreáticos y muestras del control y proporcionó 500 ng de ADN al personal del laboratorio que estaba enmascarado para el caso y el estado del control. Las muestras de ácido nucleico de archivo incluyeron 18 adenocarcinomas pancreático, 18 páncreas normales y 18 epitelios colónicos emparejados por sexo, edad, y estado de tabaquismo.

Los tipos de muestras fueron:

- 1) tejidos de PanC del registro de Pancreas SPORE en la clínica Mayo limitado a AJCC estado I y II;
- 2) páncreas de control sin PanC;
- 3) epitelio colónico del control archivado sin PanC; y
- 4) neoplasia colónico del que el ADN se había extraído y almacenado en el laboratorio BAP.

Los casos y los controles se emparejaron por sexo, edad (en incrementos de 5 años), y estado de tabaquismo (actual o anterior frente a nunca).

20 Variables principales

La variable principal fue el porcentaje de metilación de cada amplicón de 101 pares de bases individual de las regiones HCP. El porcentaje de metilación en las muestras de casos se comparó con las muestras del control después de RRBS.

Procedimientos

Se prepararon bibliotecas según procedimientos descritos previamente (véase, por ejemplo, Gu y col (2011) "Preparation of reduced representation bisulfite sequencing libraries for genome-scale DNA methylation profiling" Nature Protocols 6: 468-81) mediante la fragmentación de ADN genómico (300 ng) por digestión con 10 unidades de *MspI*, una enzima de restricción específica de metilación que reconoce los motivos que contienen CpG. Este tratamiento enriquece las muestras en su contenido de CpG y elimina zonas redundantes del genoma. Se repararon los extremos de los fragmentos digeridos y se agregaron colas A con 5 unidades de fragmento de Klenow (3'-5' exo) y se ligó durante la noche a adaptadores Illumina que contienen una de cuatro secuencias de código de barras para enlazar cada fragmento a su ID de muestra. Se llevó a cabo la selección por tamaño de los fragmentos de 160-340 pb (que tiene inserciones de 40-220 pb) usando perlas /tampón SPRI (AMPure XP, Beckman Coulter). Los cortes de tampón fueron 0,7x a 1,1 x del volumen de muestra de perlas/tampón. Las muestras se eluyeron en un volumen de 22 µl (tampón EB, Qiagen). Se utilizó la PCR para controlar la eficacia de enlazado y la calidad de fragmentos en una pequeña alícuota de la muestra. A continuación, las muestras se sometieron a dos rondas de conversión con bisulfito usando un protocolo EpiTect modificado (Qiagen). qPCR y PCR convencional (Pfu Turbo Cx hotstart, Agilent), seguido por evaluación con Bioanalyzer 2100 (Agilent) sobre alícuotas de muestra convertidas, determinaron el número óptimo de ciclos de la PCR antes de la amplificación de la biblioteca final. La PCR final se llevó a cabo en un volumen de 50 µl (5 µl de 10x tampón PCR; 1,25 µl de cada dNTP a 10 mM; 5 µl de un primer cóctel a aproximadamente 5 µM, 15 µl de molde (muestra), 1 µl de PfuTurbo Cx hotstart, y 22,75 µl de agua. El ciclado térmico comenzó con incubaciones iniciales a 95 °C durante 5 minutos y a 98 °C durante 30 segundos seguido de 16 ciclos de 98 °C durante 10 segundos, 65 °C durante 30 segundos, y a 72 °C durante 30 segundos. Tras la ciclación, las muestras se incubaron a 72 °C durante 5 minutos y se mantuvieron a 4 °C hasta la posterior elaboración y análisis. Las muestras se combinaron en cantidades equimolares en bibliotecas tetramultiplexadas según un esquema de aleatorización y se sometieron a ensayo con el bioanalizador para una comprobación de la etapa final. Las muestras también se analizaron con qPCR usando patrones phiX y cebadores específicos del adaptador.

Para la secuenciación, las muestras se cargaron en hileras de celdas de flujo de acuerdo con una asignación aleatorizada de hileras, donde hileras adicionales están reservadas para los controles internos del ensayo. La secuenciación se realizó mediante el NGS Core en la Medical Genome Facility de Mayo con el Illumina HiSeq 2000. Las lecturas fueron unidireccionales para 101 ciclos. Cada hilera de celda de flujo generó 100-120 millones de lecturas, suficiente para una mediana de cobertura de 30x a 50x de profundidad de secuenciación (basándose en el número de lecturas por CpG) para las secuencias alineadas. Se usó el programa informático convencional Illumina para analizar las lecturas junto con RRBSMAP (Xi, y col. (2012) "RRBSMAP: a fast, accurate and user-friendly alignment tool for reduced representation bisulfite sequencing" Bioinformatics 28: 430-432) y un código informático propio (SAAP-RRBS) desarrollado por el personal de Mayo Biomedical and Statistics (Sun y col. (2012) "SAAP-RRBS: streamlined analysis and annotation pipeline for reduced representation bisulfite sequencing" Bioinformatics 28: 2180-1). El análisis bioinformático consistió en 1) evaluación y limpieza de la lectura de la secuencia, 2)

alineamiento con el genoma de referencia, 3) extracción del estado de metilación, y 4) notificación y anotación de CpG.

Consideraciones estadísticas

5 La comparación primaria evaluó las diferencias en la metilación entre los casos y los controles pancreáticos para cada CpG y/o ventana de CpG apilada. La comparación secundaria evaluó las diferencias en la metilación entre los casos y los controles del colon. Se estudió la metilación diferencial de los marcadores mediante:

1. Evaluación de las distribuciones del porcentaje de metilación para cada marcador, y descartar los marcadores que estaban metilados en más del 1 % en el 10 % de los controles;
2. Ensayar la distribución de la metilación de los marcadores restantes entre casos y controles mediante la prueba de suma de rangos de Wilcoxon y ordenando los marcadores por los valores p; y
3. Usar los valores Q para estimar tasas de descubrimiento falsas (FDR) (Benjamini y col. (1995) "Multiple Testing" Journal of the Royal Statistical Society. Serie B (Metodológica) 57: 289-300; Storey y col. (2003) "Statistical significance for genomewide studies" Proc Natl Acad Sci U S A 100: 9440-5). Durante el descubrimiento, es aceptable una FDR de hasta el 25 %.

15 Análisis de datos

Se desarrolló un código de análisis de datos en el paquete informático de análisis de software R ("R: A Language and Environment for Statistical Computing" (2012), R Foundation for Statistical Computing). El flujo de trabajo comprendía las siguientes etapas:

1. Lectura de los 6.101.049 sitios CpG
- 20 2. Identificar para análisis adicional solo aquellos sitios CpG donde la profundidad de cobertura total del grupo es de 200 lecturas o más. Este corte se basó en una evaluación de potencia para detectar una diferencia entre el 20 % y el 30 % en la metilación entre dos grupos cualesquiera debido a que cualquier cosa menor de ese intervalo tiene pocas posibilidades de significación. La profundidad de cobertura del grupo mide el número de lecturas para todos los sujetos de un grupo (por ejemplo, si hay 18 sujetos por cada grupo y cada sujeto tiene 12 lecturas, entonces la profundidad de cobertura del grupo es $12 \times 18 = 216$).
- 25 3. Estimar la asociación del subtipo de enfermedad con el % de metilación usando regresión de Poisson aumentada por varianza; se determinaron los sitios CpG más discriminantes por comparación del ajuste del modelo χ^2 al percentil 95° de todos los modelos ajustados. Excluir todos los sitios CpG donde la varianza del porcentaje de metilación a través de los grupos sea 0 porque estos sitios son sitios CpG no informativos.

30 Aplicar los filtros 2 y 3 dejó un total de 1.217.523 sitios CpG.

4. Realizar la regresión logística sobre el % de metilación (basándose en el recuento real) usando los grupos definidos como Colon normal, Páncreas normal y Páncreas canceroso. Aunque la variabilidad en el % de metilación entre sujetos es mayor que el permitido por la hipótesis binómica, se usó un modelo de regresión logística superdisperso para tener en cuenta el aumento de la varianza. Este parámetro de dispersión se estimó usando la Chi cuadrada de Pearson del ajuste.
- 35 5. De estos ajustes del modelo, calcular una estadística F global para la comparación del grupo basándose en el cambio de la desviación entre modelos con y sin cada grupo como regresor. Esta desviación se escaló según la dispersión estimada del parámetro.
6. Crear islas CpG en cada cromosoma en función de la distancia entre las ubicaciones de sitios CpG. De forma aproximada, cuando la distancia entre dos ubicaciones CpG supera 100 pb, cada ubicación se define como una isla independiente. Algunas islas fueron singletes y se excluyeron.
- 40 7. A partir de la definición de isla anterior, se calculó una estadística F promedio. Cuando la estadística F supera el 95 % (es decir, 5 % superior) de todos los sitios CpG para el cromosoma concreto, se genera una cifra de resumen.

45 El análisis adicional comparó los siguientes filtros de selección:

1. corte del valor de p ANOVA $< 0,01$
2. Relaciones del % de metilación PanC respecto al páncreas normal y el colon normal > 10
3. % metilación de normales < 2 %
4. Número de CpG contiguos que cumplen los criterios ≥ 3

50 Se evaluó la ventana de metilación para incluir CpG contiguos adicionales que presentaban una metilación significativa. Después, Los candidatos se clasificaron por el nombre del gen y las regiones anotadas y por la ubicación cromosómica de las regiones no anotadas.

Resultados

55 Se cartografiaron aproximadamente 6 millones de CpG con una cobertura $\geq 10x$. Más de 500 islas CpG cumplieron los criterios de significación par la metilación diferencial. Después de aplicar los criterios de filtro anteriores, se

ES 2 712 798 T3

identificaron 107 regiones metiladas de forma diferencial (DMR) (Tabla 1).

Tabla 1: DMR

DMR N.º	anotación del gen	cromosoma	hebra	región en el cromosoma (base inicial-base final)
1	ninguno	1	F	35394805-35394875
2	ninguno	1	F	240161479-240161546
3	ninguno	1	R	156406057-156406118
4	AK055957	12	F	133484978-133485738
5	ninguno	12	R	133484979-133485739
6	APBA2	15	F	29131299-29131369
7	ninguno	2	F	71503632-71503860
8	PCBP3	21	R	47063793-47064177
9	TMEM200A	6	F	130687223-130687729
10	ninguno	9	R	120507311-120507354
11	ABCB1	7	R	87229775-87229856
12	ADAMTS17	15	R	100881373-100881437
13	ADAMTS18	16	R	77468655-77468742
14	ADCY1	7	F	45613877-45614564
15	ADCY1	7	R	45613878-45614572
16	AGFG2	7	F	100136884-100137350
17	ARHGEF7	13	F	111767862-111768355
18	AUTS2	7	R	69062531-69062585
19	BTBD11	12	F	107715014-107715095
20	BVES	6	R	105584524-105584800
21	c13orf18	13	F	46960770-46961464
22	c13orf18	13	R	46960910-46961569
23	CACNA1C	12	F	2800665-2800898
24	CBLN1	16	R	49315846-49315932
25	CBS	21	F	44496031-44496378
26	CBS	21	R	44495926-44496485
27	CD1D	1	F	158150797-158151142
28	CELF2	10	F	11059508-11060151
29	CLEC11A	19	F	51228217-51228703

ES 2 712 798 T3

(continuación)

DMR N.º	anotación del gen	cromosoma	hebra	región en el cromosoma (base inicial-base final)
30	CLEC11A	19	R	51228325-51228732
31	CNR1	6	F	88876367-88876445
32	CNR1	6	R	88875699-88875763
33	CHRH2	7	F	30721941-30722084
34	DBNL	7	F	44084171-44084235
35	DBX1	11	R	20178177-20178304
36	DHRS12	13	F	52378159-52378202
37	DLL1	6	F	170598241-170600366
38	ELMO1	7	F	37487539-37488498
39	ELMO1	7	R	37487540-37488477
40	EN1	2	R	119607676-119607765
41	EOMES	3	F	27763358-27763617
42	FBLN1	22	R	45898798-45898888
43	FEM1C	5	F	114880375-114880442
44	FER1L4	20	R	34189679-34189687
45	FKBP2	11	F	64008415-64008495
46	FLT3	13	F	28674451-28674629
47	FNIP1	5	F	131132146-131132232
48	FOXP2	7	R	113727624-113727693
49	GFRA4	20	R	3641457-3641537
50	GJC1	17	F	42907705-42907798
51	GJC1	17	R	42907752-42907827
52	GRIN2D	19	F	48946755-48946912
53	HECW1	7	R	43152309-43152375
54	HOXA1	7	R	27136030-27136245
55	IFIH1	2	R	163174541-163174659
56	IGF2BP1	17	F	47073394-47073451
57	IKZF1	7	R	50343848-50343927
58	INSM1 (región 1)	20	F	20345123-20345150
59	INSM1 (región 2)	20	F	20350520-20350532
60	KCNK12	2	F	47797332-47797371
61	KCNN2	5	F	113696984-113697057
62	KCTD15	19	R	34287890-34287972
63	LINGO3	19	F	2290471-2290541
64	LOC100126784	11	R	19733958-19734013
65	LOC63930	20	F	61637950-61638000
66	LOC642345	13	R	88323571-88323647
67	MLLT1	19	R	6236992-6237089
68	MPND	19	R	4343896-4242968
69	MYEF2	15	F	48470117-48470606
70	NDUFAB1	16	R	23607524-23607650
71	NFASC	1	F	204797781-204797859
72	NR5A1	9	F	127266951-127267032
73	PDE6B	4	F	657586-657665
74	PLAGL1	6	R	144384503-144385539
75	PRKCB	16	R	23846964-23848004
76	PRRT3	3	F	9988302-9988499
77	PTF1A	10	F	23480864-23480913
78	RASGRF2	5	R	80256215-80256313

(continuación)

DMR N.º	anotación del gen	cromosoma	hebra	región en el cromosoma (base inicial-base final)
79	RIMKLA	1	R	42846119-42846174
80	RNF216	7	F	5821188-5821283
81	RSPO3	6	F	127440526-127441039
82	RSPO3	6	R	127440492-127440609
83	RYBP	3	R	72496092-72496361
84	SCARF2	22	F	20785373-20785464
85	SHH	7	F	155597771-155597951
86	SLC35E3	12	F	69140018-69140202
87	SLC38A3	3	R	50243467-50243553
88	SLC6A3	5	R	1445384-1445473
89	SPSB4	3	F	140770135-140770193
90	SRCIN1	17	R	36762706-36762763
91	ST6GAL2	2	F	107502978-107503055
92	ST6GAL2	2	R	107503155-107503391
93	ST8SIA1	12	F	22487528-22487827
94	ST8SIA1	12	R	22487664-22487848
95	ST8SIA6	10	F	17496177-17496310
96	SUSD5	3	R	33260338-33260423
97	TOX2	20	F	42544666-42544874
98	TWIST1	7	F	19156788-19157093
99	TWIST1	7	R	19156815-19157227
100	USP3	15	R	63795435-63795636
101	USP44	12	R	95942179-95942558
102	VIM	10	F	17271896-17271994
103	VWC2	7	R	49813182-49814168
104	WT1	11	R	32460759-32460800
105	ZFP30	19	F	38146299-38146397
106	ZNF570	19	F	37958078-37958134
107	ZNF71	19	F	57106617-57106852

En la Tabla 1, las bases se numeraron según el conjunto del genoma humano, de febrero de 2009, GRCh37/hg19 (véase, por ejemplo, Rosenbloom y col. (2012) "ENCODE whole-genome data in the UCSC Genome Browser: actualización 2012" Nucleic Acids Research 40: D912-D917). Los marcadores con nombres BHLHE23 y LOC63930 se refieren al mismo marcador.

- 5 En estos candidatos, las firmas de metilación van de 3 CpG a 52 CpG. Algunos marcadores presentan metilación en ambas hebras; otros están semimetilados. Como las hebras no son complementarias después de la conversión con bisulfito, las regiones directa e inversa se contabilizaron por separado. Aunque la Tabla 1 indica la hebra en la que se encontró el marcador, la tecnología no se limita a detectar la metilación solamente en la hebra indicada. La tecnología abarca la medición de la metilación en cualquiera de las hebras directa o inversa y/o en ambas hebras
- 10 directa e inversa; y/o detectar un cambios en el estado de metilación en cualquiera de las hebras directa o inversa y/o en ambas hebras directa e inversa.

- Los niveles de metilación de los cánceres pancreáticos superan en raras ocasiones el 25 % en los CpG filtrados, lo que sugiere que los tejidos con cáncer pueden tener elevados niveles de células y/o estroma normal contaminante. Para estudiar esto, cada uno de los cánceres se secuenció para determinar las mutaciones KRAS para verificar las
- 15 frecuencias alélicas de las muestras positivas. Para el 50 % que tenía un cambio de base KRAS heterocigótico, la frecuencia del alelo mutante era al menos 4 veces menor que el correspondiente alelo natural, lo que respalda la contaminación por células y/o estroma normales.

- Se descubrió que 7 de los 107 marcadores estaban en regiones no anotadas y se encuentran en regiones sin elementos de codificación proteica. Un marcador es adyacente a una secuencia reguladora ARNnc (AK055957). De
- 20 los restantes 99 marcadores candidatos, aproximadamente 30 se habían descrito como asociados con cáncer, algunos de los cuales se clasifican como supresores tumorales. Unos pocos ejemplos:

ADCY1 Regulado defectivamente en osteosarcoma

	ELMO1	Fomenta la invasión del glioma
	HOXA2	Hipermetilado en colangioca
	RSPO3	Regulador de la señalización Wnt
	SUSD5	Media la metástasis ósea en cáncer de pulmón
5	KCNK12	Hipermetilado en cáncer de colon
	CLEC11A	GF citoblástico en leucemia
	USP3	Necesario para la progresión a fase S

Los otros 69 marcadores candidatos tenían anteriormente una baja asociación con el cáncer (por ejemplo, mutaciones y/o alteraciones en el número de copias observado en cribados de todo el genoma) o no tenían asociaciones con cáncer previamente identificadas.

Ejemplo 2 - Validación de marcadores

Para validar las DMR como marcadores del cáncer, se realizaron dos estudios de validación basados en la PCR sobre conjuntos de muestras expandidos. El primer estudio usó muestras de grupos de paciente análogos a los usados en el Ejemplo 1 (por ejemplo, PanC, páncreas normal, colon normal) y se añadieron muestras que comprenden ADN derivado de la capa leucocitaria de pacientes normales sin antecedentes de ningún cáncer. El segundo estudio utilizó una selección pan-cánceres gastrointestinales (pan-GI).

Para el primer estudio de validación, se analizó una combinación de las muestras anteriormente sometidas a RRBS y nuevas muestras depositadas para comprobar la precisión técnica y confirmar la reproducibilidad biológica, respectivamente. Se diseñaron cebadores de PCR específicos de la metilación (MSP) para cada una de las regiones de los marcadores, excluyendo solamente las hebras complementarias en casos de metilación no específica de la hebra. Un programa informático (Methprimer) ayudó al diseño semimanual de los cebadores MSP realizado por personal experimentado; los ensayos se analizaron y se optimizaron mediante qPCR con colorantes SYBR Green sobre diluciones de controles de ADN genómico metilado y no metilado de forma universal. Las secuencias de los cebadores MSP, cada uno de los cuales incluye 2-4 CpG, se diseñaron para proporcionar un medio rápido para evaluar la metilación de las muestras y se sesgaron para maximizar la eficacia de amplificación. Las secuencias de los cebadores y los parámetros físicos se proporcionan en la Tabla 2a y la Tabla 2b:

Tabla 2a: Cebadores MSP

Nombre	Longitud (nt)	Secuencia (5'→ 3')	Contenido en GC (%)	Tm	Ta	SEQ ID NO:
abcb1f	21	GAT TTT GTT CGT CGT TAG TGC	42,9	52,3	60,0	1
abcb1r	19	TCT CTA AAC CCG CGA ACG A	52,6	56,0	60,0	2
adamts17f	25	TTC GAA GTT TCG GGA TAG GAA GCG T	48,0	60,0	65,0	3
adamts17r	20	CCT ACC GAC CTT CGA ACG CG	65,0	60,3	65,0	4
adamts18f	21	GGC GGC GCG TAT TTT TTT CGC	57,1	60,6	60,0	5
adamts18r	23	CGC TAC GAT ATA AAC GAC GAC GA	47,8	56,4	60,0	6
adcy1f	19	GGT TCG GTT GTC GTA GCG C	63,2	59,0	65,0	7
adcy1r	20	CCG ACC GTA ATC CTC GAC GA	60,0	58,6	65,0	8
agfg2f	25	TTA GGT CGG GAA TCG TTA TTG TTT C	40,0	55,1	60,0	9
agfg2r	22	GTA AAT AAC CCC GCG CTA AAC G	50,0	56,5	60,0	10
arhgef7f	24	TTC GTT TGT TTT TCG GGT CGT AGC	45,8	58,1	60,0	11
arhgef7r	24	ACC ACG TAA CGA TTT ACT CGA CGA	45,8	57,8	60,0	12
auts2f	23	CGT TTT CGG ATT TGA AGT CGT TC	43,5	54,8	65,0	13
auts2r	19	CGC CTC GTC TTC CAA CGA A	57,9	57,7	65,0	14
btbd11f	19	AGG GCG TTC GGT TTT AGT C	52,6	55,1	60,0	15
btbd1r	22	AAC CGA AAA CGA CAA AAT CGA T	36,4	53,4	60,0	16
Bvesf	21	TTT GAG CGG CGG TCG TTG ATC	57,1	60,4	60,0	17

ES 2 712 798 T3

Nombre	Longitud (nt)	(continuación)		Contenido en GC (%)	Tm	Ta	SEQ NO:	ID
		Secuencia (5'→ 3')						
Bvesr	22	TCC CCG AAT CTA AAC GCT ACG A		50,0	57,8	60,0	18	
C13orf18f	25	TTT AGG GAA GTA AAG CGT CGT TTT C		40,0	55,6	60,0	19	
C13orf18r	22	AAC GAC GTC TCG ATA CCT ACG A		50,0	57,1	60,0	20	
cacna1cf	22	GGA GAG TAT TTC GGT TTT TCG C		45,5	54,2	65,0	21	
cacna1cr	24	ACA AAC AAA ATC GAA AAA CAC CCG		37,5	55,2	65,0	22	
cbln1f	23	GTT TTC GTT TCG GTC GAG GTT AC		47,8	56,2	60,0	23	
cbln1r	25	GCC ATT AAC TCG ATA AAA AAC GCG A		40,0	56,3	60,0	24	
Cbsf	25	GAT TTA ATC GTA GAT TCG GGT CGT C		44,0	55,2	65,0	25	
Cbsr	22	CCG AAA CGA ACG AAC TCA AAC G		50,0	56,8	65,0	26	
cd1df	17	GCG CGT AGC GGC GTT TC		70,6	60,7	60,0	27	
cd1dr	19	CCC ATA TCG CCC GAC GTA A		57,9	56,9	60,0	28	
celf2f	22	TCG TAT TTG GCG TTC GGT AGT C		50,0	57,0	70,0	29	
celf2r	21	CGA AAT CCA ACG CCG AAA CGA		52,4	58,4	70,0	30	
chr1 156f	24	TTG TCG TTC GTC GAA TTC GAT TTC		41,7	55,8	65,0	31	
chr1 156r	23	AAC CCG ACG CTA AAA AAC GAC GA		47,8	59,2	65,0	32	
chr1 240f	25	TTG CGT TGG TTA CGT TTT TTT ACG C		40,0	57,3	60,0	33	
chr1 240r	23	ACG CCG TAC GAA TAA CGA AAC GA		47,8	58,7	60,0	34	
chr1 353f	21	CGT TTT TCG GGT CGG GTT CGC		61,9	61,5	60,0	35	
chr1 353r	19	TCC GAC GCT CGA CTC CCG A		68,4	63,1	60,0	36	
chr12 133f	22	TCG GCG TAT TTT TCG TAG ACG C		50,0	57,6	60,0	37	
chr12 133r	24	CGC AAT CTT AAA CGT ACG CTT CGA		45,8	57,7	60,0	38	
chr15 291 (apba2)f	24	GGT TTA TAA AGA GTT CGG TTT CGC		41,7	54,4	60,0	39	
chr15 291 (apba2)r	24	AAA ACG CTA AAC TAC CCG AAT ACG		41,7	55,3	60,0	40	
chr2 715f	19	TGG GCG GGT TTC GTC GTA C		63,2	60,2	65,0	41	
chr2 715r	21	GTC CCG AAA CAT CGC AAA CGA		52,4	58,2	65,0	42	
chr6 (TMEM200A)f	130 20	GCG TTT GGA TTT TGC GTT C		55,0	58,0	60,0	43	
chr6 (TMEM200A)r	130 20	AAA ATA CGC CGC TAC CGA TA		55,0	60,6	60,0	44	
chr9 120f	20	GTT TAG GGA GTC GCG GTT AC		55,0	55,4	60,0	45	
chr9 120r	23	CAA ATC CTA CGA ACG AAC GAA CG		47,8	56,2	60,0	46	

ES 2 712 798 T3

Nombre	Longitud (nt)	(continuación)		Contenido en GC (%)	Tm	Ta	SEQ NO:	ID
		Secuencia (5'→ 3')						
clec11af	22	AGT TTG GCG TAG TCG GTA GAT C		50,0	56,4	60,0	47	
clec11ar	22	GCG CGC AAA TAC CGA ATA AAC G		50,0	57,5	60,0	48	
cnr1f	22	TCG GTT TTT AGC GTT CGT TCG C		50,0	58,4	60,0	49	
cnr1r	23	AAA CAA CGA AAC GCC AAT CCC GA		47,8	59,9	60,0	50	
crhr2f	25	TAG TTT TTG GGC GTT ATT TTC GGT C		40,0	56,1	60,0	51	
crhr2r	21	GCA ACT CCG TAC ACT CGA CGA		57,1	59,0	60,0	52	
Dbn1f	26	TTT TTC GTT TGT TTT TCG GTA TTC GC		34,6	55,5	60,0	53	
Dbn1r	22	CGA ATC CTA ACG AAC TAT CCG A		45,5	53,9	60,0	54	
dbx1f	25	TTC GGT GGA TTT TCG TAT TGA TTT C		36,0	54,0	60,0	55	
dbx1r	24	AAA CGA AAC CGC GAA CTA AAA CGA		41,7	57,6	60,0	56	
dhrs12f	22	TTA CGT GAT AGT TCG GGG TTT C		45,5	54,6	60,0	57	
dhrs12r	21	ATA AAA CGA CGC GAC GAA ACG		47,6	56,2	60,0	58	
elmo1f	24	TTT CGG GTT TTG CGT TTT ATT CGC		41,7	57,2	60,0	59	
elmo1r	28	GAA AAA AAA AAA CGC TAA AAA TAC GAC G		28,6	53,3	60,0	60	
Eomesf	21	TAG CGC GTA GTG GTC GTA GTC		57,1	58,4	60,0	61	
Eomesr	18	CCT CCG CCG CTA CAA CCG		72,2	61,5	60,0	62	
fbln1f	22	TCG TTG TTT TAG GAT CGC GTT C		45,5	55,6	60,0	63	
fbln1r	22	GAC GAA CGA TAA ACG ACG ACG A		50,0	56,9	60,0	64	
fem1cf	21	TTC GGT CGC GTT GTT CGT TAC		52,4	58,0	60,0	65	
fem1cr	25	AAA CGA AAA ACA ACT CGA ATA ACG A		32,0	53,8	60,0	66	
fer114f	18	AGT CGG GGT CGG AGT CGC		72,2	62,3	60,0	67	
fer114r	23	ATA AAT CCC TCC GAA ACC CAC GA		47,8	58,2	60,0	68	
fkbp2f	21	TCG GAA GTG ACG TAG GGT AGC		57,1	58,3	60,0	69	
fkbp2r	19	CAC ACG CCC GCT AAC ACG A		63,2	60,6	60,0	70	
flt3f	21	GCG CGT TCG GGT TTA TAT TGC		52,4	57,2	65,0	71	
flt3r	20	GAC CAA CTA CCG CTA CTC GA		55,0	56,1	65,0	72	
fnip1f	20	AGG GGA GAA TTT CGC GGT TC		55,0	57,6	65,0	73	
fnip1r	24	AAC TAA ATT AAA CCT CAA CCG CCG		41,7	55,9	65,0	74	
gfra4f	20	TTA GGA GGC GAG GTT TGC GC		60,0	60,3	65,0	75	
gfra4r	28	GAC GAA ACC GTA ACG AAA ATA AAA ACG A		35,7	56,4	65,0	76	
gjc1r	24	CGA ACT ATC CGA AAA AAC GAC GAA		41,7	55,6	65,0	77	
glc1f	22	GCG ACG CGA GCG TTA ATT TTT C		50,0	57,6	65,0	78	

ES 2 712 798 T3

Nombre	Longitud (nt)	(continuación)		Contenido en GC (%)	Tm	Ta	SEQ NO:	ID
		Secuencia (5'→ 3')						
hecw1f	23	TTC GCG TAT ATA TTC GTC GAG TC		43,5	54,2	60,0	79	
hecw1r	20	CAC GAC CAC TAT CAC GAC GA		55,0	56,5	60,0	80	
hoxa1f	22	GTA CGT CGG TTT AGT TCG TAG C		50,0	55,3	60,0	81	
hoxa1r	21	CCG AAA CGC GAT ATC AAC CGA		52,4	57,6	60,0	82	
ifih1f	20	CGG GCG GTT AGA GGG TTG TC		65,0	60,4	60,0	83	
ifih1r	26	CTC GAA AAT TCG TAA AAA CCC TCC GA		42,3	57,4	60,0	84	
igf2bp1f	29	CGA GTA GTT TTT TTT TTT ATC GTT TAG AC		27,6	52,1	65,0	85	
igf2bp1r	24	CAA AAA ACG ACA CGT AAA CGA TCG		41,7	55,2	65,0	86	
ikzf1f	24	GTT TCG TTT TGC GTT TTT TTG CGC		41,7	57,5	65,0	87	
ikzf1r	19	TCC CGA ATC GCT ACT CCG A		57,9	57,8	65,0	88	
insm1 reg1f	17	GCG GTT AGG CGG GTT GC		70,6	60,2	60,0	89	
insm1 reg1r	25	ATT ATA TCA ATC CCA AAA ACA CGC G		36,0	54,3	60,0	90	
insm1 reg2f	22	TAT TTT TCG AAT TCG AGT TCG C		36,4	51,7	60,0	91	
insm1 reg2r	22	TCA CCC GAT AAA AAC GAA AAC G		40,9	53,8	60,0	92	
kcnk12f	21	GCG TCG TTA GTA GTA CGA AGC		52,4	55,3	60,0	93	
kcnk12r	21	GCA CCT CAA CGA AAA CAC CGA		52,4	58,2	60,0	94	
kcnk2f	23	TCG AGG CGG TTA ATT TTA TTC GC		43,5	55,8	65,0	95	
kcnk2r	23	GCT CTA ACC CAA ATA CGC TAC GA		47,8	56,6	65,0	96	
Kctd15F	22	TCG GTT TCG AGG TAA GTT TAG C		45,5	54,7	60,0	97	
KcTD15r	23	CAC TTC GAA ACA AAA TTA CGC GA		39,1	54,3	60,0	98	
lingo3f	20	GGA AGC GGA CGT TTT CGT TC		55,0	56,8	65,0	99	
lingo3r	22	ACC CAA AAT CCG AAA ACG ACG A		45,5	57,3	65,0	100	
LOC100126784 (NAV2) f	19	AGG TTG CGG GCG TGA TTT C		57,9	58,8	65,0	101	
LOC100126784 (NAV2) r	20	CCA AAA CCA CGC GAA CAC GA		55,0	58,8	65,0	102	
LOC63930 (bhlhe23)f	20	GTT CGG AGT GTC GTA GTC GC		60,0	57,7	70,0	103	
LOC63930 (bhlhe23)r	21	AAT CTC GCC TAC GAA ACG ACG		52,4	57,2	70,0	104	
LOC642345f	22	GTT TAG GGA CGT TTT CGT TTT C		40,9	52,5	65,0	105	
LOC642345r	20	AAC GAA CGC TCG ATA ACC GA		50,0	56,5	65,0	106	
mllt1f	20	TTT GGG TCG GGT TAG GTC GC		60,0	59,9	60,0	107	
mllt1r	25	GAA ACC AAA AAA ACG CTA ACT CGT A		36,0	54,4	60,0	108	
Mpndf	20	CGT TGT TGG AGT TTG GCG TC		55,0	57,1	65,0	109	
Mpndr	21	TAC CCG AAC CGC GAT AAA ACG		52,4	57,5	65,0	110	
myef2f	25	GGT ATA GTT CGG TTT TTA GTC GTT C		40,0	53,6	65,0	111	

ES 2 712 798 T3

Nombre	Longitud (nt)	(continuación)		Contenido en GC (%)	Tm	Ta	SEQ NO:	ID
		Secuencia (5'→ 3')						
myef2r	24	TCT TTT CCT CCG AAA ACC GAA ACG		45,8	57,8	65,0	112	
NDUFAB1f	23	GGT TAC GGT TAG TAT TCG GAT TC		43,5	53,0	60,0	113	
NDUFAB1r	20	ATA TCA ACC GCC TAC CCG CG		60,0	59,7	60,0	114	
NFASCf	24	TTT TGT TTT AAT GCG GCG GTT GGC		45,8	59,6	65,0	115	
NFASCr	22	TAT CCG AAC TAT CCG CTA CCG A		50,0	56,9	65,0	116	
pcbp3f	19	GGT CGC GTC GTT TTC GAT C		57,9	56,6	60,0	117	
pcbp3r	17	GCC GCA AAC GCC GAC GA		70,6	62,4	60,0	118	
PDE6Bf	21	AAT CGG CGG TAG TAC GAG TAC		52,4	56,1	55,0	119	
PDE6Br	26	AAA CCA AAT CCG TAA CGA TAA TAA CG		34,6	53,9	55,0	120	
PLAGL1f	26	GAG TTT TGT TTT CGA AAT TAT TTC GC		30,8	52,4	65,0	121	
PLAGL1r	18	CCC GAA TTA CCG ACG ACG		61,1	55,7	65,0	122	
PRKCBf	21	AGG TTC GGG TTC GAC GAT TTC		52,4	57,3	70,0	123	
PRKCBr	21	AAC TCT ACA ACG CCG AAA CCG		52,4	57,7	70,0	124	
PRRT3f	23	TTA GTT CGT TTA GCG ATG GCG TC		47,8	57,4	60,0	125	
PRRT3r	20	CCG AAA CTA TCC CGC AAC GA		55,0	57,5	60,0	126	
PTF1Af	21	TTC GTC GTT TGG GTT ATC GGC		52,4	57,8	60,0	127	
PTF1Ar	23	GCC CTA AAA CTA AAA CAA CCG CG		47,8	57,1	60,0	128	
RASGRF2f	22	GGT TGT CGT TTT AGT TCG TCG C		50,0	56,6	60,0	129	
RASGRF2r	19	GCG AAA ACG CCC GAA CCG A		63,2	61,4	60,0	130	
RIMKLaf	22	TCG TTT GGG AGA CGT ATT CGT C		50,0	56,7	60,0	131	
RIMKLAr	25	ACT CGA AAA ATT TCC GAA CTA ACG A		36,0	55,0	60,0	132	
RNF216f	20	TCG GCG GTT TTC GTT ATC GC		55,0	58,4	60,0	133	
RNF216r	21	CCA CGA AAC TCG CAA CTA CGA		52,4	57,4	60,0	134	
rspo3f	25	CGT TTA TTT AGC GTA ATC GTT TCG C		40,0	55,0	65,0	135	
rspo3r	24	GAA TAA CGA ACG TTC GAC TAC CGA		45,8	56,6	65,0	136	
RYBPf	24	CGG ACG AGA TTA GTT TTC GTT AGC		45,8	55,7	60,0	137	
RYBPr	24	TCG TCA ATC ACT CGA CGA AAA CGA		45,8	58,4	60,0	138	
SCARF2f	22	TCG GTT CGT AGG TAT ACG TGT C		50,0	55,8	60,0	139	
SCARF2r	22	GCT ACT ACC AAT ACT TCC GCG A		50,0	56,4	60,0	140	
SLC35E3f	21	GTT AGA CGG TTT TAG TTT CGC		42,9	51,8	60,0	141	
SLC35E3r	20	AAA AAC CCG ACG ACG ATT CG		50,0	55,8	60,0	142	
slc38a3f	21	GTT AGA GTT CGC GTA GCG TAC		52,4	55,3	65,0	143	
slc38a3r	25	GAA AAA ACC AAC CGA ACG AAA ACG A		40,0	56,9	65,0	144	

ES 2 712 798 T3

Nombre	Longitud (nt)	(continuación)		Contenido en GC (%)	Tm	Ta	SEQ NO:	ID
		Secuencia (5'→ 3')						
slc6a3f	19	CGG GGC GTT TCG ATG TCG C		68,4	62,0	65,0	145	
slc6a3r	24	CCG AAC GAC CAA ATA AAA CCA ACG		45,8	57,0	65,0	146	
srcin 1f	22	CGT TTT ATG TTG GGA GCG TTC G		50,0	56,8	65,0	147	
srcin 1r	20	GAC CGA ACC GCG TCT AAA CG		60,0	58,5	65,0	148	
st6 gal2f	21	TAC GTA TCG AGG TTG CGT CGC		57,1	59,3	65,0	149	
st6 gal2r	25	AAA CTC TAA AAC GAA CGA AAC TCG A		36,0	54,9	65,0	150	
st8sia1f	21	TCG AGA CGC GTT TTT TGC GTC		52,4	58,7	60,0	151	
st8sia1 r	20	AAC GAT CCC GAA CCG CCG TA		60,0	61,3	60,0	152	
ST8SIA6f	21	CGA GTA GTG CGT TTT TCG GTC		52,4	56,2	60,0	153	
ST8SIA6r	22	GAC AAC AAC GAT AAC GAC GAC G		50,0	56,1	60,0	154	
SUSD5f	22	AGC GTG CGT TAT TCG GTT TTG C		50,0	59,1	65,0	155	
SUSD5r	23	ACC TAC GAT TCG TAA ACC GAA CG		47,8	56,9	65,0	156	
TOX2f	23	AGT TCG CGT TTT TTT CGG TCG TC		47,8	58,5	70,0	157	
TOX2r	21	AAC CGA CGC ACC GAC TAA CGA		57,1	61,0	70,0	158	
twist1f	22	TTG CGT CGT TTG CGT TTT TCG C		50,0	59,9	60,0	159	
twist1r	20	CAA CTC GCC AAT CTC GCC GA		60,0	60,2	60,0	160	
USP3f	18	TAT TGC GGG GAG GTG TTC		55,6	54,7	60,0	161	
USP3r	24	TCA AAA AAT AAT TAA CCG AAC CGA		29,2	51,3	60,0	162	
USP44f	24	TTA GTT TTC GAA GTT TTC GTT CGC		37,5	54,4	60,0	163	
USP44r	19	TCC GAC CCT ATC CCG ACG A		63,2	59,9	60,0	164	
VIMf	27	GAT TAG TTA ATT AAC GAT AAA GTT CGC		29,6	51,0	60,0	165	
VIMr	23	CCG AAA ACG CAT AAT ATC CTC GA		43,5	55,0	60,0	166	
vvc2f	26	TTG GAG AGT TTT TCG AAT TTT TTC GC		34,6	55,2	65,0	167	
vvc2r	19	GAA AAC CAC CCT AAC GCC G		57,9	56,6	65,0	168	
wt1f	17	CGC GGG GTT CGT AGG TC		70,6	58,5	65,0	169	
wt1r	23	CGA CAA ACA ACA ACG AAA TCG AA		39,1	54,5	65,0	170	
zfp30f	22	AGT AGC GGT TAT AGT GGC GTT C		50,0	56,7	65,0	171	
zfp30r	22	GCA TTC GCG ACG AAA ACA AAC G		50,0	58,0	65,0	172	
ZNF569f	20	GTA TTG AGG TCG GCG TTG TC		55,0	55,9	60,0	173	
ZNF569r	19	CCG CCC GAA TAA ACC GCG A		63,2	60,8	60,0	174	
ZNF71f	20	CGT AGT TCG GCG TAG TTC GC		60,0	58,2	65,0	175	
ZNF71r	21	AAC CCG CCC GAC GAC AAT ACG		61,9	62,1	65,0	176	

En la Tabla 2a, Ta es la temperatura de hibridación optimizada, y Tm es la temperatura de fusión en °C en NaCl 50 mM. Los cebadores celf2f y celf2r; LOC63930 (bhlhe23)f y LOC63930 (bhlhe23)r; PRKCBf y PRKCBr; y TOX2f y TOX2r se usaron en una reacción en dos etapas.

Especímenes

Muestras de ADN archivadas de pacientes de la clínica Mayo se usaron en ambas validaciones. Los casos y los controles se enmascararon y se emparejaron por edad y sexo. El primer conjunto de muestra incluyó ADN de 38 adenocarcinomas pancreáticos y controles (20 epitelios colónicos normales, 15 páncreas normales, y 10 capas leucocitarias normales). El segundo conjunto de muestras incluyó ADN de 38 neoplasias colorrectales 20 adenocarcinomas colorrectales y 18 adenomas > 1 cm), 19 adenocarcinomas esofágico, 10 cánceres gástricos (estómago) y 10 colangiocarcinomas.

procedimientos

El ADN archivado se repurificó usando perlas SPRI (AMPure XP-Beckman Coulter) y se cuantificaron mediante absorbancia. 1-2 µg de muestra de ADN se trataron a continuación con bisulfito de sodio y se purificaron usando el protocolo EpiTect (Qiagen). El material eluido se (10-20 ng) amplificó en un Roche 480 LightCycler usando bloques de 384 pocillos. Cada placa aceptó 4 marcadores (y patrones y controles), Usando un total de 23 placas. Los 88 ensayos MSP tuvieron perfiles de amplificación óptima diferentes y se agruparon en consecuencia. En la Tabla 2 se proporcionan las temperaturas de hibridación específica. Las reacciones de 20 µl se efectuaron usando una mezcla maestra LightCycler 480 SYBR I (Roche) y 0,5 µmoles del cebador durante 50 ciclos y se analizaron, generalmente, por el procedimiento de la segunda derivada incluido con el programa informático LightCycler. Los datos sin procesar, expresados como número de copia genómica, se normalizaron respecto a la cantidad de ADN inicial, y se tabularon. El análisis del tejido comprendió realizar una PCA (suplementada mediante validación cruzada k veces), regresión neta elástica, y construcción de gráficos de cajas de los marcadores netos elásticos no cero. De esta forma, los marcadores fueron clasificadas en conjunto. De estos candidatos, debido a la importancia de minimizar la metilación del fondo de células normales para los ensayos basados en heces y en sangre, la clasificación ponderó hacia aquellos marcadores que mostraban un valor mayor de veces de cambio diferencial entre casos y controles.

Resultados

Entre los 107 marcadores de ADN metilados con una discriminación demostrada para cánceres GI, se realizó la validación de 88, de los que se identificaron subconjuntos para la presentación de un resumen de datos más detallado.

Detección de cáncer pancreático

Un subconjunto de marcadores de la metilación fue especialmente discriminante para el cáncer pancreático: ABCB1, ADCY1, BHLHE23 (LOC63930), c13orf18, CACNA1C, chr12 133, CLEC11A, ELMO1, EOMES, GJC1, IHIF1, IKZF1, KCNK12, KCNN2, PCBP3, PRKCB, RSPO3, SCARF2, SLC38A3, ST8SIA1, TWIST1, VWC2, WT1 y ZNF71 (véase la Tabla 1). Los valores de ABC individuales (PanC versus páncreas o colon normales) para estos marcadores fueron superiores a 0,87, lo que indica una sensibilidad y especificidad clínicas superiores.

Inicialmente, los dos mejores marcadores independiente parecían ser CLEC11A y c13orf18, que tenían una sensibilidad del 95 % y del 82 % para el cáncer pancreático, respectivamente, para una especificidad del 95 %. Experimentos adicionales diseñaron cebadores adicionales para dirigirse a los CpG más específico dentro de las DMR especificadas de marcadores seleccionados. Estos cebadores adicionales mejoraron adicionalmente la discriminación. Por ejemplo, el diseño de nuevos MSP para el marcador PRKCB (sensibilidad inicial del 68 %) aumentó de forma importante la discriminación para el cáncer pancreático y consiguió una sensibilidad del 100 % a un 100 % de especificidad. Además, la mediana de la relación señal-ruido para la metilación de este marcador, que compara el tejido canceroso con el normal, fue mayor de 8000. Esto proporciona una métrica fundamental para la detección de marcadores de cáncer en muestras con altos niveles de heterogeneidad celular normal. Disponer de los perfiles de metilación basales de las DMR procedentes de los datos filtrados por RRBS permite la construcción de ensayos de detección muy sensibles y específicos. Estos resultados obtenidos con los diseños de los MSP mejorados demostraron que se puede obtener especificaciones de comportamiento similares a las otras 106 DMR con mejoras adicionales en el diseño, validación, y formatos de ensayo.

Tabla 2b: Cebadores MSP

Nombre	Longitud (nt)	Secuencia (5'→ 3')	Contenido en GC (%)	Tm	Ta	SEQ NO:	ID
dll (sentido directo)r	20	GTC GAG CGC GTT CGT TGT AC	60,0	58,9	65	177	
dll (sentido directo)r	22	GAC CCG AAA AAT AAA TCC CGA A	40,9	53,3	65	178	
dll (sentido contrario)f	24	GAT TTT TTT AGT TTG TTC GAC GGC	37,5	53,5	65	179	

(continuación)

Nombre	Longitud (nt)	Secuencia (5'→ 3')	Contenido en GC (%)	Tm	Ta	SEQ NO:	ID
dll (sentido contrario)r	25	AAA ATT ACT AAA CGC GAA ATC GAC G	36,0	54,4	65	180	
en1 (sentido directo)f	26	TAA TGG GAT GAT AAA TGT ATT CGC GG	38,5	55,2	65	181	
en1 (sentido directo)r	26	ACC GCC TAA TCC AAC TCG AAC TCG TA	50,0	61,2	65	182	
en1 (sentido contrario)f	22	GGT GTT TTT AAA GGG TCG TCG T	45,5	55,7	65	183	
en1 (sentido contrario)r	19	GAC CCG ACT CCT CCA CGT A	63,2	58,4	65	184	
foxp2 (sentido directo)f	30	GGA AGT TTA TAG TGG TTT CGG CGG GTA GGC	53,3	63,6	60	185	
foxp2 (sentido directo)r	22	GCG AAA AAC GTT CGA ACC CGC G	59,1	61,9	60	186	
grin2d (sentido directo)f	28	TGT CGT CGT CGC GTT ATT TTA GTT GTT C	42,9	59,2	60	187	
grin2d (sentido directo)r	22	AAC CGC CGT CCA AAC CAT CGT A	54,6	61,3	60	188	
nr5a1 (sentido directo)f	25	GAA GAG TTA GGG TTC GGG ACG CGA G	60,0	62,6	65	189	
nr5a1 (sentido directo)r	25	AAC GAC CAA ATA AAC GCC GAA CCG A	48,0	61,1	65	190	
nr5a1 (sentido contrario)f	25	CGT AGG AGC GAT TAG GTG GGC GTC G	64,0	64,6	60	191	
nr5a1 (sentido contrario)r	23	AAA CCA AAA CCC GAA ACG CGA AA	43,5	58,5	60	192	
shh (sentido directo)f	26	CGA TTC GGG GGA TGG ATT AGC GTT GT	53,9	62,6	65	193	
shh (sentido directo)r	30	CGA AAT CCC CCT AAC GAA AAT CTC CGA AAA	43,3	60,4	65	194	
shh (sentido contrario)f	25	CGG GGT TTT TTT AGC GGG GGT TTT C	52,0	61,0	65	195	
shh (sentido contrario)r	29	CGC GAT CCG AAA AAT AAA TTA ACG CTA CT	37,9	57,8	65	196	
spsb4 (sentido directo)f	20	AGC GGT TCG AGT TGG GAC GG	65,0	62,3	65	197	
spsb4 (sentido directo)r	24	GAA AAA CGC GAT CGC CGA AAA CGC	54,2	61,8	65	198	
spsb4 (sentido contrario)f	28	GAA GGT TAT TAA TTT AAT AGT CGC GGA A	32,1	53,7	65	199	
spsb4 (sentido contrario)r	25	AAA AAA AAC GTT CCC GAC GAC CGC G	52,0	62,4	65	200	
prkcbf (rediseño)	25	AGT TGT TTT ATA TAT CGG CGT TCG G	40,0	55,3	65	201	
prkcbr (rediseño)	23	GAC TAT ACA CGC TTA ACC GCG AA	47,8	56,9	65	202	

En la Tabla 2b, Ta es la temperatura de hibridación optimizada, y Tm es la temperatura de fusión en °C en NaCl 50 mM.

Detección de otras neoplasias gastrointestinales

- 5 A continuación, los marcadores se evaluaron en el 2º conjunto de muestras, que incluía otros cánceres y precánceres GI, como se ha indicado anteriormente. Los procedimientos, incluidas las condiciones de reacción, y la plataforma, fueron idénticos a la primera validación descrita anteriormente. Los datos se normalizaron respecto a la cantidad de ADN inicial, lo que permite que los números de copia se comparen entre las dos validaciones. El análisis consistió en PCA y validación cruzada k veces, como anteriormente.

Algunas secuencias de metilación que se identificaron mostraron grados de discriminación muy elevados, incluso como marcadores independientes. Por ejemplo, IKZF1 tuvo un 95 % de sensibilidad para adenoma y 80 % de sensibilidad para CRC, con prácticamente nada de metilación del fondo en muestras normales. Las relaciones S/N fueron mayores de 10.000 - un grado de discriminación muy pocas veces visto con cualquier clase de marcadores. El ensayo chr12.133, específico de un tramo completamente desconocido y descrito de ADN metilado, también fue adecuado para detectar todos los cánceres igualmente bien. Varios marcadores (cd1d, chr12.133, clec11a, elmo1, vvc2, zuf71) consiguieron individualmente una discriminación perfecta para el cáncer gástrico, como lo hizo twist1 para el cáncer colorrectal (Tabla 6).

Predicción del sitio del tumor

Los datos recopilados durante el desarrollo de las realizaciones de la tecnología demostraron que los estados de metilación de marcadores de ADN concretos predicen de forma precisa el sitio de la neoplasia. En este análisis, se usó un modelo de regresión con particiones recursivas en un análisis de árbol de decisiones basado en combinaciones de marcadores que se comportan de manera complementaria para generar una clasificación del sitio sólida.

En particular, se realizaron análisis estadísticos para validar los resultados de sensibilidad y especificidad de combinaciones de marcadores. Por ejemplo, usando un modelo de "bosque aleatorio" (véase, por ejemplo, Breiman (2001) "Random Forests" Machine Learning 45: 5-32), se construyeron modelos de árbol usando regresión de árboles con particiones recursivas, por ejemplo, tal como se implementa en el paquete rPart en el programa informático estadístico R. La regresión de árboles con particiones recursivas es una técnica de regresión que intenta minimizar una función de pérdida y de esta forma maximiza el contenido de información en problemas de clasificación. El árbol se construye mediante el siguiente procedimiento: en primer lugar, se encuentra la variable individual que mejor divide los datos en dos grupos. Los datos se separan, y a continuación este procedimiento se aplica por separado a cada subgrupo, y así de forma recursiva, hasta que cada uno de los subgrupos alcanza un tamaño mínimo o hasta que no se pueden hacer mejoras. La segunda etapa del procedimiento consiste en utilizar validación cruzada para reducir todo el árbol. Una estimación del riesgo por validación cruzada se calcula a partir de un conjunto de subárboles anidados, y se produce un modelo final a partir del subárbol con menor estimación del riesgo. Véase, por ejemplo, Therneau (2012) "An Introduction to Recursive Partitioning Using RPART Routines", available at The Comprehensive R Archive Network; Breiman y col. (1983) "Classification and Regression Trees" Wadsworth, Belmont, CA; Clark y col. (1992) "Tree-based models" in J.M. Chambers y T.J. Hastie, eds., Statistical Models in S, capítulo 9. Wadsworth y Brooks/Cole, Pacific Grove, CA; Therneau (1983) "A short introduction to recursive partitioning" Orion Technical Report 21, Universidad de Stanford, Departamento de estadística; Therneau y col. (1997) "An introduction to recursive partitioning using the rpart routines" Division of Biostatistics 61, Clínica Mayo.

Como se usa en este análisis, la clasificación es Lesión GI superior vs. Lesión GI inferior vs. Muestras normales. En cada nodo de la regresión, todas las variables se consideran para la entrada, pero solamente la variable de mayor disminución en el riesgo del resultado previsto se introduce. Se añaden nodos posteriores al árbol hasta que no se produce ningún cambio en el riesgo. Para evitar sobreajustes, se usó la regresión de bosque aleatorio. En este enfoque, se generaron 500 árboles de predicción usando análisis con sustitución de muestras y selección aleatoria de variables. Para determinar la importancia de la variable i -ésima, la i -ésima variable se deja de lado, y se comparan las correspondientes tasas de error para el ajuste completo (incluidos todos los datos) vs. el ajuste reducido (todos los datos salvo la i -ésima variable) usando las 500 predicciones.

Se construyó un bosque de 500 árboles para estudiar la potencia predictiva de los marcadores candidatos para discriminar entre el tejido normal, las lesiones gastrointestinales superiores y las lesiones gastrointestinales inferiores. Este procedimiento se realiza con un grado de solidez muy alto. En primer lugar, para cada creación de árbol, se toma una muestra del análisis con sustitución de la base de datos para crear un conjunto de entrenamiento, y todas las observaciones no seleccionadas se utilizan como conjunto de validación. En cada rama del árbol, se usa un subconjunto aleatorio de marcadores, que se evalúa para determinar el mejor marcador para uso en ese determinado nivel del árbol. En consecuencia, todos los marcadores tienen la misma oportunidad de seleccionarse. La técnica proporciona una validación y evaluación rigurosa de la importancia relativa de cada marcador. Cada uno de los 500 árboles puede "votar" acerca de la clase a la que pertenece una muestra en concreto, siendo el ganador el que obtiene mayoría de votos. La tasa de clasificación incorrecta estimada se estima para todas las muestras no utilizadas en un árbol concreto.

Para analizar la importancia relativa de un marcador dado, el conjunto de validación se vuelve a usar. Aquí, una vez que un árbol se ha ajustado, los datos de validación se hacen descender por el árbol, y se anota el índice de clasificación correcto. Después, los valores de los marcadores se permutan aleatoriamente para el m -ésimo marcador, se hace descender por el árbol, y se vuelve a anotar la clasificación correcta. Si un marcador tiene mucha importancia, los datos reales proporcionan una mejor clasificación que los datos permutados aleatoriamente. La clasificación incorrecta de los datos permutados se conoce como la disminución media en la precisión. Si un marcador no es importante, los datos reales proporcionarán una clasificación similar a los datos aleatoriamente permutados. La Figura 1 es una representación gráfica de la importancia de los marcadores según se determina por la disminución media en la precisión. Las líneas verticales se encuentran al 2,5 % y al 5 %. Estos datos indican que, por ejemplo, para clec11a, la disminución media en la precisión estimada es de aproximadamente un 12 %, lo que

indica que, cuando se permutan aleatoriamente los resultados de este marcador, la precisión global de la predicción disminuye en un 12 %. La Figura 1 lista los marcadores en orden de importancia.

La tasa de clasificación incorrecta global estimada para los 500 árboles del bosque fue 0,0989. Los resultados del procedimiento de votación en los 500 árboles del bosque se resumen en la Tabla 3, y se expande por subtipo en la Tabla 4. En las tablas, el tipo de muestra de tejido se relaciona en la primera columna (por ejemplo, no canceroso ("Normal"), cáncer gastrointestinal superior ("Superior"), o cáncer gastrointestinal inferior ("Inferior") en la Tabla 3; adenoma ("Ad"), colon normal ("Colo Normal"), cáncer colorrectal ("CCR"), cáncer esofágico ("Eso C"), cáncer pancreático ("Pan C"), páncreas normal ("Pan Normal"), y cáncer de estómago ("Estómago C") en la Tabla 4). Se proporciona una clasificación cuantitativa de las muestras mediante el análisis como un número en las columnas 1, 2, o 3, para su clasificación como cáncer gastrointestinal superior (columna 1), cáncer gastrointestinal inferior (columna 2), o tejido normal (columna 3), respectivamente. Los números proporcionan una medición que indica la tasa de éxito del clasificador (por ejemplo, el número de veces que el clasificador clasifica el tipo de muestra en la primera columna como el tipo que se indica en la primera fila).

Tabla 3

	1	2	3	error de clase
Superior	59,00	1,00	7,00	0,12
Inferior	3,00	33,00	2,00	0,13
Normal	1,00	0,00	44,00	0,02

columna 1 = GI superior; columna 2 = GI inferior; Columna 3 = normal

Tabla 4

Tipo de muestra	UGIC*	Previsto en el modelo CRN**	Normal
UGIC*			
Cáncer de páncreas	35	0	3
Cáncer de esófago	15	0	3
Cáncer de estómago	9	1	0
CRN**			
Cáncer de colon	2	16	2
Adenoma de colon	1	17	0
Controles			
Páncreas Normal	0	0	15
Colon Normal	0	0	20
Capa leucocitaria normal	1	0	9

* UGIC = Cáncer GI superior, ** CRN = CRC + Adenoma ≥ 1 cm

Un análisis adicional demostró que una combinación de dos marcadores predijo con precisión el sitio del tumor en >90 % de las muestras, las 17 principales combinaciones de marcadores predijeron con precisión el sitio del tumor en >80 % de las muestras, y las 49 principales combinaciones de marcadores predijeron con precisión el sitio del tumor en el 70 % de las muestras. Esta observación de que múltiples combinaciones de marcadores de la metilación del ADN predicen con precisión el sitio del tumor demuestra la solidez de la tecnología.

Usando los dos marcadores principales en el árbol de decisión con particiones recursivas, todos los tejidos normales se clasificaron correctamente como normales, todos los cánceres gástricos se clasificaron correctamente como GI superior, casi todos los cánceres esofágicos y pancreáticos se clasificaron correctamente como GI superior, y casi todos los cánceres colorrectales y precánceres (adenomas) se clasificaron correctamente como GI inferior. Durante el desarrollo de las realizaciones de la tecnología proporcionada en el presente documento, los análisis estadísticos se centraron en un conjunto de marcadores específico que consiste en clec11a, c13orf18, kcnn2, abcb1, slc38a3, cdcl, ikzf1, adcy1, chr12133, rsop3 y twist1. En particular, los análisis estadísticos anteriormente descritos estaban dirigidos a identificar conjuntos de marcadores (por ejemplo, con dos o más marcadores) que proporciona una mayor potencia para identificar el cáncer y/o discriminar entre cánceres. La Tabla 5 resume la precisión para cada conjunto de marcadores emparejados, concretamente proteínas clec11a, c13orf18, kcnn2, abcb1, slc38a3, cdcl, ikzf1, adcy1, chr12133, rsop3 y twist1. De acuerdo con este análisis, el par de marcadores compuesto por cleclla y twist1 es el más informativo, pero otras diversas combinaciones tienen precisiones similares.

Tabla 5 - Precisión para la predicción del sitio usando varias combinaciones de marcadores

precisión	marcadores
90,7	clec11a twist1

ES 2 712 798 T3

(continuación)

precisión	marcadores
90,7	clec11a twist1
88,7	clec11a chr12.133
88,7	clec11a rspo3
88	clec11a ikzf1
86,7	clec11a adcy1
84,7	twist1 c13orf18
84	clec11a ed1d
83,3	twist1 abcb1
83,3	c13orf18 chr12.133
83,3	abcb1 chr12.133
83,3	abcb1 rspo3
82	c13orf18 rspo3
81,3	abcb1 ikzf1
80,7	abcb1 adcy1
80	twist1 kcnn2
80	c13orf18 adcy1
80	cd1d rspo3
79,3	c13orf18 cd1d
79,3	kcnn2 adcy1
79,3	kcnn2 rspo3
79,3	cd1d ikzf1
78,7	c13orf18 ikzf1
77,3	kcnn2 ikzf1
77,3	abcb1 cd1d
76,7	twist1 cd1d
76,7	kcnn2 chr12.133
76,7	chr12.133 rspo3
76	cd1d chr12.133
75,3	twist1 rspo3
75,3	kcnn2 cd1d
74,7	twist1 ikzf1
74	twist1 slc38a3
74	slc38a3 ikzf1
74	slc38a3 chr12.133
73,3	twist1 chr12.133
73,3	slc38a3 adcy1
73,3	adcy1 rspo3
72,7	slc38a3 rspo3
72	cd1d adcy1
72	ikzf1 chr12.133
72	adcy1 chr12.133
71,3	ikzf1 adcy1
70,7	clec11a c13orf18
70,7	clec11a kcnn2
70,7	clec11a abcb1
70,7	clec11a slc38a3
70,7	ikzf1 rspo3
70	twist1 adcy1
70	kcnn2 abcb1

(continuación)

precisión	marcadores
68	slc38a3 cd1d
66,7	c13orf18 abcb1
65,3	c13orf18 kcn2
65,3	kcn2 slc38a3
64,7	c13orf18 slc38a3
56	abcb1 slc38a3

Ejemplo 3 - Análisis de la ABC de los marcadores individuales

El análisis estadístico incluyó el análisis de los componentes principales para identificar combinaciones lineales no correlacionadas de los marcadores cuya varianza explica el mayor porcentaje de variabilidad observado en los datos originales. El análisis determinó la ponderación relativa de cada marcador para discriminar entre grupos de tratamiento. Como resultado de este análisis, los valores de la ABC en el punto final se determinaron para un subconjunto de los marcadores que mida la potencia de cada marcador para discriminar un cáncer específico (esofágico, estómago, pancreático, colorrectal, y adenoma) de 1) los otros tipos de cáncer y de 2) muestras normales (por ejemplo, que no comprenden tejido canceroso o no son de un paciente que tiene cáncer o que puede desarrollar cáncer). En la Tabla 6 se proporcionan estos datos.

Tabla 6 - Valores de la ABC para un subconjunto de los marcadores

	BMP3	NDRG4	abcb1	adcy1
Eso C. vs. Otro	0,51	0,58	0,67	0,39
Eso C. vs. Normal	0,82	0,86	0,83	0,63
Estómago C. vs. Otro	0,72	0,70	0,87	0,65
Estómago C. vs. Normal	0,91	0,95	1,00	0,86
Pan C. vs. Otro	0,59	0,66	0,73	0,69
Pan C. vs. Normal	0,90	0,90	0,91	0,94
CRC. vs. Otro	0,74	0,59	0,46	0,69
CRC. vs. Normal	0,91	0,87	0,72	0,86
Ad. vs. Otro	0,74	0,71	0,35	0,71
Ad. vs. Normal	0,96	0,94	0,61	0,99

	c13orf18	cacna1c	cd1d	chr12.133
Eso C. vs. Otro	0,60	0,27	0,52	0,52
Eso C. vs. Normal	0,75	0,42	0,85	0,86
Estómago C. vs. Otro	0,78	0,70	0,75	0,81
Estómago C. vs. Normal	0,88	0,96	1,00	1,00
Pan C. vs. Otro	0,81	0,85	0,73	0,57
Pan C. vs. Normal	0,89	0,96	0,94	0,86
CRC. vs. Otro	0,37	0,56	0,67	0,73
CRC. vs. Normal	0,51	0,75	0,88	0,89
Ad. vs. Otro	0,21	0,42	0,54	0,72
Ad. vs. Normal	0,35	0,53	0,88	0,99

	clec11a	elmo1	eomes	glc1
Eso C. vs. Otro	0,55	0,46	0,37	0,51
Eso C. vs. Normal	0,81	0,76	0,54	0,69
Estómago C. vs. Otro	0,84	0,76	0,70	0,74
Estómago C. vs. Normal	1,00	1,00	0,89	0,88
Pan C. vs. Otro	0,89	0,62	0,70	0,54
Pan C. vs. Normal	0,98	0,93	0,87	0,73
CRC. vs. Otro	0,31	0,71	0,61	0,64
CRC. vs. Normal	0,56	0,83	0,79	0,80

ES 2 712 798 T3

(continuación)

Ad. vs. Otro	0,35	0,70	0,59	0,65
Ad. vs. Normal	0,59	0,92	0,77	0,82

	ihif1	kcnk12	kcnk2	loc63930
Eso C. vs. Otro	0,11	0,39	0,68	0,40
Eso C. vs. Normal	0,10	0,66	0,84	0,69
Estómago C. vs. Otro	0,80	0,65	0,76	0,65
Estómago C. vs. Normal	0,98	0,90	0,91	0,88
Pan C. vs. Otro	0,91	0,71	0,76	0,61
Pan C. vs. Normal	0,97	0,94	0,91	0,88
CRC. vs. Otro	0,50	0,71	0,46	0,84
CRC. vs. Normal	0,58	0,93	0,67	0,95
Ad. vs. Otro	0,21	0,67	0,30	0,69
Ad. vs. Normal	0,22	0,92	0,47	0,93

	prkcb	rspo3	scarf2	slc38a3
Eso C. vs. Otro	0,44	0,42	0,13	0,34
Eso C. vs. Normal	0,62	0,68	0,21	0,50
Estómago C. vs. Otro	0,71	0,64	0,70	0,81
Estómago C. vs. Normal	0,85	0,86	0,82	0,97
Pan C. vs. Otro	0,74	0,57	0,93	0,83
Pan C. vs. Normal	0,90	0,93	0,94	0,96
CRC. vs. Otro	0,56	0,80	0,49	0,57
CRC. vs. Normal	0,71	0,93	0,57	0,73
Ad. vs. Otro	0,46	0,82	0,26	0,32
Ad. vs. Normal	0,66	1,00	0,34	0,47

	twist1	vwc2	wt1	znf71
Eso C. vs. Otro	0,42	0,52	0,35	0,70
Eso C. vs. Normal	0,74	0,83	0,66	0,90
Estómago C. vs. Otro	0,58	0,78	0,70	0,89
Estómago C. vs. Normal	0,92	1,00	0,91	1,00
Pan C. vs. Otro	0,67	0,58	0,76	0,50
Pan C. vs. Normal	0,94	0,92	0,98	0,79
CRC. vs. Otro	0,83	0,72	0,69	0,63
CRC. vs. Normal	1,00	0,90	0,92	0,91
Ad. vs. Otro	0,70	0,76	0,64	0,64
Ad. vs. Normal	0,95	0,98	0,89	0,90

	st8sia1	ikzf1	pcbp3	PCA1
Eso C. vs. Otro	0,45	0,55	0,51	0,47
Eso C. vs. Normal	0,64	0,88	0,86	0,79
Estómago C. vs. Otro	0,77	0,74	0,76	0,81
Estómago C. vs. Normal	0,92	0,97	0,97	0,99
Pan C. vs. Otro	0,65	0,49	0,64	0,72
Pan C. vs. Normal	0,93	0,85	0,86	0,96
CRC. vs. Otro	0,58	0,81	0,67	0,68
CRC. vs. Normal	0,74	0,94	0,90	0,96
Ad. vs. Otro	0,67	0,80	0,63	0,62
Ad. vs. Normal	0,84	0,99	0,86	0,98

Ejemplo 4 - Esófago de Barrett y cáncer esofágico

El desarrollo de cáncer esofágico está estrechamente ligado a la metaplasia epitelial de Barrett y el adenocarcinoma pancreático surge de metaplasias discretas en las células de la mucosa. Véase, por ejemplo, Biankin y col (2003) "Molecular pathogenesis of precursor lesions of pancreatic ductal adenocarcinoma" *Pathology* 35:14-24; Cameron y col (1995) "Adenocarcinoma of the esophagogastric junction and Barrett's esophagus" *Gastroenterology* 109: 1541-1546.

Para disminuir de forma significativa la creciente incidencia de adenocarcinoma esofágico, se necesitan procedimientos eficaces para seleccionar la población para el precursor crítico del esófago de Barrett (EB). Se han propuesto herramientas mínimamente invasivas o no invasivas para el cribado del EB, pero se han visto dificultados por la falta de marcadores óptimamente sensibles y de marcadores específicos. Los marcadores de cribado deseados discriminan el EB de la mucosa esofagogástrica normal. Algunos genes con metilación anómala están asociados como marcadores candidatos para el Be (véase, por ejemplo, *Gastroenterology* 2011; 140: S-222).

En consecuencia, durante el desarrollo de las realizaciones de la tecnología, se realizaron experimentos para evaluar los marcadores del ADN metilado seleccionados para discriminar el EB del esófago escamoso (SE) adyacente y cardias gástrico (GC) y del SE y GC en controles sanos.

Se reclutaron pacientes con y sin EB conocido antes de una endoscopia superior rutinaria. Los casos de EB tenían >1 cm de longitud de la mucosa columnar perimetral, con metaplasia intestinal confirmada histológicamente; los controles no tenían BE como se determinó endoscópicamente. Se obtuvieron biopsias en los casos de EB, GC (1 cm bajo la línea Z), y SE (>2 cm por encima de EB), y en los controles de GC (como para EB) y SE (5 cm por encima de la línea Z), y se congelaron de forma rápida. Las muestras de biopsia se procesaron en lote, y se sometieron a ensayo de forma enmascarada. Después de la extracción del ADN y el tratamiento con bisulfito, la metilación de los genes diana se analizó mediante la PCR específica de metilación para los marcadores APC, HPP1, SFRP1, y ensayo QuARTS para los marcadores BMP3 y NDRG4. β -actina se cuantificó como un marcador del control para el ADN humano total.

Entre 25 casos de BE y 22 controles, las medianas de edad fueron 67 (intervalo 39-83) y 50 (intervalo 20-78), respectivamente, y los varones representaron el 72% y el 46 % de los sujetos en los grupos de EB y del control, respectivamente. La mediana de la longitud del EB fue de 6 cm (intervalo 2-14 cm). Salvo para APC, la mediana de los niveles de los marcadores metilados fue significativa y sustancialmente (por ejemplo, 200-1100 veces) mayor en el EB que en el SE y GC adyacentes, o respecto a los SE y GC normales. Las sensibilidades del EB para varias especificidades se muestran para cada marcador (Tabla 7). Los marcadores metilados fueron significativamente superiores en el GC adyacente al EB que en el GC de los controles normales. El APC metilado fue mayor en el EB que en SE, pero no distingue el EB del GC. A diferencia de los marcadores metilados, las distribuciones de β -actina fueron similares para todos los grupos de tejidos. Los niveles de los marcadores aumentaron con la longitud del EB para NDRG4, SFRP1, BMP3, y HPP1 ($p = 0,01, 0,01, 0,02, y 0,04$, respectivamente). Los factores que no afectan significativamente los niveles de marcadores incluyen la edad, el sexo, la inflamación, y la presencia de displasia (ninguna (8), grado bajo (6), grado alto (11)).

Por tanto, estos datos se muestran que los marcadores de ADN metilado seleccionado discriminan fuertemente el EB de GC y SE, y proporcionan aplicaciones de cribado útiles.

Tabla 7

	Sensibilidad para el EB, %				
	<i>NDRG4</i>	<i>SFRP1</i>	<i>BMP3</i>	<i>HPP1</i>	<i>APC</i>
Corte de especificidad*					
100 %	96	96	84	84	0
95 %	96	96	92	88	8
90 %	96	96	92	92	8
*Basado en los datos combinados de SE y GC de los controles normales					

Ejemplo 5 - Los marcadores de ADN metilados en el jugo pancreático discriminan el cáncer pancreático de la pancreatitis crónica y controles normales

El análisis del jugo pancreático se ha explorado como un enfoque mínimamente invasivo para la detección precoz del cáncer pancreático (PC). Sin embargo, la citología y muchos marcadores moleculares en el jugo pancreático han demostrado no ser sensibles o no han logrado distinguir entre PC de la pancreatitis crónica (véase, por ejemplo, *J Clin Oncol* 2005; 23: 4524). Se realizaron experimentos para comprobar que el ensayo de genes con metilación anómala puede representar un enfoque más preciso para detección del PC que el jugo pancreático (véase, por ejemplo, *Cancer Res* 2006; 66: 1208). En particular, los datos recogidos para evaluar los marcadores de ADN metilados se analizaron en el jugo pancreático para discriminar casos de pacientes con PC de los controles con

pancreatitis crónica (CP) o un páncreas normal (NP).

Un panel de 110 pacientes (66 PC, 22 CP, 22 controles NP) se sometieron a recogida del jugo pancreático estimulada con secretina durante una prueba endoscópica con ultrasonidos. Los diagnósticos fueron histológicamente confirmados para PC y se basaron en radiografías para CP y NP. El jugo se congeló rápidamente y se almacenó a -80 °C. Se realizaron ensayos de forma enmascarada sobre las muestras descongeladas por lotes. Los marcadores de ADN metilado candidatos se seleccionaron mediante secuenciación del metiloma completo en un estudio de tejidos independiente. Una vez que el ADN se extrajo del jugo pancreático y se trató con bisulfito, se determinó la metilación génica mediante la PCR específica de metilación para CD1D, CLEC11A, y KCNN2, o por QuARTS para BMP3 y NDRG4. Se estudiaron las mutaciones KRAS (7 total) mediante QuARTS (la presencia de cualquier mutación KRAS se consideró como un positivo). β -actina, un marcador del ADN humano, también se analizó mediante QuARTS, para proporcionar un control de la cantidad de ADN.

Respectivamente para PC, CP, y NP, la mediana de la edad fue 67 (intervalo 43-90), 64 (intervalo 44-86), y 60 (intervalo 35-78); los hombres representaron 56, 68, y 21 % de estos grupos, respectivamente. Todos los marcadores discriminaron PC de NP pero en cantidad variable. La ABC fue 0,91 (95 % IC, 0,85-0,97), 0,85 (0,77-0,94), 0,85 (0,76-0,94), 0,78 (0,67-0,89), y 0,75 (0,64-0,87) para CD1D, NDRG4, CLEC11A, KCNN2 y BMP3 metilado, respectivamente, y 0,75 (0,64-0,86) para KRAS mutante. La discriminación para PC mediante CD1D fue significativamente superior que mediante KRAS ($p=0,01$), KCNN2 ($p=0,02$), o BMP3 ($p<0,01$). Se muestran las tasas de positividad para PC y CP para cada marcador al 95 y 100 % de los corte de especificidad normales (Tabla 8); la tasa de positividad en CP (positivos falsos) fue menor para CD1D y mayor para KRAS. Los niveles de marcadores no se vieron significativamente afectados por el sitio de PC (cabeza, cuerpo, cola) o etapa (N0 vs. N1). Los niveles de β -actina fueron similares entre los grupos de pacientes.

Estos datos muestran que los marcadores de ADN metilado discriminan PC de CP y NP cuando se someten a ensayo para el jugo pancreático, por ejemplo, jugo pancreático estimulado por secretina. En particular, el CD1D metilado fue significativamente más sensible para PC y mostró sustancialmente menos positivos falsos con CP de lo que lo hizo KRAS mutante.

Tabla 8

	Tasas de positividad, %			
	Para especificidad 95 %*		Para especificidad 100 %*	
	PC	CP	PC	CP
Marcadores de la metilación				
<i>CD1D</i>	75	9	63	5
<i>NDRG4</i>	67	14	56	5
<i>CLEC11A</i>	56	18	38	5
<i>KCNN2</i>	33	18	33	18
<i>BMP3</i>	31	9	23	5
Marcador de mutación				
<i>KRAS</i>	55	41	53	32

*Cortes de especificidad basados en los datos NP

Ejemplo 6 - Paneles de marcadores de ADN sensibles para la detección del cáncer pancreático mediante análisis en jugo pancreático

El análisis del jugo pancreático representa un enfoque mínimamente invasivo para la detección de cáncer pancreático (PC) y del precáncer. Se ha descubierto que los marcadores de ADN metilado específicos en el jugo pancreático discriminan entre PC y la pancreatitis crónica (CP) y el páncreas normal (Gastroenterology 2013;144:S-90), pero nuevos marcadores y combinaciones de marcadores siguen por descubrir.

Se han realizado experimentos para evaluar el valor de los marcadores del ADN metilado recientemente descubiertos y de los KRAS mutantes analizados en solitario y combinados en el jugo pancreático para discriminar entre PC y la pancreatitis crónica (CP) y los controles de referencia (CON).

167 pacientes (85 PC, 30 CP, 23 con neoplasia mucinosa intraductal premaligna (IPMN), 29 CON) que se habían sometido a recogida de jugo pancreático estimulada por secretina durante EUS fueron objeto de estudio. Los diagnósticos se basaron histológicamente para PC, radiográficamente para PC, e histológica o radiográficamente para IPMN. La especificidad se basó en CON, que incluía pacientes con factores de riesgo para PC, valor elevado de enzimas pancreáticas, o síntomas GI pero páncreas radiográficamente normales. Las muestras de jugo archivadas a -80 °C se analizaron por lotes de forma enmascarada. Sobre el ADN extraído de 200 μ l de jugo pancreático, se determinó la metilación de los genes después del tratamiento con bisulfito mediante amplificación de la diana y la señal en tiempo real específica de alelo cuantitativa (QuARTS) para analizar *ADCY1*, *CD1D*, *BMP3*,

PRKCB, KCNK12, C13ORF18, IKZF1, CLEC11A, TWIST1, NDRG4, ELMO, y 55957. Las mutaciones de KRAS mutante (7 total) y β -actina (un marcador del ADN humano total) también se analizaron mediante QuARTS. Para los datos cuantitativos, se siguió un algoritmo para conseguir la discriminación óptima mediante un panel que combina todos los marcadores.

5 Respectivamente para PC, CP, IPMN, y CON: la mediana de edad fue 67 (IQR 60-77), 66 (55-77), 66 (60-76) y 70 (62-77); los hombres comprendían 52, 53, 49, y 72 %. Para los cortes de especificidad respectivos del 90 % y 95 %: el panel de marcadores combinado consiguió las sensibilidades más altas para PC (88 % y 77 %); *ADCY1*, el marcador individual más sensible, detectó 84 % y 71 %. Otros marcadores individuales detectaron PC pero en cantidad variable (tabla). La discriminación por área bajo la curva COR fue mayor según el panel que para cualquier otro marcador individual ($p < 0,05$), salvo *ADCY1* (tabla). Para el 90 % de especificidad, el panel detectó el 44 % de todos los IPMN y el 75 % (3/4) del subconjunto con un elevado grado de displasia. Las tasas de positividad fueron sustancialmente menores para CP que para PC para todos los marcadores que se muestran en la Tabla 7 ($p < 0,0001$). Para el 100 % de especificidad, el panel fue positivo para el 58 % de PC, 17 % de IPMN, y 13 % de CP. En consecuencia, estos resultados demuestran que un panel de novedosos marcadores de ADN metilado y KRAS mutante analizado en el jugo pancreático consiguió alta sensibilidad para PC.

Tabla 7: Tasas de positividad para marcadores en jugo pancreático de pacientes con cáncer pancreático (PC), neoplasia mucinosa papilar intraductal (IPMN), y pancreatitis crónica pancreatitis (CP).

	ABC (PC vs Con)	Tasas de positividad, %					
		Para especificidad 90 %*			Para especificidad 95 %*		
		PC	IPMN	CP	PC	IPMN	CP
Marcadores de la metilación**							
ADCY1	0,89	84	39	30	71	35	17
C13ORF18	0,82	67	17	13	52	13	3
PRKCB	0,82	62	9	20	42	4	17
CD1D	0,82	61	22	10	46	17	10
KCNK12	0,82	54	17	10	25	13	3
BMP3	0,81	49	13	7	27	13	0
IKZF1	0,80	71	22	20	54	22	17
Marcador de mutación							
KRAS	0,80	59	22	13	58	17	10
Todos los marcadores							
Panel***	0,91	88	44	37	77	39	23

*Cortes de especificidad basados en los datos del control de referencia (CON).
 **Se muestran los 7 principales marcadores del ADN metilado individuales.
 ***Salvo para *ADCY1*, el panel tuvo una ABC significativamente más elevada que los marcadores del ADN metilado individuales ($p < 0,05$).

Ejemplo 7: Detección del cáncer pancreático con una muestra de heces usando el marcador CD1D.

Muestras de heces de 45 individuos con cáncer pancreático, y 45 individuos que no tenían cáncer pancreático se recogieron y se analizaron para determinar la presencia del marcador CD1D. El cáncer pancreático se detectó correctamente usando el marcador CD1D en heces.

Ejemplo 8: Novedosos marcadores de la metilación del ADN asociados con el cáncer pancreático en etapa inicial.

Introducción al estudio:

En estudios de tejido de control del caso independientes, se realizaron experimentos para identificar novedosos y fuertemente discriminantes marcadores de la metilación para PanC usando RRBS durante la fase de descubrimiento y PCR específica de metilación (MSP) para la fase de validación.

Población del estudio:

Tras su autorización por el Comité de revisión institucional de la Clínica Mayo, se identificaron muestras de tejidos a partir de registros de cáncer existentes. La población accesible incluyó los que se han sometido a una pancreatectomía distal, pancreaticoduodenotomía, colectomía o biopsia de colon con un espécimen congelado archivado. Todos los tejidos se revisaron por un patólogo gastrointestinal experto para confirmar la clasificación correcta. Las muestras del caso PanC incluyeron tejidos con adenocarcinoma pancreático ductal limitado a la enfermedad en la etapa inicial (AJCC estadio I y II) (Edge SBB, D.R.; Compton, C.C.; Fritz, A.G.; Greene, F.L.; Trotti,

A. (Eds.), editor. AJCC Cancer Staging Manual. 7ª ed: Springer, Nueva York; 2010). Se excluyeron las neoplasias derivadas de lesiones IPMN. Se estudiaron dos grupos del control. El primero, denominado "páncreas normal", incluyó los márgenes de resección histológica normal de bajo riesgo (cistadenoma seroso) o neoplasias pancreáticas focales (tumores neuroendocrinos). El segundo grupo del control incluyó tejidos epiteliales colónicos procedentes de pacientes con una confirmación de no padecer PanC o neoplasia colónica. Los casos y ambos controles se emparejaron por sexo, edad (en incrementos de 5 años) y estado de tabaquismo (actual o anterior frente a nunca). En un laboratorio centralizado, los tejidos de los casos y del control se microdesecaron, y el ADN se extrajo usando una técnica fenol-cloroformo, proporcionando al menos 500 ng de ADN. La identificación de los casos, el emparejamiento y la extracción de ADN se realizaron por personal independiente para mantener el enmascaramiento del personal de laboratorio por caso y estado del control.

Secuenciación con representación reducida de bisulfito:

La preparación de la biblioteca (Gu H, Bock C, Mikkelsen TS, Jager N, Smith ZD, Tomazou E, y col. Genome-scale DNA methylation mapping of clinical samples at single-nucleotide resolution. Nat Methods. 2010;7:133-6): El ADN genómico (300 ng) se fragmentó por digestión con 10 unidades de MspI, una enzima de restricción específica de metilación que reconoce los motivos que contienen CpG. Esto enriquece las muestras en su contenido de CpG y elimina zonas redundantes del genoma. Se repararon los extremos de los fragmentos digeridos y se agregaron colas A con 5 unidades de fragmento de Klenow (3'-5' exo-) y, se ligó durante la noche a adaptadores de TruSeq (Illumina, San Diego CA) que contenían de una a cuatro secuencias de código de barras para enlazar cada fragmento a su ID de muestra). Se llevó a cabo la selección por tamaño de los fragmentos de 160-340 pb (inserciones de 40-220 pb) usando perlas/tampón Agencourt AMPure XP SPRI (Beckman Coulter, Brea CA). Los cortes de tampón fueron 0,7x a 1,1x volúmenes de muestra de perlas/tampón. El volumen fue de 22 ul (tampón EB - Qiagen, Germantown MD) Se utilizó la PCR para controlar la eficacia de enlazado y la calidad de fragmentos en una pequeña alícuota de la muestra. A continuación, las muestras se sometieron a conversión con bisulfito usando un protocolo EpiTect modificado (Qiagen). qPCR y PCR convencional (Pfu Turbo Cx hotstart, Agilent, Santa Clara CA) seguido por evaluación con Bioanalyzer 2100 (Agilent) sobre alícuotas de muestra convertidas determinaron el número óptimo de ciclos de la PCR antes de la amplificación de la biblioteca final. Condiciones de la PCR final: 50 ul rxn: tampón 10X 5 ul, 1,25 ul de 10 mM de cada dNTP, 5 ul de cóctel de cebadores (~ 5 uM), 15 ul de molde (muestra), 1 ul de PfuTurbo Cx hotstart, 22,75 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 (cantidades equimolares) en bibliotecas tetramultiplexadas según el esquema de aleatorización y se sometieron a ensayo con el bioanalizador para una comprobación de la etapa final, y con qPCR usando patrones phiX y cebadores específicos de adaptadores.

Secuenciación y bioinformática:

Las muestras se cargaron en hileras de celdas de flujo de acuerdo con una asignación aleatorizada de hileras, donde hileras adicionales están reservadas para los controles internos del ensayo. La secuenciación se realizó mediante el Next Generation Sequencing Core en la Medical Genome Facility de Mayo con el Illumina HiSeq 2000. Las lecturas fueron unidireccionales para 101 ciclos. Cada hilera de celda de flujo generó 100-120 millones de lecturas, suficiente para una mediana de cobertura de 30-50 veces de profundidad de secuenciación (el número de lecturas por CpG) para las secuencias alineadas. Se usó el programa informático convencional Standard Illumina para la asignación de bases y generación de la lectura de secuencia en formato fastq. Como se ha descrito anteriormente (Sun Z, Baheti S, Middha S, Kanwar R, Zhang Y, Li X, y col. SAAP-RRBS: streamlined analysis and annotation pipeline for reduced representation bisulfite sequencing. Bioinformatics. 2012;28:2180-1), SAAP-RRBS, un código informático para el análisis y anotación directos para una secuenciación con representación reducida de bisulfito, se usó para en el alineamiento de secuencias y la extracción de la metilación.

Estudios de validación mediante PCR específica de metilación:

Visión general: Se realizaron dos estudios de validación basados en la PCR sobre conjuntos de muestras expandidos para confirmar la precisión y la reproducibilidad de los candidatos diferencialmente metilados observados. El primero, un estudio de *validación interna*, se realizó sobre muestras no emparejadas no enmascaradas usando réplicas biológicas y técnicas de PanC y colon normal, y réplicas técnicas de páncreas normal. Esta etapa se llevó a cabo con el fin de asegurarse de que los sitios de metilación diferencial identificados mediante la filtración de datos RRBS, donde el % de metilación era la unidad del análisis, se reflejaría en la MSP, donde la unidad de análisis es el número absoluto de copias genómicas de la secuencia diana, corregida por la concentración del ADN inicial para cada muestra. El segundo, experimento de *validación externa*, utilizó MSP para estudiar los candidatos principales en muestras aleatoriamente asignadas, emparejadas, enmascaradas independientes de PanC, páncreas benigno, y colon normal.

Diseño de cebadores: Se diseñaron para cada marcador cebadores para dirigirse a las secuencias metiladas modificadas con bisulfito para cada gen diana (IDT, Coralville IA) y una región sin sitios citosina-fofato-guanina en el gen de la β -actina, como referencia del tratamiento con bisulfito y ADN inicial. El diseño se realizó bien con el programa informático Methprimer (University of California, San Francisco CA) o por procedimientos semimanuales (por H.Z y W.R.T). A continuación, los ensayos se analizaron y se optimizaron mediante qPCR con colorantes SYBR Green (Life Technologies, Grand Island NY) sobre diluciones de controles de ADN genómico metilado y no metilado

de forma universal.

PCR específica de metilación: las reacciones MSP se llevaron a cabo sobre ADN extraído de tejido como se ha descrito anteriormente (Kisiel JB, Yab TC, Taylor WR, Chari ST, Petersen GM, Mahoney DW, y col. Stool DNA testing for the detection of pancreatic cancer: assessment of methylation marker candidates. Cancer. 2012;118:2623-31). En resumen, el ADN se trató con bisulfito usando el kit EZ DNA Methylation (Zymo Research, Orange, CA) y se eluyó en tampón. Un μ l de ADN tratado con bisulfito se usó como molde para la cuantificación de metilación con una PCR en tiempo real basada en fluorescencia, realizada con mezcla maestra SYBR Green (Roche, Mannheim Alemania). Las reacciones se realizaron en un Roche 480 LightCyclers (Indianapolis, IN), donde ADN metilado universal CpGenome tratado con bisulfito con bisulfito (Millipore, Billerica, MA) se usó 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 bajo petición.

Análisis estadístico

RRBS: La comparación primaria de interés fue la diferencia de metilación entre los casos y los controles pancreáticos en cada CpG cartografiado. Las islas CpG se definen mediante una relación entre la CpG esperada y la observada que supera 0,6 (Gardiner-Garden M, Frommer M. CpG islands in vertebrate genomes. Journal of molecular biology 1987;196:261-82). Sin embargo, para este modelo, se crearon unidades apiladas de "regiones diferencialmente metiladas (DMR)" según análisis de CpG basándose en la distancia entre ubicaciones de sitios CpG para cada cromosoma. Como la distancia entre cualquier CpG dado supera la ubicación posterior o anterior en más de 100 pb, se creó un nuevo identificador para islas. Las islas con un único CpG se excluyeron. El resultado secundario fue la misma comparación entre casos y controles de colon. Los sitios CpG individuales se consideraron par el análisis diferencial solamente si la profundidad total de la cobertura por grupo de enfermedad ~~ad 20a~~ lecturas (aproximadamente igual a una promedio de 10 lecturas por sujeto) y la varianza en el % de metilación fue mayor que cero (los sitios CpG no informativos con varianza 0 fueron excluidos). Los criterios para la lectura de la profundidad se basaron en la deseada potencia estadística para detectar una diferencia del 10 % en la tasa de metilación entre dos grupos cualesquiera en los que el tamaño de la muestra de individuos para cada grupo era 18.

La significación estadística se determinó mediante regresión logística del porcentaje de metilación por DMR (usando los recuentos actuales) definiéndose los grupos como PanC, páncreas normal y colon normal. Para tener en cuenta las profundidades de lectura variables para los sujetos individuales, se usó un modelo de regresión logística superdisperso, donde el parámetro de dispersión se estimó usando la Chi cuadrada de Pearson estadística de los residuos del modelo ajustado. Para evaluar la metilación específica de la hebra, las regiones directa e inversa se analizaron por separado. Las DMR se ordenaron a continuación según su nivel de significación y se consideraron una región marcadora viable si la tasa de metilación en los controles ~~era~~ % pero ≥ 10 % en PanC. Cada DMR significativa se consideró un marcador candidato.

Para el estudio de validación interna, el resultado primario fue el área bajo la curva de características del operador (ABC) para cada marcador. Esta se calculó usando regresión logística (JMP versión 9.0.1, SAS Institute, Cary NC) para modelar la intensidad el número de copias corregido para la concentración para cada marcador con PanC en comparación con el páncreas normal y colon normal. Los marcadores con los valores de ABC más altos y la relación más estrecha para la mediana del número de copias entre casos y controles se seleccionaron para el estudio de validación externa. El resultado primario del estudio de validación externa fue el ABC para cada marcador representado gráficamente contra la intensidad de la señal de cada marcador, medido según el log de la relación de la mediana corregida para el número de copias en comparación con los controles. Con dieciocho casos, hubo >80 % de potencia para detectar un área bajo la curva de 0,85 o mayor para la hipótesis nula de 0,5 con un nivel de significación bilateral de 0,05. El criterio de valoración secundario fue el ABC de combinaciones de dos marcadores, medidos mediante regresión logística, en los que ambos marcadores fueron necesarios para asociarse independientemente con los casos PanC.

Descubrimiento de marcadores RRBS

Extractos de ADN emparejados, enmascarados, asignados aleatoriamente de 18 tumores de cáncer pancreático, 18 tejidos pancreáticos benignos del control y 18 tejidos epiteliales de colon normal, se secuenciaron por RRBS. La mediana de edad fue 61 (intervalo 52- 65), 61 % eran mujeres, y el 44 % eran fumadores actuales o anteriores. Se capturaron un total de 6.101.049 sitios de CpG en cualquiera de las muestras con al menos 10X de cobertura. Tras seleccionar solamente sitios CpG donde la cobertura de grupos y los criterios de varianza se cumplían, un total de 1.217.523 sitios CpG se consideraron adicionalmente para el análisis. Aproximadamente 500 DMR cumplieron los criterios de significación par la metilación diferencial. Entre estos, los inventores identificaron 107 regiones candidatas con suficientes firmas de metilación para un diseño de cebadores MSP. Las firmas de metilación van de 3 CpG vecinos a 52 CpG. Los niveles de metilación de los cánceres pancreáticos superan en raras ocasiones el 25 % en los CpG filtrados, lo que refleja niveles elevados de células estromales contaminantes. Esto se confirmó después de secuenciar cada uno de los cánceres para mutaciones *KRAS* para comprobar las frecuencias alélicas de las muestras positivas; para el 50 % de los especímenes PanC que tenían un cambio de base *KRAS* heterocigótico, la frecuencia del alelo mutante era al menos 4 veces menor que el correspondiente alelo natural.

Validación interna

Basándose en el número de CpG vecinos en cada firma de metilación del gen candidato, se diseñaron cebadores para 87 de los 107 marcadores candidatos. A continuación se usó MSP para analizar los candidatos en la muestra de ADN procedentes de 20 lesiones PanC adicionales no enmascaradas, 10 muestras de epitelio colónico normal adicionales (réplicas biológicas) así como, muestras de ADN restantes de las 18 lesiones de PanC secuenciadas, 15 de los tejidos pancreáticos benignos secuenciados y 10 de las muestras de colon normal secuenciadas (réplicas técnicas). Con el diseño de cebadores de primer paso, 74 de los 87 marcadores se amplificaron correctamente. Con el rediseño, los restantes 13 cebadores se amplificaron correctamente y se analizaron en 12 muestras de PanC sin enmascarar y 10 muestras de colon normales. La β -actina se amplificó en todas las muestras. Con cualquiera de los MSP de primer paso o segundo paso, 31 de 87 marcadores candidatos tuvieron un valor de ABC > 0,85. Basándose en la magnitud de las diferencias en la mediana del número de copias genómicas entre los casos y los controles para cada marcador candidato, 23 se seleccionaron para la validación externa en muestras independientes. Estos fueron *ABCB1*, *ADCY1*, *BMP3*, *C13ORF18*, *CACNA1C*, *CD1D*, *CHR12:133484978-133485738 (CHR12 133)*, *CLEC11A*, *ELMO1*, *FOXP2*, *GRIN2D*, *IKZF1*, *KCNK12*, *KCNN2*, *NDRG4*, *PRKCB*, *RSPO3*, *SCARF2*, *SHH*, *SLC38A3*, *TWIST1*, *VWC3* y *WT1*.

Validación externa

ADN emparejado, enmascarado, asignado aleatoriamente procedente de 18 muestras PanC, 18 muestras pancreáticas benignas y 36 muestras de epitelio de colon normal se analizaron mediante MSP para los 23 principales candidatos. La mediana de edad de este subconjunto fue 60 (intervalo 54-64). La mayoría (55 %) de las muestras eran de hombres y el 61 % eran fumadores o ex-fumadores. La β -actina se amplificó en todas las muestras. 9 de 23 candidatos mostraron una excelente asociación con PanC. Los valores individuales de ABC para *CACNA1C*, *CHR12.133*, *WT1*, *GRIN2D*, *ELMO1*, *TWIST1*, *C13ORF18*, *KCNN2*, y *CLEC11A* fueron 0,95, 0,95, 0,94, 0,94, 0,93, 0,92, 0,91, 0,90 y 0,90, respectivamente. Se observó buena asociación con otros 9 candidatos; los valores de ABC para *PRKCB*, *CD1D*, *SLC38A3*, *ABCB1*, *KCNK12*, *VWC2*, *RSPO3*, *SHH* y *ADCY1* fueron 0,89, 0,88, 0,86, 0,86, 0,86, 0,85, 0,85, 0,85 y 0,84 respectivamente.

El log de la relación entre la mediana de los valores de los casos y del control para cada marcador se representó gráficamente contra el ABC. Ocho marcadores, *SHH*, *KCNK12*, *PRKCB*, *CLEC11*, *C13ORF18*, *TWIST1*, *ELMO1* y *CHR12.133* tuvieron cada uno un valor de ABC mayor de 0,85, y mostraron un número de copias genómicas mayor 1,5 log (>30 veces) para los casos que para los controles. *KCNK12*, *PRKCB*, *ELMO1* y *CHR12.133* mostraron una diferencia mayor de 2 log (> 100-veces).

Análisis de la complementariedad

Entre las 231 posibles combinaciones de 2 marcadores, ambos marcadores fueron muy significativos en 30 (13 %) modelos emparejados de asociación con PanC. De estos, 18 (8 %) mostraron mejora en el ABC. Notablemente, entre varios marcadores complementarios, *C13ORF18* mejoró la precisión de *CACNA1C*, *WT1*, *GRIN2D*, *SLC38A3* y *SCARF2* con valores del ABC de 0,99, 0,99, 0,97, 0,96, y 0,95, respectivamente, para cada combinación. Considerando el ABC de *SHH* como marcador individual era 0,85, mejoró el rendimiento de otros 6 marcadores cuando se emparejó. El ABC de *CACNA1C*, *WT1*, *SLC38A3*, *ABCB1*, *VWC2* y *RSPO3* mejoró a 0,96, 0,95, 0,92, 0,98, 0,88 y 0,95, respectivamente cuando se combinó en modelos con *SHH*. De las 18 combinaciones de marcadores más sólidas, se pudieron analizar 9 combinaciones de forma emparejada procedentes del conjunto de datos de validación interna. De estas, 7 pares (78 %) siguieron siendo muy significativo en ambos conjuntos de datos.

Ejemplo 9: Marcadores de ADN metilado muy discriminantes para la detección del esófago de Barret

Para disminuir la creciente incidencia de adenocarcinoma esofágico, se necesitan procedimientos eficaces para seleccionar la población para el precursor crítico del esófago de Barrett (EB). Se han propuesto herramientas mínimamente invasivas o no invasivas pero se han visto dificultadas por la falta de marcadores óptimamente sensibles y de marcadores específicos. Se llevaron a cabo experimentos y se identificaron *BMP3* y *NDRG4* metilados de forma anómala como marcadores candidatos discriminantes para el EB.

Un objetivo de dichos experimentos era evaluar de forma prospectiva la precisión de *BMP3* y *NDRG4* metilados para identificar EB usando biopsias endoscópicas (Fase 1) y cepillados del esófago y el cardias completos para simular los dispositivos de muestreo no endoscópico (Fase 2).

Se alistaron los casos con EB y los controles sin EB antes de la endoscopia. Los casos de EB tenían >1 cm de la mucosa columnar perimetral, con metaplasia intestinal confirmada; los controles no tenían EB endoscópicamente. En la Fase 1, se obtuvieron biopsias en los casos de EB, cardias gástrico ((CG); 1 cm por debajo de la línea Z) y epitelio escamoso ((SE); >2 cm por encima de EB) y en controles de CG (como para EB) y SE (5 cm por encima de la línea Z); a continuación se congelaron rápidamente. Las muestras de biopsia se procesaron en lote, y se sometieron a ensayo de forma enmascarada. En la Fase 2, se obtuvieron los especímenes utilizando un cepillo citológico endoscópico de alta capacidad (Hobbs Medical, Stafford Springs CT); el cardias, el EB (en casos), y la longitud completa del esófago se cepillaron para simular un dispositivo de muestreo del tipo de una esponja tragada.

Después de la extracción del ADN y el tratamiento con bisulfito, se evaluó la metilación en los genes diana mediante amplificación de la señal y la diana en tiempo real cuantitativa específica de alelo. β -actina también se cuantificó como un marcador para el ADN humano total.

Se estudiaron prospectivamente 100 sujetos. Fase 1: Entre 40 casos de BE y 40 controles: la mediana de edad fue de 65 (cuartiles 55-77) y 54 (37-69) y los hombres comprendieron 78 % y 48 %, respectivamente. La mediana de la longitud del EB fue de 6 cm (intervalo 3-10). La mediana de los niveles de los marcadores metilados fue sustancialmente superior (34-600 veces) en el EB que en el SE y GC adyacentes, o que en SE y GC normales. A diferencia de los marcadores metilados, las distribuciones de β -actina fueron similares para todos los grupos de tejidos. Ambos niveles de marcadores aumentaron con la longitud de EB y la edad, $p < 0,001$ mientras que solo NDRG4 aumentó significativamente con la presencia de displasia (ninguno (19), grado bajo (9), grado alto (11); $p = 0,003$). Los factores que no afectan significativamente los niveles de marcadores incluyen el sexo y la inflamación. Fase 2: Entre 10 casos de BE y 10 controles, la mediana de edad fue de 64 (59-70) y 66 (49, 71) y los hombres comprendieron el 80 y el 30 % respectivamente. La mediana de la longitud del EB fue de 2 cm (intervalo 1-4). La discriminación de EB por los marcadores fue extraordinaria con la ABC de 1,0 para NDRG4 y 0,99 para BMP3; los niveles fueron >100 veces mayores en los casos respecto de los controles (Figura 2).

Estos experimentos demuestran que marcadores de ADN metilados seleccionados discriminan mucho EB de CG y SE normales, en especímenes de biopsia y cepillados. La Tabla 9 muestra las asociaciones entre la función y la biología del cáncer de los marcadores de ADN metilados seleccionados.

Tabla 8: Niveles de marcadores (números de copias de marcadores ajustados para la beta actina) para las biopsias de BMP3 y NDRG4 en casos de BE (cardias, de Barrett, escamoso) y los controles (cardias, escamoso).

	BMP3		NDRG4	
	Controles normales	Casos de Barrett	Controles normales	Casos de Barrett
Escamoso	0,8	5,6	1,0	4,9
Q1, Q3	0,3, 2,2	0,7, 14,8	0,5, 2,7	1,5, 10,9
P90, P95	7,0, 23,0	25,5, 50,3	5,0, 13,7	32,0, 64,1
EB		300,2		390,6
Q1, Q3		137,1, 659,5		146,6, 763,5
P90, P95		1083,1, 1219,0		921,8, 1006,6
Cardias	0,5	8,2	2,3	11,5
Q1, Q3	0,3, 1,9	2,8, 40,3	1,0, 6,3	5,0, 48,3
P90, P95	10,3, 16,4	190,7, 431,5	13,1, 15,4	116,7, 345,0
Compuesto	1,3	131,4	2,3	136,5
Q1, Q3	0,4, 3,8	67,1, 242,7	1,1, 5,3	68,9, 272,3
P90, P95	10,0, 15,3	402,9, 417,9	8,1, 12,5	344,0, 383,3
Valor P	$< 0,0001$		$< 0,0001$	

Tabla 9. Asociaciones entre la función y la biología del cáncer de los principales marcadores candidatos

DMR	Símbolo	Nombre del gen	Función de la proteína	Asociación con cáncer	Referencia (referencia completa bajo la tabla)
Chr7: 87229775-87229856	ABCB1	Casete de unión a ATP, subfamilia B, miembro 1	Proteína asociada al transportador de membrana	Resistencia multifármaco a quimioterapia	Lee, y col. 2013
Chr7: 45613877-45614564	ADCY1	Adenilato ciclasa 1	Señalización transmembrana	Metilación asociada con el cáncer pancreático	Vincent, y col. 2011
Chr13: 46960770-46961464	C13ORF18	análogo a KIAA0226	Sin caracterizar	Metilación asociada con el cáncer pancreático, neoplasia cervical	Vincent, y col. 2011; Yan, y col. 2009
Chr12:	CACNA1C	Canal del calcio,	Media el la	Metilación	Vincent, y col. 2011

ES 2 712 798 T3

2800665-2800898		dependiente de tensión, Tipo L, subunidad alfa 1C	entrada de iones calcio en la célula	asociada con el cáncer pancreático	
-----------------	--	---	--	--	--

(continuación)

DMR	Símbolo	Nombre del gen	Función de la proteína	Asociación con cáncer	Referencia (referencia completa bajo la tabla)
Chr1: 158150797- 158151142	CD1D	Molécula CD1D	Glicoproteína transmembrana que media la presentación de los antígenos a los linfocitos T	Diana de novedoso tratamientos del cáncer basado en inmunoterapia; expresado por el meduloblastoma	Liu, y col.
Chr19: 51228217-51228703	CLEC11	lectina 11 de tipo C	dominio de lectina de tipo C, sin caracterizar	Ninguno	
Chr12: 133484978- 133485738	(Chr12- 133)	-	Sin caracterizar	-	-
Chr7: 37487539-37488498	ELMO1	Pinocitosis y motilidad celular, 1	Interacción con proteínas de citoquinesis, promoción de motilidad celular y fagocitosis	Promoción de diseminación metastásica	Li, y col.
Chr7: 113727624- 113727693	FOXP2	Secuencia Forkhead P2	Factor de transcripción, expresado en cerebro, pulmón, intestino	Expresado en subconjuntos del cáncer de próstata, linfoma y mieloma múltiple	Stumm, y col., Campbell, y col.
Chr19: 48946755-48946912	GRIN2D	Receptor de glutamato, ionotrópico, N-metil-D-aspartato 2D	receptor de NMDA, neurotransmisión	Metilación asociada con el cáncer pancreático, mutante en el cáncer de mama	Vincent, y col. 2011, Jiao, y col.
Chr7: 50343848-50343927	IKZF1	familia del dedo de cinc IKAROS 1	Proteína de unión al ADN asociada con el remodelado de la cromatina	Mutante en leucemias	Asai, y col.
Chr2: 47797332-47797371	KCNK12	Canal del potasio, subfamilia K, miembro 12	Canal del potasio no funcional	Metilación asociada con el cáncer pancreático y de colon	Vincent, y col. 2011, Kober, y col.
Chr5: 113696984- 113697057	KCNN2	Canal activado por calcio de conductancia intermedia/pequeña del potasio, subfamilia N, miembro 2	Canal del potasio, dependiente de tensión, activado por calcio	Expresado en exceso en cáncer de próstata	Campes, y col.

(continuación)

DMR	Símbolo	Nombre del gen	Función de la proteína	Asociación con cáncer	Referencia (referencia completa bajo la tabla)
Chr16: 584973 95-58497458	NDRG4	N-myc gen regulado de forma defectiva, miembro de la familia 4	Proteína se señalización citosólica necesaria para la progresión del ciclo celular	Metilada en cáncer pancreático y de colon cáncer de colon	Kisiel, y col., Ahlquist, y col.
Chr16: 23846964-23848004	PRKCB	Proteína quinasa C, beta	Quinasa específica de serina y treonina implicada en la señalización celular	Metilación asociada con el cáncer pancreático, diana que se puede convertir en fármaco en el sarcoma de Ewing	Vincent, y col. 2011, Surdez, y col.
Chr6: 127440526-127441039	RSPO3	R-espondina, tipo 3	Proteína reguladora en la ruta de señalización Wnt/ β -catenina	Metilación asociada con el cáncer pancreático, expresión elevada en cánceres de colon	Vincent, y col. 2011, Seshigiri, y col.
Chr22: 20785373-20785464	SCARF2	Secuestrante del receptor de clase F, miembro 2	Media en la unión y la degradación de las lipoproteínas de baja densidad	Metilación asociada con el cáncer pancreático, metilación y expresión reducida en cáncer gástrico	Vincent, y col. 2011, Zhao, y col.
Chr7: 155597771-155597951	SHH	Sonic hedgehog	Embriogénesis	Metilación asociada con el cáncer pancreático, epigenéticamente reprimida en el síndrome MEN1; la señalización hedgehog media la invasión en el cáncer pancreático	Vincent, y col. 2011, Gurung, y col.
Chr3: 50243467 - 50243553	SLC38A3	Transportador de soluto, familia 38, miembro 3	Sin caracterizar	Expresión disminuida en cáncer de pulmón	Person, y col.
Chr7: 19156788-19157093	TWIST1	Factor de transcripción 1 del giro básico hélice-bucle-hélice	Factor de transcripción expresado en tejido placentario y mesodérmico	Metilación asociada con el cáncer pancreático, cáncer biliar, cáncer urotelial	Vincent, y col. 2011, Shin, y col.
Chr7: 49813182-49814168	VWC2	dominio del factor C de von Willebrand factor C que contiene 2	antagonistas de la proteína morfogénica ósea secretada	Metilación asociada con el cáncer pancreático	Vincent, y col. 2011

(continuación)

DMR	Símbolo	Nombre del gen	Función de la proteína	Asociación con cáncer	Referencia (referencia completa bajo la tabla)
Chr11: 32460759-32460800	WT1	Tumor de Wilms 1	Factor de transcripción del motivo en dedo de cinc	Metilación asociada con cánceres pancreático, de próstata, ovario, y mama	Vincent, y col. 2011, Jacobs, y col.

Lee WK, Chakraborty PK, Thevenod F. Pituitary homeobox 2 (PITX2) protects renal cancer cell lines against doxorubicin toxicity by transcriptional activation of the multidrug transporter ABCB1. *International journal of cancer Journal international du cancer*. 2013;133:556-67.

5 Vincent A, Omura N, Hong SM, Jaffe A, Eshleman J, Goggins M. Genome-wide analysis of promoter methylation associated with gene expression profile in pancreatic adenocarcinoma. *Clinical cancer research: an official journal of the American Association for Cancer Research*. 2011;17:4341-54.

10 Yang N, Eijsink JJ, Lendvai A, Volders HH, Klip H, Buikema HJ, y col. Methylation markers for CCNA1 and C13ORF18 are strongly associated with high-grade cervical intraepithelial neoplasia and cervical cancer in cervical scrapings. *Cancer epidemiology, biomarkers & prevention: a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*. 2009;18:3000-7.

Liu D, Song L, Brawley VS, Robison N, Wei J, Gao X, y col. Medulloblastoma expresses CD1d and can be targeted for immunotherapy with NKT cells. *Clin Immunol*. 2013;149:55-64.

15 Li H, Yang L, Fu H, Yan J, Wang Y, Guo H, y col. Association between Galphai2 and ELMO1/Dock180 connects chemokine signalling with Rac activation and metastasis. *Nat Commun*. 2013;4:1706.

Stumm L, Burkhardt L, Steurer S, Simon R, Adam M, Becker A, y col. Strong expression of the neuronal transcription factor FOXP2 is linked to an increased risk of early PSA recurrence in ERG fusion-negative cancers. *Journal of clinical pathology*. 2013;66:563-8.

20 Campbell AJ, Lyne L, Brown PJ, Launchbury RJ, Bignone P, Chi J, y col. Aberrant expression of the neuronal transcription factor FOXP2 in neoplastic plasma cells. *British journal of haematology*. 2010;149:221-30.

Jiao X, Wood LD, Lindman M, Jones S, Buckhaults P, Polyak K, y col. Somatic mutations in the Notch, NF-KB, PIK3CA, and Hedgehog pathways in human breast cancers. *Genes, chromosomes & cancer*. 2012;51:480-9.

Asai D, Imamura T, Suenobu S, Saito A, Hasegawa D, Deguchi T, y col. IKZF1 deletion is associated with a poor outcome in pediatric B-cell precursor acute lymphoblastic leukemia in Japan. *Cancer Med*. 2013;2:412-9.

25 Kober P, Bujko M, Oledzki J, Tysarowski A, Siedlecki JA. Methyl-CpG binding column-based identification of nine genes hypermethylated in colorectal cancer. *Molecular carcinogenesis*. 2011;50:846-56.

Camoses MJ, Paulo P, Ribeiro FR, Barros-Silva JD, Almeida M, Costa VL, y col. Potential downstream target genes of aberrant ETS transcription factors are differentially affected in Ewing's sarcoma and prostate carcinoma. *PLoS ONE*. 2012;7:e49819.

30 Kisiel JB, Yab TC, Taylor WR, Chari ST, Petersen GM, Mahoney DW, y col. Stool DNA testing for the detection of pancreatic cancer: assessment of methylation marker candidates. *Cancer*. 2012;118:2623-31.

Ahlquist DA, Zou H, Domanico M, Mahoney DW, Yab TC, Taylor WR, y col. Next-Generation Stool DNA Test Accurately Detects Colorectal Cancer and Large Adenomas. *Gastroenterology*. 2012;142:248-56.

Surdez D, Benetkiewicz M, Perrin V, Han ZY, Pierron G, Ballet S, y col. Targeting the EWSR1-FLI1 oncogene-induced protein kinase PKC-beta abolishes ewing sarcoma growth. *Cancer research*. 2012;72:4494-503.

35 Seshagiri S, Stawiski EW, Durinck S, Modrusan Z, Storm EE, Conboy CB, y col. Recurrent R-spondin fusions in colon cancer. *Nature*. 2012;488:660-4.

Zhao J, Liang Q, Cheung KF, Kang W, Lung RW, Tong JH, y col. Genome-wide identification of Epstein-Barr virus-driven promoter methylation profiles of human genes in gastric cancer cells. *Cancer*. 2013;119:304-12.

40 Gurung B, Feng Z, Iwamoto DV, Thiel A, Jin G, Fan CM, y col. Menin epigenetically represses Hedgehog signaling in MEN1 tumor syndrome. *Cancer research*. 2013;73:2650-8.

Person RJ, Tokar EJ, Xu Y, Orihuela R, Ngalame NN, Waalkes MP. Chronic cadmium exposure *in vitro* induces cancer cell characteristics in human lung cells. *Toxicol Appl Pharmacol*. 2013.

45 Shin SH, Lee K, Kim BH, Cho NY, Jang JY, Kim YT, y col. Bile-based detection of extrahepatic cholangiocarcinoma with quantitative DNA methylation markers and its high sensitivity. *The Journal of molecular diagnostics: JMD*. 2012;14:256-63.

Jacobs DI, Mao Y, Fu A, Kelly WK, Zhu Y. Dysregulated methylation at imprinted genes in prostate tumor tissue detected by methylation microarray. *BMC Urol*. 2013;13:37.

Ejemplo 10: Un panel de marcadores de microARN y ADN basado en heces para la detección del cáncer pancreático

50 Dada la extraordinaria letalidad del cáncer pancreático (PC), se necesitan procedimientos no invasivos prácticos para la detección por cribado presintomático. Los microARN (miARN) tienen la expresión alterada en el PC.

Se llevaron a cabo experimentos que tenían como objetivo explorar la factibilidad del miR-1290 en heces para la

detección del PC.

Se analizaron muestras de heces del archivo de 58 casos de PC y 64 controles sanos emparejados por la edad, género, y antecedentes de tabaquismo. Se llevó a cabo la detección del miARN siguiendo la estrategia de la reacción en cadena de la polimerasa mediante transcripción inversa cuantitativa de tallo-bucle (qRT-PCR). La cuantificación del miARN se basó en la medición de las copias absolutas por nanogramo del ARN extraído. Los marcadores de ADN (*BMP3* metilado, *KRAS* mutante y β -*actina*) se capturaron en forma híbrida y se amplificaron como se ha descrito (Cancer 2012, 118:2623). Se usó un modelo de regresión logística por etapas, limitado a 5 variables, para desarrollar un panel de marcadores optimizado basado en miR-1290, marcadores de ADN, y edad. Las áreas ajustadas a la edad bajo la curva COR (ABC) para cada uno de los modelos se compararon utilizando los procedimientos de DeLong y col. Se evaluó la asociación de miR-1290 con factores clínicos utilizando la prueba de la suma de rangos de Wilcoxon.

Las distribuciones de miR-1290 fueron significativamente mayores en heces procedentes de casos de PC que de los controles ($P = 0,0002$). Los niveles de miR-1290 en heces no se vieron afectados por la edad, el sexo, el sitio del tumor o la etapa del tumor. La ABC de miR-1290 en heces fue de 0,74 (95 % del IC: 0,65 - 0,82, Figura 3) para la detección de PC en comparación con una ABC de 0,81 (0,73 - 0,89) mediante el panel de marcadores de ADN en heces. La adición de miR-1290 a los marcadores de ADN se ha demostrado creciente ($P = 0,0007$) con una ABC de 0,87 (0,81 - 0,94). Añadir el miR-1290 al panel de ADN aumentó la sensibilidad de la prueba a través del intervalo completo de especificidades incluyendo la región crítica de 90-100 %. La sensibilidad al PC del panel de marcadores combinado fue del 64 % (50 % - 76 %) al 95 % (87 % - 99 %) de especificidad, y del 79 % (67 % - 89 %) al 85 % (74 % - 92 %) de especificidad.

Estos experimentos identificaron el miR-1290 en heces como un marcador para el PC.

Ejemplo 11 - Identificación de marcadores usando RRBS

Durante el desarrollo de la tecnología proporcionada en el presente documento, se recogieron datos de un estudio de control del caso para demostrar que una estrategia de búsqueda en todo el genoma identifica marcadores novedosos e informativos.

Población del estudio, adquisición de especímenes, y muestras

La población diana eran pacientes con cáncer de páncreas tratados en la clínica Mayo. La población accesible incluye los que se han sometido a una pancreatectomía distal, una pancreaticoduodenectomía, o una colectomía con un espécimen de resección archivado, y un diagnóstico patológico confirmado. El ADN del epitelio colónico se había extraído previamente de especímenes microdisecados en el laboratorio de Biospecimens Accessioning Processing (BAP) usando un protocolo de fenol-cloroformo. El personal de Pancreas SPORE usó los datos de las variables emparejadas de estas muestras para seleccionar las muestras de registro de tejidos. Estas fueron revisadas por un patólogo experto para confirmar el caso y controlar el estado y excluir casos de neoplasias producidas a partir de IPMN, que puede tener una biología subyacente distinta. El personal de SPORE dispuso la microdissección en el laboratorio BAP y la extracción del ADN en los casos pancreáticos y muestras del control y proporcionó 500 ng de ADN al personal del laboratorio que estaba enmascarado para el caso y el estado del control. Las muestras de ácido nucleico de archivo incluyeron 18 adenocarcinomas pancreático, 18 páncreas normales y 18 epitelios colónicos emparejados por sexo, edad, y estado de tabaquismo.

Los tipos de muestras fueron:

- 1) tejidos de PanC del registro de Pancreas SPORE en la clínica Mayo limitado a AJCC estado I y II;
- 2) páncreas de control sin PanC;
- 3) epitelio colónico del control archivado sin PanC; y
- 4) neoplasia colónico del que el ADN se había extraído y almacenado en el laboratorio BAP.

Los casos y los controles se emparejaron por sexo, edad (en incrementos de 5 años), y estado de tabaquismo (actual o anterior frente a nunca).

procedimientos

Se prepararon bibliotecas según procedimientos descritos previamente (véase, por ejemplo, Gu y col (2011) "Preparation of reduced representation bisulfite sequencing libraries for genome-scale DNA methylation profiling" Nature Protocols 6: 468-81) mediante la fragmentación de ADN genómico (300 ng) por digestión con 10 unidades de *MspI*, una enzima de restricción específica de metilación que reconoce los motivos que contienen CpG. Este tratamiento enriquece las muestras en su contenido de CpG y elimina zonas redundantes del genoma. Se repararon los extremos de los fragmentos digeridos y se agregaron colas A con 5 unidades de fragmento de Klenow (3'-5' exo) y se ligó durante la noche a adaptadores Illumina que contienen una de cuatro secuencias de código de barras para enlazar cada fragmento a su ID de muestra. Se llevó a cabo la selección por tamaño de los fragmentos de 160-340 pb (que tiene inserciones de 40-220 pb) usando perlas /tampón SPRI (AMPure XP, Beckman Coulter). Los cortes de tampón fueron 0,7x a 1,1 x del volumen de muestra de perlas/tampón. Las muestras se eluyeron en un volumen de

22 µl (tampón EB, Qiagen). Se utilizó la PCR para controlar la eficacia de enlazado y la calidad de fragmentos en una pequeña alícuota de la muestra. A continuación, las muestras se sometieron a dos rondas de conversión con bisulfito usando un protocolo EpiTect modificado (Qiagen). qPCR y PCR convencional (Pfu Turbo Cx hotstart, Agilent), seguido por evaluación con Bioanalyzer 2100 (Agilent) sobre alícuotas de muestra convertidas, determinaron el número óptimo de ciclos de la PCR antes de la amplificación de la biblioteca final. La PCR final se llevó a cabo en un volumen de 50 µl (5 µl de 10x tampón PCR; 1,25 µl de cada dNTP a 10 mM; 5 µl de un primer cóctel a aproximadamente 5 µM, 15 µl de molde (muestra), 1 µl de PfuTurbo Cx hotstart, y 22,75 µl de agua. El ciclado térmico comenzó con incubaciones iniciales a 95 °C durante 5 minutos y a 98 °C durante 30 segundos seguido de 16 ciclos de 98 °C durante 10 segundos, 65 °C durante 30 segundos, y a 72 °C durante 30 segundos. Tras la ciclación, las muestras se incubaron a 72 °C durante 5 minutos y se mantuvieron a 4 °C hasta la posterior elaboración y análisis. Las muestras se combinaron en cantidades equimolares en bibliotecas tetramultiplexadas según un esquema de aleatorización y se sometieron a ensayo con el bioanalizador para una comprobación de la etapa final. Las muestras también se analizaron con qPCR usando patrones phiX y cebadores específicos del adaptador.

Para la secuenciación, las muestras se cargaron en hileras de celdas de flujo de acuerdo con una asignación aleatorizada de hileras, donde hileras adicionales están reservadas para los controles internos del ensayo. La secuenciación se realizó mediante el NGS Core en la Medical Genome Facility de Mayo con el Illumina HiSeq 2000. Las lecturas fueron unidireccionales para 101 ciclos. Cada hilera de celda de flujo generó 100-120 millones de lecturas, suficiente para una mediana de cobertura de 30x a 50x de profundidad de secuenciación (basándose en el número de lecturas por CpG) para las secuencias alineadas. Se usó el programa informático convencional Illumina para analizar las lecturas junto con RRBSMAP (Xi, y col. (2012) "RRBSMAP: a fast, accurate and user-friendly alignment tool for reduced representation bisulfite sequencing" *Bioinformatics* 28: 430-432) y un código informático propio (SAAP-RRBS) desarrollado por el personal de Mayo Biomedical and Statistics (Sun y col. (2012) "SAAP-RRBS: streamlined analysis and annotation pipeline for reduced representation bisulfite sequencing" *Bioinformatics* 28: 2180-1). El análisis bioinformático consistió en 1) evaluación y limpieza de la lectura de la secuencia, 2) alineamiento con el genoma de referencia, 3) extracción del estado de metilación, y 4) notificación y anotación de CpG.

Consideraciones estadísticas:

La comparación primaria de interés es las diferencias en la metilación entre los casos y los controles para cada CpG y/o ventana de CpG apilada. El resultado secundario es la misma comparación entre casos y capa leucocitaria normal y controles de colon. Se estudió la metilación diferencial de los marcadores mediante:

1. Evaluación de las distribuciones del porcentaje de metilación para cada marcador, y descartar los marcadores con más de un 2,5 % de fondo metilado en los controles del colon y en la capa leucocitaria normal
2. Analizar la distribución de la metilación de los marcadores restantes entre casos y controles mediante la prueba de suma de rangos de Wilcoxon y ordenando los marcadores por los valores p.
3. Usar los valores Q para estimar las tasas de descubrimiento falsas (FDR) (Benjamini y col. (1995) "Multiple Testing" *Journal of the Royal Statistical Society. Serie B (Metodológica)* 57: 289-300; Storey y col. (2003) "Statistical significance for genomewide studies" *Proc Natl Acad Sci U S A* 100: 9440-5). Durante el descubrimiento, es aceptable una FDR de hasta el 25 %.

Análisis de datos

Se desarrolló un código de análisis de datos en el paquete informático de análisis de software R ("R: A Language and Environment for Statistical Computing" (2012), R Foundation for Statistical Computing). El flujo de trabajo comprendía las siguientes etapas:

1. Lectura de todos los sitios CpG
2. Considerar solo aquellos sitios CpG donde la profundidad de cobertura total del grupo es de 200 lecturas o más. Esto se basa en una evaluación de potencia para detectar una diferencia entre el 20 % y el 30 % en la metilación entre dos grupos cualesquiera; cualquier cosa menor tiene pocas posibilidades de significación. Por lo tanto, si hay 18 sujetos por cada grupo y cada sujeto tiene 12 lecturas, la profundidad de cobertura del grupo es $12 \times 18 = 216$.
3. Excluir todos los sitios CpG donde la varianza del % metilación a través de los grupos sea 0 (sitios CpG no informativos).
4. Realizar una regresión logística sobredispersa para el % de metilación (basándose en el recuento real) usando los grupos definidos como Colon/Capa leucocitaria normal, control específico de la enfermedad, y cáncer específico de interés (casos) para determinar la significación estadística del % de metilación para los análisis primario y secundario. Se usó un modelo logístico sobredisperso porque la variabilidad en el % de metilación entre sujetos es mayor que el permitido por la hipótesis binómica. Este parámetro de dispersión se estimó usando la Chi cuadrada de Pearson del ajuste.
5. Generar los valores del área bajo la curva de características operativas del receptor (COR). El área bajo la curva COR es una medida de la precisión predictiva del % de metilación específica del sujeto y se estimó del a partir del análisis primario (casos vs. control de a enfermedad) y el análisis secundario (casos vs. colon/capa

leucocitaria normal), independientemente.

6. De una manera similar al punto 5, también se estimaron las veces de cambio (FC, una medida de la separación entre casos y controles) para el análisis primario y secundario usando la relación del % de metilación entre los casos y el correspondiente grupo del control.

7. 4-6 anterior se realizó en sitios CpG individuales así como en regiones CpG metiladas. Estas regiones se definieron para cada cromosoma como un grupo de al menos 5 sitios CpG con una distancia de aproximadamente 100 pares de bases (pb) con un % de metilación media < 2,5 % en controles de colon/capa leucocitaria normales

8. Las regiones CpG que muestran promesa para una validación técnica y biológica se identificaron por tener una diferencia en la metilación estadísticamente significativa, un FC grande, y un valor de ABC elevado para el análisis primario y secundario.

Análisis post-R:

1. Regiones CpGs y CpG individuales clasificadas por el valor p, FC, y ABC. Los cortes fueron <0,01, >20, y >0,85 respectivamente, aunque estos se ajustaron a menudo dependiendo de la solidez de los datos. Por ejemplo, el tejido neoplásico muy heterogéneo dio como resultado valores del % de metilación inferiores, que a su vez afectan a la filtración. Las comparaciones primarias y secundarias pueden clasificarse juntas o por separado dependiendo de los requerimientos de especificidad de la aplicación. Los epitelios colónicos normales se incluyeron como un control para los marcadores sin cobertura adecuados para el ensayo de las heces. Si se está ensayando el jugo pancreático, el tejido colónico es innecesario. Esto puede dar como resultado un conjunto completamente diferente de marcadores.

2. Regiones de marcadores clasificadas basadas en los requerimientos de la plataforma de ensayo. En la actualidad, la PCR específica de la metilación (MSP), o plataformas de discriminación similares donde la discriminación está basada en la especificidad de la hibridación del cebador, es la plataforma de elección. Para esta metodología, es imperativo tener 2-5 regiones CpG discriminadas por oligonucleótido en un tramo amplificable del ADN. Para los ensayos de heces, este requerimiento es incluso más restrictivo puesto que los amplicones deben ser cortos (<100pb). La selección de marcadores, por lo tanto, debe hacerse sobre la base de tramos contiguos cortos de CpG muy discriminadas. Si la plataforma evoluciona a una tecnología basada en secuencias, los requisitos de distribución de CpG en una región pueden ser completamente diferentes.

Resultados

Extractos de ADN emparejados, enmascarados, asignados aleatoriamente de 18 tumores de cáncer pancreático, 18 tejidos pancreáticos benignos del control y 18 tejidos epiteliales colónicos normales, se secuenciaron por RRBS. La mediana de edad fue 61 (intervalo 52- 65), 61 % eran mujeres, y el 44 % eran fumadores actuales o anteriores. Se cartografiaron aproximadamente 6 millones de CpG con una cobertura ≥10x. Más de 2000 regiones CpG cumplieron los criterios de significación por la metilación diferencial. Después de aplicar los criterios de filtro anteriores, se identificaron 449 regiones metiladas de forma diferencial (DMR) (Tabla 10). La Tabla 11 presenta las 449 regiones diferencialmente metiladas (DMR) identificadas ordenadas por área decreciente bajo la curva COR (ABC).

En estos marcadores, las firmas de metilación van de 3 CpG a 56 CpG. Los niveles de metilación de los cánceres pancreáticos superan en raras ocasiones el 25 % en los CpG filtrados, lo que sugiere que los tejidos con cáncer pueden tener elevados niveles de células y/o estroma normal contaminante. Para estudiar esto, cada uno de los cánceres se secuenció para determinar las mutaciones KRAS para verificar las frecuencias alélicas de las muestras positivas. Para el 50 % que tenía un cambio de base KRAS heterocigótico, la frecuencia del alelo mutante era al menos 4 veces menor que el correspondiente alelo natural, lo que respalda la contaminación por células y/o estroma normales.

Se descubrió que 58 de los 449 marcadores estaban en regiones no anotadas y se encuentran en regiones sin elementos de codificación proteica. De los restantes 391 marcadores candidatos, aproximadamente 225 se habían descrito como asociados con cáncer, algunos de los cuales se clasifican como supresores tumorales. Los otros 166 marcadores candidatos tenían anteriormente una baja asociación con el cáncer (por ejemplo, mutaciones y/o alteraciones en el número de copias observado en cribados de todo el genoma) o no tenían asociaciones con cáncer previamente identificadas.

Tabla 10: DMR

N.º marcador	Cromosoma	Coordenadas del cromosoma	Anotación
1	chr7	87229775-87229856	ABCB1
2	chr2	207307687-207307794	ADAM23
3	chr15	100881373-100881437	ADAMTS17
4	chr16	77468655-77468742	ADAMTS18
5	chr19	41224781-41225006	ADCK4
6	chr7	45613877-45614572	ADCY1

ES 2 712 798 T3

7	chr2	70994498-70994755	ADD2
8	chr14	105190863-105191031	ADSSL1
9	chr10	116064516-116064600	AFAP1L2

ES 2 712 798 T3

(continuación)

N.º marcador	Cromosoma	Coordenadas del cromosoma	Anotación
10	chr4	87934353-87934488	AFF1
11	chr2	100720494-100720679	AFF3
12	chr7	100136884-100137350	AGFG2
13	chr9	116151083-116151315	ALAD
14	chr14	103396870-103396920	AMN
15	chr19	10206736-10206757	ANGPTL6
16	chr19	17438929-17438974	ANO8
17	chr15	90358267-90358400	ANPEP
18	chr15	29131299-29131369	APBA2
19	chr19	45430362-45430458	APOC1P1
20	chr13	111767862-111768355	ARHGEF7
21	chr7	98990897-98990989	ARPC1B
22	chr22	51066374-51066431	ARSA
23	chr9	120175665-120176057	ASTN2
24	chr1	203619509-203619829	ATP2B4
25	chr7	69062853-69062972	AUTS2
26	chr8	104152963-104152974	BAALC
27	chr11	64052053-64052132	BAD
28	chr10	121411207-121411375	BAG3
29	chr7	98029116-98029383	BAIAP2L1
30	chr9	135462730-135462765	BARHL1
31	chr10	133795124-133795423	BNIP3
32	chr12	107715014-107715095	BTBD11
33	chr6	105584524-105584800	BVES
34	chr10	21816267-21816490	C10orf140
35	chr12	21680381-21680438	C12orf39
36	chr12	21680681-21680817	C12orf39
37	chr12	117174873-117175030	C12orf49
38	chr13	46960767-46961669	C13orf18
39	chr14	50099743-50099930	C14orf104
40	chr19	16772631-16772712	C19orf42
41	chr20	31061389-31061649	C20orf112
42	chr5	175665232-175665311	C5orf25
43	chr6	42858890-42859092	C6orf226
44	chr9	139735581-139735683	C9orf172
45	chr12	2800756-2800899	CACNA1C
46	chr3	54156904-54156987	CACNA2D3
47	chr11	115373179-115373281	CADM1
48	chr16	89007413-89007432	CBFA2T3
49	chr16	49316205-49316258	CBLN1
50	chr21	44495919-44495933	CBS
51	chr17	77810085-77810206	CBX4
52	chr17	8649567-8649665	CCDC42
53	chr11	64110001-64110069	CCDC88B
54	chr14	91883473-91883674	CCDC88C
55	chr14	99946756-99946806	CCNK
56	chr1	158150797-158151205	CD1D
57	chr5	175969660-175969699	CDHR2
58	chr7	39989959-39990020	CDK13

ES 2 712 798 T3

(continuación)

N.º marcador	Cromosoma	Coordenadas del cromosoma	Anotación
59	chr16	80837397-80837505	CDYL2
60	chr10	11059508-11060151	CELF2
61	chr22	47130339-47130459	CERK
62	chr2	233389020-233389049	CHRNA
63	chr7	73245708-73245798	CLDN4
64	chr19	51228217-51228732	CLEC11A
65	chr3	139654045-139654132	CLSTN2
66	chr7	155302557-155302639	CNPY1
67	chr6	88875699-88875763	CNR1
68	chr6	88876367-88876445	CNR1
69	chr6	88876701-88876726	CNR1
70	chr2	165698520-165698578	COBLL1
71	chr6	75794978-75795024	COL12A1
72	chr12	48398051-48398093	COL2A1
73	chr12	48398306-48398375	COL2A1
74	chr18	449695-449798	COLEC12
75	chr7	30721980-30722020	CRHR2
76	chr16	84875643-84875772	CRISPLD2
77	chr7	151127086-151127195	CRYGN
78	chr10	126812450-126812653	CTBP2
79	chr20	56089440-56089547	CTCFL
80	chr2	219261190-219261327	CTDSP1
81	chr2	80530326-80530374	CTNNA2
82	chr22	43044555-43044737	CYB5R3
83	chr19	1406516-1406625	DAZAP1
84	chr7	44084171-44084235	DBNL
85	chr11	20178177-20178304	DBX1
86	chr4	151000325-151000356	DCLK2
87	chr4	151000358-151000403	DCLK2
88	chr4	183817058-183817157	DCTD
89	chr13	52378159-52378202	DHRS12
90	chr8	13014567-13014682	DLC1
91	chr11	84432067-84432186	DLG2
92	chr6	170598276-170598782	DLL1
93	chr19	39989824-39989852	DLL3
94	chr19	12996198-12996321	DNASE2
95	chr2	230578698-230578802	DNER
96	chr2	225907414-225907537	DOCK10
97	chr18	32073971-32074004	DTNA
98	chr2	233352345-233352605	ECEL1
99	chr7	37487539-37488596	ELMO1
100	chr20	39995010-39995051	EMILIN3
101	chr19	48833763-48833967	EMP3
102	chr2	119607676-119607765	EN1
103	chr3	27763358-27763617	EOMES
104	chr3	27763909-27763981	EOMES
105	chr12	132435207-132435428	EP400
106	chr19	16473958-16474095	EPS15L1
107	chr6	152129293-152129450	ESR1
108	chr3	185825887-185826002	ETV5

ES 2 712 798 T3

(continuación)

N.º marcador	Cromosoma	Coordenadas del cromosoma	Anotación
109	chr9	140201493-140201583	EXD3
110	chr6	133562127-133562229	EYA4
111	chr1	160983607-160983768	F11R
112	chr20	821836-821871	FAM110A
113	chr22	45898798-45898888	FBLN1
114	chr9	97401449-97401602	FBP1
115	chr16	750679-750715	FBXL16
116	chr5	15500208-15500399	FBXL7
117	chr5	15500663-15500852	FBXL7
118	chr5	114880375-114880442	FEM1C
119	chr20	34189488-34189693	FER1L4
120	chr14	53417493-53417618	FERMT2
121	chr2	219849962-219850042	FEV
122	chr17	7339280-7339492	FGF11
123	chr19	49256413-49256451	FGF21
124	chr10	103538848-103539033	FGF8
125	chr11	64008415-64008495	FKBP2
126	chr11	128564106-128564209	FLI1
127	chr10	102985059-102985130	FLJ41350
128	chr13	28674451-28674629	FLT3
129	chr1	240255240-240255264	FMN2
130	chr5	131132146-131132232	FNIP1
131	chr6	108882636-108882682	FOXO3
132	chr3	71478053-71478206	FOXP1
133	chr7	113724864-113725006	FOXP2
134	chr7	113727624-113727693	FOXP2
135	chr5	160975098-160975142	GABRB2
136	chr12	51786085-51786218	GALNT6
137	chr5	179780839-179780955	GFPT2
138	chr20	3641457-3641537	GFRA4
139	chr17	4462834-4463034	GGT6
140	chr17	4463796-4464037	GGT6
141	chr17	42907549-42907807	GJC1
142	chr8	144358251-144358266	GL14
143	chr16	4377510-4377615	GLIS2
144	chr12	56881329-56881414	GLS2
145	chr6	24776486-24776667	GMNN
146	chr19	3095019-3095055	GNA11
147	chr22	19710910-19710984	GP1BB
148	chr22	19711364-19711385	GP1BB
149	chr2	131485151-131485219	GPR148
150	chr2	165477564-165477609	GRB14
151	chr2	165477839-165477886	GRB14
152	chr17	73390467-73390597	GRB2
153	chr19	48918266-48918311	GRIN2D
154	chr19	48946755-48946912	GRIN2D
155	chr13	114018369-114018421	GRTP1
156	chr12	13254503-13254606	GSG1
157	chr7	43152309-43152375	HECW1
158	chr7	139440133-139440341	HIPK2

ES 2 712 798 T3

(continuación)

N.º marcador	Cromosoma	Coordenadas del cromosoma	Anotación
159	chr6	34205664-34206018	HMGA1
160	chr12	121416542-121416670	HNF1A
161	chr20	42984244-42984427	HNF4A
162	chr20	43040031-43040119	HNF4A
163	chr5	177632203-177632260	HNRNPAB
164	chr7	27136030-27136245	HOXA1
165	chr2	176971915-176971968	HOXD11
166	chr19	35540057-35540200	HPN
167	chr2	163174366-163174659	IFIH1
168	chr17	47073421-47073440	IGF2BP1
169	chr11	133797643-133797789	IGSF9B
170	chr7	50343838-50344029	IKZF1
171	chr7	50344414-50344453	IKZF1
172	chr20	20345123-20345150	INSM1
173	chr20	20350520-20350532	INSM1
174	chr15	76632356-76632462	ISL2
175	chr2	182321880-182322022	ITGA4
176	chr2	182322168-182322198	ITGA4
177	chr2	173293542-173293644	ITGA6
178	chr19	2097386-2097437	IZUMO4
179	chr21	27011846-27011964	JAM2
180	chr2	47797260-47797371	KCNK12
181	chr10	79397895-79397945	KCNMA1
182	chr5	113696524-113696682	KCNN2
183	chr5	113696971-113697058	KCNN2
184	chr1	154733071-154733232	KCNN3
185	chr8	99439457-99439482	KCNS2
186	chr19	34287890-34287972	KCTD15
187	chr12	121905558-121905792	KDM2B
188	chr8	136469529-136469873	KHDRBS3
189	chr16	85646495-85646594	KIAA0182
190	chr18	46190841-46190970	KIAA0427
191	chr4	37245694-37245718	KIAA1239
192	chr17	72350351-72350403	KIF19
193	chr2	149633039-149633137	KIF5C
194	chr22	50987245-50987312	KLHDC7B
195	chr12	53298237-53298384	KRT8
196	chr19	54974004-54974086	LENG9
197	chr1	180198528-180198542	LHX4
198	chr19	2290471-2290541	LINGO3
199	chr11	19733958-19734013	LOC100126784
200	chr19	58513829-58513851	LOC100128398
201	chr17	43324999-43325188	LOC100133991
202	chr17	43325784-43325960	LOC100133991
203	chr2	109745715-109745742	LOC100287216
204	chr1	178063099-178063167	LOC100302401
205	chr12	53447992-53448072	LOC283335
206	chr1	45769962-45770141	LOC400752
207	chr20	61637950-61638000	LOC63930
208	chr13	88323571-88323647	LOC642345

ES 2 712 798 T3

(continuación)

N.º marcador	Cromosoma	Coordenadas del cromosoma	Anotación
209	chr6	111873064-111873162	LOC643749
210	chr5	87956937-87956996	LOC645323
211	chr5	87970260-87970568	LOC645323
212	chr5	87970751-87970850	LOC645323
213	chr12	85430135-85430175	LRRIQ1
214	chr19	497878-497933	MADCAM1
215	chr5	71404528-71404563	MAP1B
216	chr2	39665069-39665282	MAP4K3
217	chr1	156406057-156406118	MAX.chr1.156406057-156406118
218	chr1	23894874-23894919	MAX.chr1.23894874-23894919
219	chr1	240161479-240161546	MAX.chr1.240161479-240161546
220	chr1	244012804-244012986	MAX.chr1.244012804-244012986
221	chr1	35394690-35394876	MAX.chr1.35394690-35394876
222	chr1	35395179-35395201	MAX.chr1.35395179-35395201
223	chr1	39044345-39044354	MAX.chr1.39044345-39044354
224	chr10	101282185-101282257	MAX.chr10.101282185-101282257
225	chr10	127033272-127033428	MAX.chr10.127033272-127033428
226	chr11	120382450-120382498	MAX.chr11.120382450-120382498
227	chr11	47421719-47421776	MAX.chr11.47421719-47421776
228	chr12	133484978-133485066	MAX.chr12.133484978-133485066
229	chr12	133485702-133485739	MAX.chr12.133485702-133485739
230	chr12	54151078-54151153	MAX.chr12.54151078-54151153
231	chr12	58259413-58259475	MAX.chr12.58259413-58259475
232	chr13	25322044-25322165	MAX.chr13.25322044-25322165
233	chr13	29394692-29394771	MAX.chr13.29394692-29394771
234	chr14	100751586-100751695	MAX.chr14.100751586-100751695
235	chr14	61123624-61123707	MAX.chr14.61123624-61123707
236	chr14	89507100-89507162	MAX.chr14.89507100-89507162
237	chr15	40361431-40361644	MAX.chr15.40361431-40361644
238	chr15	89942904-89943197	MAX.chr15.89942904-89943197
239	chr16	25042924-25043187	MAX.chr16.25042924-25043187
240	chr16	85230248-85230405	MAX.chr16.85230248-85230405
241	chr17	1835463-1835690	MAX.chr17.1835463-1835690
242	chr17	60218266-60218449	MAX.chr17.60218266-60218449
243	chr17	76337726-76337824	MAX.chr17.76337726-76337824
244	chr19	11805543-11805639	MAX.chr19.11805543-11805639
245	chr19	22034747-22034887	MAX.chr19.22034747-22034887
246	chr19	32715650-32715707	MAX.chr19.32715650-32715707
247	chr19	5805881-5805968	MAX.chr19.5805881-5805968
248	chr2	127783183-127783233	MAX.chr2.127783183-127783233
249	chr2	232530964-232531124	MAX.chr2.232530964-232531124
250	chr2	239957125-239957163	MAX.chr2.239957125-239957163
251	chr2	43153331-43153424	MAX.chr2.43153331-43153424
252	chr2	71503632-71503860	MAX.chr2.71503632-71503860
253	chr20	43948422-43948484	MAX.chr20.43948422-43948484
254	chr21	47063798-47063877	MAX.chr21.47063798-47063877
255	chr22	17849540-17849622	MAX.chr22.17849540-17849622
256	chr22	38732124-38732211	MAX.chr22.38732124-38732211
257	chr22	42764974-42765049	MAX.chr22.42764974-42765049
258	chr22	46974925-46975007	MAX.chr22.46974925-46975007

ES 2 712 798 T3

(continuación)

N.º marcador	Cromosoma	Coordenadas del cromosoma	Anotación
259	chr22	50342922-50343232	MAX.chr22.50342922-50343232
260	chr3	132273353-132273532	MAX.chr3.132273353-132273532
261	chr3	193858771-193858843	MAX.chr3.193858771-193858843
262	chr3	24563009-24563117	MAX.chr3.24563009-24563117
263	chr3	75411368-75411473	MAX.chr3.75411368-75411473
264	chr4	26828422-26828522	MAX.chr4.26828422-26828522
265	chr4	8965831-8965868	MAX.chr4.8965831-8965868
266	chr5	142100518-142100780	MAX.chr5.142100518-142100780
267	chr6	169613138-169613249	MAX.chr6.169613138-169613249
268	chr6	64168133-64168268	MAX.chr6.64168133-64168268
269	chr7	129794337-129794536	MAX.chr7.129794337-129794536
270	chr7	1705957-1706065	MAX.chr7.1705957-1706065
271	chr7	28893550-28893569	MAX.chr7.28893550-28893569
272	chr7	47650711-47650882	MAX.chr7.47650711-47650882
273	chr7	64408106-64408135	MAX.chr7.64408106-64408135
274	chr9	108418404-108418453	MAX.chr9.108418404-108418453
275	chr9	120507310-120507354	MAX.chr9.120507310-120507354
276	chr5	89769002-89769411	MBLAC2
277	chr12	51319165-51319319	METTTL7A
278	chr2	191272534-191272765	MFSD6
279	chr19	6236947-6237089	MLLT1
280	chr6	168333306-168333467	MLLT4
281	chr8	89339567-89339662	MMP16
282	chr17	2300399-2300476	MNT
283	chr7	156802460-156802490	MNX1
284	chr19	4343896-4242968	MPND
285	chr16	56715756-56716025	MT1X
286	chr15	48470062-48470503	MYEF2
287	chr15	48470606-48470725	MYEF2
288	chr5	16936010-16936058	MYO10
289	chr3	39851068-39851989	MYRIP
290	chr13	33001061-33001251	N4BP2L1
291	chr4	2060477-2060624	NAT8L
292	chr12	125002129-125002192	NCOR2
293	chr16	23607524-23607650	NDUFAB1
294	chr10	105338596-105338843	NEURL
295	chr1	204797773-204797785	NFASC
296	chr2	233877877-233878027	NGEF
297	chr18	31803017-31803114	NOL4
298	chr9	139438534-139438629	NOTCH1
299	chr5	32714270-32714325	NPR3
300	chr9	127266951-127267032	NR5A1
301	chr11	124615979-124616029	NRGN
302	chr11	124616860-124617005	NRGN
303	chr20	327754-327871	NRSN2
304	chr8	99952501-99952533	OSR2
305	chr5	76926598-76926703	OTP
306	chr3	8809858-8809865	OXTR
307	chr19	14172823-14172948	PALM3
308	chr6	52268531-52268702	PAQR8

ES 2 712 798 T3

(continuación)

N.º marcador	Cromosoma	Coordenadas del cromosoma	Anotación
309	chr20	21686466-21686563	PAX1
310	chr21	47063793-47064177	PCBP3
311	chr7	100203461-100203600	PCOLCE
312	chr4	657555-657666	PDE6B
313	chr7	544848-545022	PDGFA
314	chr2	239194812-239194946	PER2
315	chr19	43979400-43979435	PHLDB3
316	chr6	144384503-144385539	PLAGL1
317	chr2	28844174-28844270	PLB1
318	chr1	242687719-242687746	PLD5
319	chr12	6419210-6419489	PLEKHG6
320	chr22	50745629-50745727	PLXNB2
321	chr2	105471752-105471787	POU3F3
322	chr13	79177868-79177951	POU4F1
323	chr1	203044913-203044929	PPFIA4
324	chr22	50825886-50825981	PPP6R2
325	chr17	74519328-74519457	PRCD
326	chr7	601162-601552	PRKAR1B
327	chr16	23846964-23847339	PRKCB
328	chr16	23847507-23847617	PRKCB
329	chr16	23847825-23848168	PRKCB
330	chr22	18923785-18923823	PRODH
331	chr22	45099093-45099304	PRR5
332	chr3	9988302-9988499	PRRT3
333	chr1	11538685-11538738	PTCHD2
334	chr1	11539396-11539540	PTCHD2
335	chr10	23480864-23480913	PTF1A
336	chr19	5340273-5340743	PTPRS
337	chr2	1747034-1747126	PXDN
338	chr2	1748338-1748444	PXDN
339	chr7	4923056-4923107	RADIL
340	chr19	15568448-15568639	RASAL3
341	chr5	80256215-80256313	RASGRF2
342	chr17	77179784-77179887	RBFOX3
343	chr4	40516823-40516984	RBM47
344	chr4	57775698-57775771	REST
345	chr10	43572798-43572896	RET
346	chr10	121302439-121302501	RGS10
347	chr16	318717-318893	RGS11
348	chr1	241520322-241520334	RGS7
349	chr1	42846119-42846174	RIMKLA
350	chr21	43189031-43189229	RIPK4
351	chr7	5821188-5821283	RNF216
352	chr19	23941063-23941142	RPSAP58
353	chr19	23941384-23941670	RPSAP58
354	chr16	29118636-29118891	RRN3P2
355	chr6	127440492-127441039	RSPO3
356	chr17	42392669-42392701	RUNDC3A
357	chr6	45345446-45345595	RUNX2
358	chr6	45387405-45387456	RUNX2

ES 2 712 798 T3

(continuación)

N.º marcador	Cromosoma	Coordenadas del cromosoma	Anotación
359	chr3	72496092-72496361	RYBP
360	chr22	20785373-20785464	SCARF2
361	chr8	145561664-145561696	SCRT1
362	chr7	54826636-54826706	SEC61G
363	chr10	38691448-38691521	SEPT7L
364	chr4	154712157-154712232	SFRP2
365	chr7	155597793-155597973	SHH
366	chr4	77610781-77610824	SHROOM3
367	chr21	38120336-38120558	SIM2
368	chr15	68115602-68115675	SKOR1
369	chr17	6949717-6949778	SLC16A11
370	chr11	35441199-35441260	SLC1A2
371	chr19	59025337-59025385	SLC27A5
372	chr2	27486089-27486170	SLC30A3
373	chr12	69140018-69140206	SLC35E3
374	chr12	46661132-46661306	SLC38A1
375	chr3	50243467-50243553	SLC38A3
376	chr7	150760388-150760530	SLC4A2
377	chr5	1445384-1445473	SLC6A3
378	chr2	40679298-40679326	SLC8A1
379	chr5	506178-506343	SLC9A3
380	chr20	61284095-61284194	SLCO4A1
381	chr5	101631546-101631731	SLCO4C1
382	chr10	98945242-98945493	SLIT1
383	chr13	88330094-88330355	SLITRK5
384	chr15	66999854-67000014	SMAD6
385	chr10	112064230-112064280	SMNDC1
386	chr6	84419007-84419072	SNAP91
387	chr17	36508733-36508891	SOCS7
388	chr4	7367687-7367825	SORCS2
389	chr17	70116754-70116823	SOX9
390	chr4	57687746-57687764	SPINK2
391	chr3	140770014-140770193	SPSB4
392	chr17	36762706-36762763	SRCIN1
393	chr6	43141954-43142058	SRF
394	chr7	105029460-105029585	SRPK2
395	chr16	70415312-70415673	ST3GAL2
396	chr2	107502978-107503055	ST6GAL2
397	chr2	107503155-107503391	ST6GAL2
398	chr12	22487528-22487848	ST8SIA1
399	chr10	17496177-17496310	ST8SIA6
400	chr2	242447608-242447724	STK25
401	chr3	120626999-120627116	STXBP5L
402	chr3	33260338-33260423	SUSD5
403	chr16	19179713-19179744	SYT17
404	chr12	115122614-115122632	TBX3
405	chr19	3606372-3606418	TBXA2R
406	chr10	70359250-70359439	TET1
407	chr16	4310204-4310233	TFAP4
408	chr21	32930371-32930409	TIAM1

ES 2 712 798 T3

(continuación)

N.º marcador	Cromosoma	Coordenadas del cromosoma	Anotación
409	chr4	942190-942382	TMEM175
410	chr6	130686773-130686820	TMEM200a
411	chr6	130687200-130687735	TMEM200a
412	chr3	185215700-185215782	TMEM41A
413	chr20	42544780-42544835	TOX2
414	chr9	140091343-140091644	TPRN
415	chr8	126441476-126441519	TRIB1
416	chr5	14143759-14143880	TRIO
417	chr22	38148620-38148716	TRIOBP
418	chr7	19156788-19157227	TWIST1
419	chr7	19157436-19157533	TWIST1
420	chr4	41259387-41259594	UCHL1
421	chr15	63795401-63795636	USP3
422	chr17	9548120-9548325	USP43
423	chr12	95942077-95942558	USP44
424	chr10	17271896-17271994	VIM
425	chr7	49813135-49814168	VWC2
426	chr7	151078646-151078674	WDR86
427	chr12	49372205-49372274	WNT1
428	chr11	32460759-32460800	WT1
429	chr19	4061206-4061360	ZBTB7A
430	chr8	144623045-144623088	ZC3H3
431	chr2	145274698-145274874	ZEB2
432	chr19	38146299-38146397	ZFP30
433	chr16	88521287-88521377	ZFPM1
434	chr4	2298384-2298498	ZFYVE28
435	chr4	2415252-2415286	ZFYVE28
436	chr20	45986341-45986684	ZMYND8
437	chr22	22862957-22862983	ZNF280B
438	chr6	43336449-43336545	ZNF318
439	chr19	53661819-53662279	ZNF347
440	chr16	88497041-88497148	ZNF469
441	chr19	57019064-57019137	ZNF471
442	chr19	2842178-2842235	ZNF555
443	chr19	37958078-37958134	ZNF570
444	chr8	125985552-125985847	ZNF572
445	chr19	53696101-53696195	ZNF665
446	chr19	53696497-53696704	ZNF665
447	chr19	20149796-20149923	ZNF682
448	chr19	57106617-57106967	ZNF71
449	chr7	6655380-6655652	ZNF853

En la Tabla 10, las bases se numeraron según el conjunto del genoma humano, de febrero de 2009, GRCh37/hg19 (véase, por ejemplo, Rosenbloom y col. (2012) "ENCODE whole-genome data in the UCSC Genome Browser: actualización 2012" *Nucleic Acids Research* 40: D912-D917). Los marcadores con nombres BHLHE23 y LOC63930 se refieren al mismo marcador.

5

Tabla 11.

Cromosoma	Coordenadas del cromosoma	Anotación	Área bajo la curva COR
chr12	53298237-53298384	KRT8	1,00

ES 2 712 798 T3

(continuación)			
Cromosoma	Coordenadas del cromosoma	Anotación	Área bajo la curva COR
chr7	129794337-129794536	MAX.chr7.129794337-129794536	1,00
chr10	101282185-101282257	MAX.chr10.101282185-101282257	0,99
chr10	126812450-126812653	CTBP2	0,99
chr9	116151083-116151315	ALAD	0,99
chr8	13014567-13014682	DLC1	0,99
chr7	139440133-139440341	HIPK2	0,99
chr3	39851068-39851989	MYRIP	0,99
chr19	4061206-4061360	ZBTB7A	0,99
chr16	84875643-84875772	CRISPLD2	0,99
chr6	52268531-52268702	PAQR8	0,99
chr2	239194812-239194946	PER2	0,99
chr17	1835463-1835690	MAX.chr17.1835463-1835690	0,99
chr5	506178-506343	SLC9A3	0,99
chr20	31061389-31061649	C20orf112	0,98
chr9	139438534-139438629	NOTCH1	0,98
chr15	48470606-48470725	MYEF2	0,98
chr12	125002129-125002192	NCOR2	0,98
chr4	7367687-7367825	SORCS2	0,98
chr19	6236947-6237089	MLLT1	0,98
chr7	544848-545022	PDGFA	0,98
chr7	98029116-98029383	BAIAP2L1	0,98
chr4	2415252-2415286	ZFYVE28	0,98
chr12	6419210-6419489	PLEKHG6	0,98
chr22	50825886-50825981	PPP6R2	0,97
chr20	45986341-45986684	ZMYND8	0,97
chr5	142100518-142100780	MAX.chr5.142100518-142100780	0,97
chr19	16473958-16474095	EPS15L1	0,97
chr16	29118636-29118891	RRN3P2	0,97
chr6	75794978-75795024	COL12A1	0,97
chr9	139735581-139735683	C9orf172	0,97
chr17	4462834-4463034	GGT6	0,97
chr17	4463796-4464037	GGT6	0,96
chr12	95942077-95942558	USP44	0,96
chr20	42984244-42984427	HNF4A	0,96
chr7	47650711-47650882	MAX.chr7.47650711-47650882	0,96
chr4	942190-942382	TMEM175	0,96
chr7	73245708-73245798	CLDN4	0,96
chr22	46974925-46975007	MAX.chr22.46974925-46975007	0,96
chr10	127033272-127033428	MAX.chr10.127033272-127033428	0,96
chr3	132273353-132273532	MAX.chr3.132273353-132273532	0,96
chr4	26828422-26828522	MAX.chr4.26828422-26828522	0,96
chr20	61284095-61284194	SLCO4A1	0,96
chr19	35540057-35540200	HPN	0,96
chr22	45099093-45099304	PRR5	0,95
chr17	60218266-60218449	MAX.chr17.60218266-60218449	0,95
chr6	168333306-168333467	MLLT4	0,95
chr10	105338596-105338843	NEURL	0,95
chr9	120175665-120176057	ASTN2	0,95

ES 2 712 798 T3

Cromosoma	Coordenadas del cromosoma	(continuación)	Área bajo la curva COR
		Anotación	
chr4	183817058-183817157	DCTD	0,95
chr6	108882636-108882682	FOXO3	0,95
chr7	27136030-27136245	HOXA1	0,95
chr19	14172823-14172948	PALM3	0,95
chr3	75411368-75411473	MAX.chr3.75411368-75411473	0,94
chr6	64168133-64168268	MAX.chr6.64168133-64168268	0,94
chr16	318717-318893	RGS11	0,94
chr20	43040031-43040119	HNF4A	0,94
chr7	49813135-49814168	VWC2	0,94
chr16	85230248-85230405	MAX.chr16.85230248-85230405	0,94
chr22	38148620-38148716	TRIOBP	0,94
chr5	89769002-89769411	MBLAC2	0,94
chr1	158150797-158151205	CD1D	0,93
chr19	1406516-1406625	DAZAP1	0,93
chr12	121416542-121416670	HNF1A	0,93
chr17	76337726-76337824	MAX.chr17.76337726-76337824	0,93
chr13	88330094-88330355	SLITRK5	0,93
chr19	54974004-54974086	LENG9	0,93
chr22	47130339-47130459	CERK	0,92
chr7	601162-601552	PRKAR1B	0,92
chr2	70994498-70994755	ADD2	0,92
chr15	40361431-40361644	MAX.chr15.40361431-40361644	0,92
chr19	15568448-15568639	RASAL3	0,92
chr6	24776486-24776667	GMNN	0,92
chr18	449695-449798	COLEC12	0,92
chr7	150760388-150760530	SLC4A2	0,92
chr21	38120336-38120558	SIM2	0,91
chr15	66999854-67000014	SMAD6	0,91
chr2	28844174-28844270	PLB1	0,91
chr11	115373179-115373281	CADM1	0,91
chr21	47063793-47064177	PCBP3	0,91
chr2	1748338-1748444	PXDN	0,91
chr21	47063798-47063877	MAX.chr21.47063798-47063877	0,91
chr16	56715756-56716025	MT1X	0,90
chr4	87934353-87934488	AFF1	0,90
chr9	140091343-140091644	TPRN	0,90
chr5	15500208-15500399	FBXL7	0,90
chr19	48833763-48833967	EMP3	0,90
chr6	43141954-43142058	SRF	0,90
chr3	185215700-185215782	TMEM41A	0,90
chr1	160983607-160983768	F11R	0,90
chr12	58259413-58259475	MAX.chr12.58259413-58259475	0,90
chr2	47797260-47797371	KCNK12	0,89
chr16	4377510-4377615	GLIS2	0,89
chr15	63795401-63795636	USP3	0,89
chr13	33001061-33001251	N4BP2L1	0,89
chr3	120626999-120627116	STXBP5L	0,89
chr7	19156788-19157227	TWIST1	0,89
chr18	46190841-46190970	KIAA0427	0,89

ES 2 712 798 T3

Cromosoma	Coordenadas del cromosoma	(continuación)	Área bajo la curva COR
		Anotación	
chr7	100203461-100203600	PCOLCE	0,88
chr19	51228217-51228732	CLEC11A	0,88
chr19	17438929-17438974	ANO8	0,88
chr12	2800756-2800899	CACNA1C	0,88
chr6	34205664-34206018	HMGA1	0,88
chr15	76632356-76632462	ISL2	0,88
chr6	111873064-111873162	LOC643749	0,88
chr10	70359250-70359439	TET1	0,88
chr2	39665069-39665282	MAP4K3	0,88
chr2	43153331-43153424	MAX.chr2.43153331-43153424	0,87
chr22	17849540-17849622	MAX.chr22.17849540-17849622	0,87
chr2	233877877-233878027	NGEF	0,87
chr8	89339567-89339662	MMP16	0,87
chr13	46960767-46961669	C13orf18	0,87
chr6	170598276-170598782	DLL1	0,87
chr4	40516823-40516984	RBM47	0,87
chr3	139654045-139654132	CLSTN2	0,87
chr2	27486089-27486170	SLC30A3	0,87
chr17	74519328-74519457	PRCD	0,86
chr2	163174366-163174659	IFIH1	0,86
chr4	41259387-41259594	UCHL1	0,86
chr7	45613877-45614572	ADCY1	0,86
chr7	98990897-98990989	ARPC1B	0,86
chr3	54156904-54156987	CACNA2D3	0,86
chr16	49316205-49316258	CBLN1	0,86
chr3	71478053-71478206	FOXP1	0,86
chr5	87956937-87956996	LOC645323	0,86
chr21	43189031-43189229	RIPK4	0,86
chr12	22487528-22487848	ST8SIA1	0,86
chr20	42544780-42544835	TOX2	0,86
chr20	821836-821871	FAM110A	0,86
chr16	4310204-4310233	TFAP4	0,86
chr11	64110001-64110069	CCDC88B	0,85
chr8	136469529-136469873	KHDRBS3	0,85
chr10	102985059-102985130	FLJ41350	0,85
chr2	176971915-176971968	HOXD11	0,85
chr12	51319165-51319319	METTL7A	0,85
chr22	50342922-50343232	MAX.chr22.50342922-50343232	0,85
chr7	155597793-155597973	SHH	0,85
chr4	154712157-154712232	SFRP2	0,84
chr19	57019064-57019137	ZNF471	0,84
chr5	87970260-87970568	LOC645323	0,84
chr6	130686773-130686820	TMEM200a	0,84
chr9	140201493-140201583	EXD3	0,84
chr12	53447992-53448072	LOC283335	0,84
chr22	43044555-43044737	CYB5R3	0,84
chr19	49256413-49256451	FGF21	0,84
chr17	77810085-77810206	CBX4	0,84
chr7	156802460-156802490	MNX1	0,84

ES 2 712 798 T3

Cromosoma	Coordenadas del cromosoma	(continuación)	Área bajo la curva COR
		Anotación	
chr7	151127086-151127195	CRYGN	0,83
chr6	169613138-169613249	MAX.chr6.169613138-169613249	0,83
chr2	71503632-71503860	MAX.chr2.71503632-71503860	0,83
chr20	21686466-21686563	PAX1	0,83
chr2	173293542-173293644	ITGA6	0,83
chr7	87229775-87229856	ABCB1	0,83
chr2	207307687-207307794	ADAM23	0,83
chr12	21680381-21680438	C12orf39	0,83
chr15	89942904-89943197	MAX.chr15.89942904-89943197	0,83
chr10	43572798-43572896	RET	0,83
chr19	5805881-5805968	MAX.chr19.5805881-5805968	0,83
chr19	53661819-53662279	ZNF347	0,83
chr22	38732124-38732211	MAX.chr22.38732124-38732211	0,83
chr11	124615979-124616029	NRGN	0,83
chr2	100720494-100720679	AFF3	0,83
chr19	497878-497933	MADCAM1	0,82
chr5	14143759-14143880	TRIO	0,82
chr18	32073971-32074004	DTNA	0,82
chr15	48470062-48470503	MYEF2	0,82
chr3	50243467-50243553	SLC38A3	0,82
chr16	70415312-70415673	ST3GAL2	0,82
chr11	35441199-35441260	SLC1A2	0,82
chr12	51786085-51786218	GALNT6	0,82
chr2	232530964-232531124	MAX.chr2.232530964-232531124	0,81
chr22	19710910-19710984	GP1BB	0,81
chr19	2097386-2097437	IZUMO4	0,81
chr11	20178177-20178304	DBX1	0,81
chr7	37487539-37488596	ELMO1	0,81
chr11	128564106-128564209	FLI1	0,81
chr7	105029460-105029585	SRPK2	0,81
chr10	103538848-103539033	FGF8	0,81
chr11	124616860-124617005	NRGN	0,81
chr19	57106617-57106967	ZNF71	0,81
chr9	97401449-97401602	FBP1	0,81
chr5	113696971-113697058	KCNN2	0,80
chr19	53696497-53696704	ZNF665	0,80
chr1	45769962-45770141	LOC400752	0,80
chr14	91883473-91883674	CCDC88C	0,80
chr17	43324999-43325188	LOC100133991	0,80
chr16	23846964-23847339	PRKCB	0,80
chr19	11805543-11805639	MAX.chr19.11805543-11805639	0,80
chr12	117174873-117175030	C12orf49	0,80
chr20	39995010-39995051	EMILIN3	0,80
chr5	87970751-87970850	LOC645323	0,80
chr7	4923056-4923107	RADIL	0,80
chr19	23941063-23941142	RPSAP58	0,80
chr6	45387405-45387456	RUNX2	0,80
chr17	6949717-6949778	SLC16A11	0,80
chr2	165477564-165477609	GRB14	0,80

ES 2 712 798 T3

Cromosoma	Coordenadas del cromosoma	(continuación)	Área bajo la curva COR
		Anotación	
chr20	34189488-34189693	FER1L4	0,80
chr22	50745629-50745727	PLXNB2	0,79
chr7	155302557-155302639	CNPY1	0,79
chr7	19157436-19157533	TWIST1	0,79
chr1	203619509-203619829	ATP2B4	0,79
chr2	230578698-230578802	DNER	0,79
chr19	23941384-23941670	RPSAP58	0,79
chr17	73390467-73390597	GRB2	0,79
chr15	68115602-68115675	SKOR1	0,79
chr17	2300399-2300476	MNT	0,79
chr13	79177868-79177951	POU4F1	0,79
chr19	59025337-59025385	SLC27A5	0,79
chr9	135462730-135462765	BARHL1	0,78
chr8	125985552-125985847	ZNF572	0,78
chr5	175665232-175665311	C5orf25	0,78
chr6	42858890-42859092	C6orf226	0,78
chr12	21680681-21680817	C12orf39	0,78
chr14	50099743-50099930	C14orf104	0,78
chr5	175969660-175969699	CDHR2	0,78
chr16	80837397-80837505	CDYL2	0,78
chr19	12996198-12996321	DNASE2	0,78
chr13	28674451-28674629	FLT3	0,78
chr1	154733071-154733232	KCNN3	0,78
chr1	35395179-35395201	MAX.chr1.35395179-35395201	0,78
chr19	5340273-5340743	PTPRS	0,78
chr3	33260338-33260423	SUSD5	0,78
chr2	145274698-145274874	ZEB2	0,78
chr13	25322044-25322165	MAX.chr13.25322044-25322165	0,78
chr2	80530326-80530374	CTNNA2	0,78
chr12	56881329-56881414	GLS2	0,78
chr3	24563009-24563117	MAX.chr3.24563009-24563117	0,78
chr7	6655380-6655652	ZNF853	0,78
chr4	2298384-2298498	ZFYVE28	0,77
chr5	177632203-177632260	HNRNPAB	0,77
chr22	19711364-19711385	GP1BB	0,77
chr2	165477839-165477886	GRB14	0,77
chr13	29394692-29394771	MAX.chr13.29394692-29394771	0,77
chr14	103396870-103396920	AMN	0,77
chr12	132435207-132435428	EP400	0,77
chr8	99439457-99439482	KCNS2	0,77
chr7	5821188-5821283	RNF216	0,77
chr17	9548120-9548325	USP43	0,77
chr3	185825887-185826002	ETV5	0,77
chr12	121905558-121905792	KDM2B	0,77
chr3	193858771-193858843	MAX.chr3.193858771-193858843	0,77
chr19	53696101-53696195	ZNF665	0,77
chr7	69062853-69062972	AUTS2	0,77
chr1	242687719-242687746	PLD5	0,76
chr20	43948422-43948484	MAX.chr20.43948422-43948484	0,76

ES 2 712 798 T3

Cromosoma	Coordenadas del cromosoma	(continuación)	Área bajo la curva COR
		Anotación	
chr6	84419007-84419072	SNAP91	0,76
chr17	43325784-43325960	LOC100133991	0,76
chr19	41224781-41225006	ADCK4	0,76
chr5	15500663-15500852	FBXL7	0,76
chr20	20350520-20350532	INSM1	0,76
chr1	23894874-23894919	MAX.chr1.23894874-23894919	0,76
chr1	11538685-11538738	PTCHD2	0,76
chr14	105190863-105191031	ADSSL1	0,76
chr22	22862957-22862983	ZNF280B	0,76
chr17	72350351-72350403	KIF19	0,76
chr7	50343838-50344029	IKZF1	0,76
chr2	191272534-191272765	MFSD6	0,76
chr17	47073421-47073440	IGF2BP1	0,76
chr10	133795124-133795423	BNIP3	0,75
chr5	101631546-101631731	SLCO4C1	0,75
chr12	133485702-133485739	MAX.chr12.133485702-133485739	0,75
chr22	18923785-18923823	PRODH	0,75
chr20	56089440-56089547	CTCFL	0,75
chr6	43336449-43336545	ZNF318	0,75
chr14	61123624-61123707	MAX.chr14.61123624-61123707	0,75
chr7	30721980-30722020	CRHR2	0,75
chr17	7339280-7339492	FGF11	0,75
chr11	84432067-84432186	DLG2	0,75
chr2	233352345-233352605	ECEL1	0,75
chr3	27763358-27763617	EOMES	0,75
chr5	160975098-160975142	GABRB2	0,75
chr1	244012804-244012986	MAX.chr1.244012804-244012986	0,75
chr16	25042924-25043187	MAX.chr16.25042924-25043187	0,75
chr4	57775698-57775771	REST	0,75
chr6	127440492-127441039	RSPO3	0,75
chr8	145561664-145561696	SCRT1	0,75
chr8	144623045-144623088	ZC3H3	0,75
chr12	48398051-48398093	COL2A1	0,75
chr2	182321880-182322022	ITGA4	0,75
chr9	120507310-120507354	MAX.chr9.120507310-120507354	0,74
chr6	133562127-133562229	EYA4	0,74
chr2	127783183-127783233	MAX.chr2.127783183-127783233	0,74
chr11	47421719-47421776	MAX.chr11.47421719-47421776	0,74
chr19	10206736-10206757	ANGPTL6	0,74
chr2	225907414-225907537	DOCK10	0,74
chr1	35394690-35394876	MAX.chr1.35394690-35394876	0,74
chr4	2060477-2060624	NAT8L	0,74
chr2	1747034-1747126	PXDN	0,74
chr6	45345446-45345595	RUNX2	0,74
chr7	50344414-50344453	IKZF1	0,74
chr1	180198528-180198542	LHX4	0,74
chr14	53417493-53417618	FERMT2	0,74
chr17	77179784-77179887	RBFOX3	0,74
chr10	98945242-98945493	SLIT1	0,74

ES 2 712 798 T3

Cromosoma	Coordenadas del cromosoma	(continuación)	
		Anotación	Área bajo la curva COR
chr2	40679298-40679326	SLC8A1	0,74
chr12	48398306-48398375	COL2A1	0,74
chr22	50987245-50987312	KLHDC7B	0,73
chr12	54151078-54151153	MAX.chr12.54151078-54151153	0,73
chr7	28893550-28893569	MAX.chr7.28893550-28893569	0,73
chr10	38691448-38691521	SEPT7L	0,73
chr1	203044913-203044929	PPFIA4	0,73
chr22	51066374-51066431	ARSA	0,73
chr7	113724864-113725006	FOXP2	0,73
chr12	13254503-13254606	GSG1	0,73
chr11	19733958-19734013	LOC100126784	0,73
chr1	39044345-39044354	MAX.chr1.39044345-39044354	0,73
chr3	9988302-9988499	PRRT3	0,73
chr22	20785373-20785464	SCARF2	0,73
chr6	130687200-130687735	TMEM200a	0,73
chr12	46661132-46661306	SLC38A1	0,73
chr19	20149796-20149923	ZNF682	0,73
chr11	133797643-133797789	IGSF9B	0,73
chr2	105471752-105471787	POU3F3	0,72
chr5	179780839-179780955	GFPT2	0,72
chr8	99952501-99952533	OSR2	0,72
chr19	16772631-16772712	C19orf42	0,72
chr2	119607676-119607765	EN1	0,72
chr12	49372205-49372274	WNT1	0,72
chr5	113696524-113696682	KCNN2	0,72
chr17	8649567-8649665	CCDC42	0,72
chr7	1705957-1706065	MAX.chr7.1705957-1706065	0,71
chr2	149633039-149633137	KIF5C	0,71
chr19	2842178-2842235	ZNF555	0,71
chr10	121302439-121302501	RGS10	0,71
chr21	44495919-44495933	CBS	0,71
chr10	11059508-11060151	CELF2	0,71
chr19	48946755-48946912	GRIN2D	0,71
chr12	133484978-133485066	MAX.chr12.133484978-133485066	0,71
chr5	16936010-16936058	MYO10	0,71
chr17	42392669-42392701	RUNDC3A	0,71
chr16	88521287-88521377	ZFPM1	0,71
chr4	37245694-37245718	KIAA1239	0,71
chr16	23847507-23847617	PRKCB	0,71
chr5	76926598-76926703	OTP	0,71
chr18	31803017-31803114	NOL4	0,71
chr2	182322168-182322198	ITGA4	0,70
chr15	90358267-90358400	ANPEP	0,70
chr12	107715014-107715095	BTBD11	0,70
chr16	89007413-89007432	CBFA2T3	0,70
chr4	151000325-151000356	DCLK2	0,70
chr6	152129293-152129450	ESR1	0,70
chr19	38146299-38146397	ZFP30	0,70
chr1	204797773-204797785	NFASC	0,70

ES 2 712 798 T3

Cromosoma	Coordenadas del cromosoma	(continuación)	Área bajo la curva COR
		Anotación	
chr22	42764974-42765049	MAX.chr22.42764974-42765049	0,70
chr2	165698520-165698578	COBLL1	0,70
chr8	144358251-144358266	GL14	0,70
chr2	219261190-219261327	CTDSP1	0,70
chr2	239957125-239957163	MAX.chr2.239957125-239957163	0,70
chr10	121411207-121411375	BAG3	0,69
chr2	233389020-233389049	CHRND	0,69
chr14	99946756-99946806	CCNK	0,69
chr11	120382450-120382498	MAX.chr11.120382450-120382498	0,69
chr16	750679-750715	FBXL16	0,69
chr15	100881373-100881437	ADAMTS17	0,69
chr1	11539396-11539540	PTCHD2	0,69
chr2	242447608-242447724	STK25	0,69
chr16	23847825-23848168	PRKCB	0,69
chr17	42907549-42907807	GJC1	0,69
chr19	48918266-48918311	GRIN2D	0,69
chr10	79397895-79397945	KCNMA1	0,69
chr5	71404528-71404563	MAP1B	0,69
chr19	43979400-43979435	PHLDB3	0,69
chr17	70116754-70116823	SOX9	0,69
chr16	88497041-88497148	ZNF469	0,69
chr2	131485151-131485219	GPR148	0,69
chr8	126441476-126441519	TRIB1	0,68
chr4	151000358-151000403	DCLK2	0,68
chr19	39989824-39989852	DLL3	0,68
chr14	89507100-89507162	MAX.chr14.89507100-89507162	0,68
chr12	115122614-115122632	TBX3	0,68
chr19	58513829-58513851	LOC100128398	0,68
chr5	32714270-32714325	NPR3	0,68
chr3	140770014-140770193	SPSB4	0,68
chr6	88875699-88875763	CNR1	0,68
chr4	657555-657666	PDE6B	0,68
chr16	19179713-19179744	SYT17	0,67
chr3	8809858-8809865	OXTR	0,67
chr10	116064516-116064600	AFAP1L2	0,67
chr4	77610781-77610824	SHROOM3	0,67
chr6	88876367-88876445	CNR1	0,67
chr7	151078646-151078674	WDR86	0,67
chr2	109745715-109745742	LOC100287216	0,67
chr14	100751586-100751695	MAX.chr14.100751586-100751695	0,67
chr21	32930371-32930409	TIAM1	0,67
chr4	57687746-57687764	SPINK2	0,67
chr2	219849962-219850042	FEV	0,66
chr20	327754-327871	NRSN2	0,66
chr1	178063099-178063167	LOC100302401	0,66
chr19	45430362-45430458	APOC1P1	0,66
chr13	111767862-111768355	ARHGEF7	0,66
chr19	37958078-37958134	ZNF570	0,66
chr19	32715650-32715707	MAX.chr19.32715650-32715707	0,66

ES 2 712 798 T3

Cromosoma	Coordenadas del cromosoma	(continuación)	
		Anotación	Área bajo la curva COR
chr8	104152963-104152974	BAALC	0,66
chr19	3095019-3095055	GNA11	0,66
chr19	3606372-3606418	TBXA2R	0,66
chr12	69140018-69140206	SLC35E3	0,66
chr4	8965831-8965868	MAX.chr4.8965831-8965868	0,66
chr17	36508733-36508891	SOCS7	0,66
chr16	85646495-85646594	KIAA0182	0,65
chr7	54826636-54826706	SEC61G	0,65
chr9	108418404-108418453	MAX.chr9.108418404-108418453	0,65
chr7	64408106-64408135	MAX.chr7.64408106-64408135	0,65
chr10	21816267-21816490	C10orf140	0,65
chr7	39989959-39990020	CDK13	0,65
chr1	240255240-240255264	FMN2	0,65
chr13	114018369-114018421	GRTP1	0,65
chr13	88323571-88323647	LOC642345	0,65
chr5	80256215-80256313	RASGRF2	0,65
chr10	112064230-112064280	SMNDC1	0,65
chr12	85430135-85430175	LRRIQ1	0,65
chr1	241520322-241520334	RGS7	0,65
chr19	22034747-22034887	MAX.chr1 9.22034747-22034887	0,65
chr21	27011846-27011964	JAM2	0,65
chr11	64052053-64052132	BAD	0,65
chr1	42846119-42846174	RIMKLA	0,64
chr10	17271896-17271994	VIM	0,64
chr13	52378159-52378202	DHRS12	0,63
chr3	27763909-27763981	EOMES	0,63
chr7	100136884-100137350	AGFG2	0,62
chr6	88876701-88876726	CNR1	0,62
chr19	2290471-2290541	LINGO3	0,62
chr6	105584524-105584800	BVES	0,61
chr16	23607524-23607650	NDUFAB1	0,61
chr11	64008415-64008495	FKBP2	0,60
chr20	3641457-3641537	GFRA4	0,59
chr19	4343896-4242968	MPND	0,59
chr2	107503155-107503391	ST6GAL2	0,59
chr1	240161479-240161546	MAXchr1.240161479-240161546	0,57
chr6	144384503-144385539	PLAGL1	0,57
chr3	72496092-72496361	RYBP	0,57
chr5	131132146-131132232	FNIP1	0,55
chr17	36762706-36762763	SRCIN1	0,55
chr11	32460759-32460800	WT1	0,55
chr9	127266951-127267032	NR5A1	0,53
chr7	44084171-44084235	DBNL	0,46
chr15	29131299-29131369	APBA2	0,44
chr5	114880375-114880442	FEM1C	0,44
chr19	34287890-34287972	KCTD15	0,44
chr16	77468655-77468742	ADAMTS18	
chr22	45898798-45898888	FBLN1	
chr7	113727624-113727693	FOXP2	

Cromosoma	Coordenadas del cromosoma	(continuación) Anotación	Área bajo la curva COR
chr7	43152309-43152375	HECW1	
chr20	20345123-20345150	INSM1	
chr20	61637950-61638000	LOC63930	
chr1	156406057-156406118	MAXchr1. 156406057-156406118	
chr10	23480864-23480913	PTF1A	
chr5	1445384-1445473	SLC6A3	
chr2	107502978-107503055	ST6GAL2	
chr10	17496177-17496310	ST8SIA6	

5 Varias modificaciones y variaciones de las composiciones, procedimientos, y usos de la tecnología descritos serán evidentes para el experto en la materia sin separarse del espíritu y del ámbito de la tecnología como se ha descrito. Aunque la tecnología se ha descrito con respecto a realizaciones específicas ilustrativas, se deberá entender que la invención tal como se reivindica no deberá quedar indebidamente limitada a dichas realizaciones específicas. De hecho, está previsto que las diferentes modificaciones de los modos descritos para realizar la Invención que son evidentes para los expertos en los campos de la farmacología, bioquímica, ciencia médica, o campos relacionados, estén incluidas en el ámbito de las siguientes reivindicaciones.

LISTADO DE SECUENCIAS

- <110> Mayo Foundation for Medical Education and Research
- 10 <120> DETECCIÓN DE NEOPLASIAS
- <130> EXCTM-32992/WO-2/ORD
- <150> US 61/784.429
- <151> 14/03/2013
- <160> 202
- 15 <170> PatentIn versión 3.5
- <210> 1
- <211> 21
- <212> ADN
- <213> Secuencia artificial
- 20 <220>
- <223> Sintética
- <400> 1
- gattttgttc gtcgtagtg c 21
- <210> 2
- 25 <211> 19
- <212> ADN
- <213> Secuencia artificial
- <220>
- <223> Sintética
- 30 <400> 2
- tctctaaacc cgccaacga 19
- <210> 3
- <211> 25
- <212> ADN
- 35 <213> Secuencia artificial
- <220>

<223> Sintética
 <400> 3
 ttcgaagttt cgggatagga agcgt 25
 5
 <210> 4
 <211> 20
 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Sintética
 10
 <400> 4
 cctaccgacc ttcgaacgcg 20
 <210> 5
 <211> 21
 <212> ADN
 15
 <213> Secuencia artificial
 <220>
 <223> Sintética
 <400> 5
 ggcggcgcgt attttttcg c 21
 20
 <210> 6
 <211> 23
 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Sintética
 25
 <400> 6
 cgctacgata taaacgacga cga 23
 <210> 7
 <211> 19
 <212> ADN
 <213> Secuencia artificial
 30
 <220>
 <223> Sintética
 <400> 7
 ggttcggtg tcgtagcgc 19
 35
 <210> 8
 <211> 20
 <212> ADN
 <213> Secuencia artificial
 40
 <220>
 <223> Sintética
 <400> 8
 ccgaccgtaa tcctcgacga 20
 45
 <210> 9
 <211> 25
 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Sintética
 50
 <400> 9

	ttaggtcggg aatcgttatt gtttc	25
	<210> 10 <211> 22 <212> ADN <213> Secuencia artificial	
5	<220> <223> Sintética	
	<400> 10 gtaaataacc ccgcgctaaa cg	22
10	<210> 11 <211> 24 <212> ADN <213> Secuencia artificial	
15	<220> <223> Sintética	
	<400> 11 ttcgtttgtt ttccgggtcg tagc	24
20	<210> 12 <211> 24 <212> ADN <213> Secuencia artificial	
	<220> <223> Sintética	
25	<400> 12 accacgtaac gatttactcg acga	24
	<210> 13 <211> 23 <212> ADN <213> Secuencia artificial	
30	<220> <223> Sintética	
	<400> 13 cgttttcgga ttgaagtcg ttc	23
35	<210> 14 <211> 19 <212> ADN <213> Secuencia artificial	
	<220> <223> Sintética	
40	<400> 14 cgcctcgtct tccaacgaa	19
	<210> 15 <211> 19 <212> ADN <213> Secuencia artificial	
45	<220> <223> Sintética	
	<400> 15 agggcgttcg gtttagtc	19
50	<210> 16	

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

5

<220>
 <223> Sintética

<400> 16
 aaccgaaaac gacaaaatcg at 22

10

<210> 17
 <211> 21
 <212> ADN
 <213> Secuencia artificial

<220>
 <223> Sintética

15

<400> 17
 ttgagcggc ggtcgtgat c 21

<210> 18
 <211> 22
 <212> ADN
 <213> Secuencia artificial

20

<220>
 <223> Sintética

<400> 18
 tccccgaatc taaacgctac ga 22

25

<210> 19
 <211> 25
 <212> ADN
 <213> Secuencia artificial

<220>
 <223> Sintética

30

<400> 19
 tttaggaag taaagcgtcg ttttc 25

<210> 20
 <211> 22
 <212> ADN
 <213> Secuencia artificial

35

<220>
 <223> Sintética

<400> 20
 aacgacgtct cgatacctac ga 22

40

<210> 21
 <211> 22
 <212> ADN
 <213> Secuencia artificial

<220>
 <223> Sintética

45

<400> 21
 ggagagtatt toggttttc gc 22

<210> 22
 <211> 24
 <212> ADN
 <213> Secuencia artificial

50

<220>
 <223> Sintética

 <400> 22
 acaaacaaaa tcgaaaaaca cccg 24

 5 <210> 23
 <211> 23
 <212> ADN
 <213> Secuencia artificial

 <220>
 <223> Sintética

 10 <400> 23
 gtttcggtt cggtcgaggt tac 23

 <210> 24
 <211> 25
 <212> ADN
 <213> Secuencia artificial

 15 <220>
 <223> Sintética

 <400> 24
 gccattaact cgataaaaa cgcg 25

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

 20 <220>
 <223> Sintética

 <400> 25
 gatttaatcg tagattcggg tcgtc 25

 <210> 26
 <211> 22
 <212> ADN
 <213> Secuencia artificial

 25 <220>
 <223> Sintética

 <400> 26
 ccgaaacgaa cgaactcaaa cg 22

 <210> 27
 <211> 17
 <212> ADN
 <213> Secuencia artificial

 30 <220>
 <223> Sintética

 <400> 27
 gcgcgtagcg gcgttcc 17

 <210> 28
 <211> 19
 <212> ADN
 <213> Secuencia artificial

 35 <220>
 <223> Sintética

 <400> 27
 gcgcgtagcg gcgttcc 17

 <210> 28
 <211> 19
 <212> ADN
 <213> Secuencia artificial

 40 <220>
 <223> Sintética

 <400> 27
 gcgcgtagcg gcgttcc 17

 <210> 28
 <211> 19
 <212> ADN
 <213> Secuencia artificial

 45 <220>
 <223> Sintética

 <400> 27
 gcgcgtagcg gcgttcc 17

 <210> 28
 <211> 19
 <212> ADN
 <213> Secuencia artificial

 50 <220>
 <223> Sintética

	<400> 28 cccatatcgc cgcacgtaa	19
5	<210> 29 <211> 22 <212> ADN <213> Secuencia artificial	
	<220> <223> Sintética	
10	<400> 29 tcgtatttgg cgttcgtag tc	22
	<210> 30 <211> 21 <212> ADN <213> Secuencia artificial	
15	<220> <223> Sintética	
	<400> 30 cgaaatccaa cgccgaaacg a	21
20	<210> 31 <211> 24 <212> ADN <213> Secuencia artificial	
	<220> <223> Sintética	
25	<400> 31 ttgtcgttcg tcgaattcga ttc	24
	<210> 32 <211> 23 <212> ADN <213> Secuencia artificial	
30	<220> <223> Sintética	
	<400> 32 aaccgcgacgc taaaaaacga cga	23
35	<210> 33 <211> 25 <212> ADN <213> Secuencia artificial	
40	<220> <223> Sintética	
	<400> 33 ttgcgttggc tacgttttt tacgc	25
45	<210> 34 <211> 23 <212> ADN <213> Secuencia artificial	
	<220> <223> Sintética	
50	<400> 34 acgccgtacg aataacgaaa cga	23

<210> 35
 <211> 21
 <212> ADN
 <213> Secuencia artificial
 5 <220>
 <223> Sintética
 <400> 35
 cgtttttcgg gtcgggttcg c 21
 10 <210> 36
 <211> 19
 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Sintética
 15 <400> 36
 tccgacgctc gactccga 19
 <210> 37
 <211> 22
 <212> ADN
 20 <213> Secuencia artificial
 <220>
 <223> Sintética
 <400> 37
 tcggcgtatt ttctagac gc 22
 25 <210> 38
 <211> 24
 <212> ADN
 <213> Secuencia artificial
 <220>
 30 <223> Sintética
 <400> 38
 cgcaatctta aacgtacgct tcga 24
 <210> 39
 <211> 24
 <212> ADN
 35 <213> Secuencia artificial
 <220>
 <223> Sintética
 <400> 39
 40 ggtttataaa gagtcggtt tcgc 24
 <210> 40
 <211> 24
 <212> ADN
 <213> Secuencia artificial
 45 <220>
 <223> Sintética
 <400> 40
 aaaacgctaa actaccgaa tacg 24
 50 <210> 41
 <211> 19
 <212> ADN

<213> Secuencia artificial
 <220>
 <223> Sintética
 <400> 41
 5 tggcggggtt tcgtgtac 19
 <210> 42
 <211> 21
 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Sintética
 <400> 42
 10 gtccgaaac atcgaaacg a 21
 <210> 43
 <211> 19
 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Sintética
 <400> 43
 20 gcgtttgat ttgcgttc 19
 <210> 44
 <211> 20
 <212> ADN
 25 <213> Secuencia artificial
 <220>
 <223> Sintética
 <400> 44
 30 aaaatagcc gctaccgata 20
 <210> 45
 <211> 20
 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Sintética
 <400> 45
 35 gtttagggag tcgcggttac 20
 <210> 46
 <211> 23
 <212> ADN
 40 <213> Secuencia artificial
 <220>
 <223> Sintética
 <400> 46
 45 caaatcctac gaacgaacga acg 23
 <210> 47
 <211> 22
 <212> ADN
 <213> Secuencia artificial
 50 <220>

<223> Sintética
 <400> 47
 agtttgcggt agtcggtaga tc 22
 5
 <210> 48
 <211> 22
 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Sintética
 10
 <400> 48
 ggcgcgcaaat accgaataaa cg 22
 <210> 49
 <211> 22
 <212> ADN
 15
 <213> Secuencia artificial
 <220>
 <223> Sintética
 <400> 49
 tcggtttta gcgttcggtc gc 22
 20
 <210> 50
 <211> 23
 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Sintética
 25
 <400> 50
 aaacaacgaa acgccaatcc cga 23
 <210> 51
 <211> 25
 <212> ADN
 <213> Secuencia artificial
 30
 <220>
 <223> Sintética
 <400> 51
 tagtttttg gcgttatttt cggtc 25
 35
 <210> 52
 <211> 21
 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Sintética
 40
 <400> 52
 gcaactccgt aactcgacg a 21
 <210> 53
 <211> 26
 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Sintética
 50
 <400> 53

ttttcgttt gttttcggg attcgc 26
 <210> 54
 <211> 22
 <212> ADN
 <213> Secuencia artificial
 5
 <220>
 <223> Sintética
 <400> 54
 cgaatcctaa cgaactatcc ga 22
 10
 <210> 55
 <211> 25
 <212> ADN
 <213> Secuencia artificial
 15
 <220>
 <223> Sintética
 <400> 55
 ttcggtggat ttcgtattg atttc 25
 20
 <210> 56
 <211> 24
 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Sintética
 <400> 56
 aaacgaaacc gcgaactaaa acga 24
 25
 <210> 57
 <211> 22
 <212> ADN
 <213> Secuencia artificial
 30
 <220>
 <223> Sintética
 <400> 57
 ttacgtgata gttcggggtt tc 22
 35
 <210> 58
 <211> 21
 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Sintética
 <400> 58
 ataaaacgac gcgacgaaac g 21
 40
 <210> 59
 <211> 24
 <212> ADN
 <213> Secuencia artificial
 45
 <220>
 <223> Sintética
 <400> 59
 tttcgggttt tgcgtttat tcgc 24
 50
 <210> 60

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

 <220>
 5 <223> Sintética

 <400> 60
 gaaaaaaaaa aacgctaaaa atacgacg 28

 <210> 61
 <211> 21
 10 <212> ADN
 <213> Secuencia artificial

 <220>
 <223> Sintética

 <400> 61
 15 tagcgcgtag tggcgtagt c 21

 <210> 62
 <211> 18
 <212> ADN
 <213> Secuencia artificial

 20 <220>
 <223> Sintética

 <400> 62
 cctccgccgc tacaaccg 18

 <210> 63
 <211> 22
 25 <212> ADN
 <213> Secuencia artificial

 <220>
 <223> Sintética

 30 <400> 63
 tcgtgtttt aggatcgct tc 22

 <210> 64
 <211> 22
 <212> ADN
 35 <213> Secuencia artificial

 <220>
 <223> Sintética

 <400> 64
 gacgaacgat aaacgacgac ga 22

 40 <210> 65
 <211> 21
 <212> ADN
 <213> Secuencia artificial

 <220>
 <223> Sintética

 45 <400> 65
 ttcggtcgcg ttgtcgta c 21

 <210> 66
 <211> 25
 50 <212> ADN
 <213> Secuencia artificial

<220>
 <223> Sintética

 <400> 66
 aaacgaaaa caactcgaat aacga 25

 5 <210> 67
 <211> 18
 <212> ADN
 <213> Secuencia artificial

 <220>
 <223> Sintética

 10 <400> 67
 agtcggggtc ggagtcgc 18

 <210> 68
 <211> 23
 <212> ADN
 <213> Secuencia artificial

 15 <220>
 <223> Sintética

 <400> 68
 ataaatccct ccgaaacca cga 23

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

 20 <220>
 <223> Sintética

 <400> 69
 tcggaagtga cgtagggtag c 21

 <210> 70
 <211> 19
 <212> ADN
 <213> Secuencia artificial

 25 <220>
 <223> Sintética

 <400> 70
 cacacgccg ctaacacga 19

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

 30 <220>
 <223> Sintética

 <400> 71
 gcgcggtcgg gtttatattg c 21

 <210> 72
 <211> 20
 <212> ADN
 <213> Secuencia artificial

 35 <220>
 <223> Sintética

 <400> 71
 gcgcggtcgg gtttatattg c 21

 <210> 72
 <211> 20
 <212> ADN
 <213> Secuencia artificial

 40 <220>
 <223> Sintética

 <400> 71
 gcgcggtcgg gtttatattg c 21

 <210> 72
 <211> 20
 <212> ADN
 <213> Secuencia artificial

 45 <220>
 <223> Sintética

 <400> 71
 gcgcggtcgg gtttatattg c 21

 <210> 72
 <211> 20
 <212> ADN
 <213> Secuencia artificial

 50 <220>
 <223> Sintética

<400> 72
 gaccaactac cgctactcga 20

 <210> 73
 <211> 20
 5 <212> ADN
 <213> Secuencia artificial

 <220>
 <223> Sintética

 <400> 73
 10 aggggagaat ttcgcggttc 20

 <210> 74
 <211> 24
 <212> ADN
 <213> Secuencia artificial

 15 <220>
 <223> Sintética

 <400> 74
 aactaaatta aacctcaacc gccg 24

 20 <210> 75
 <211> 20
 <212> ADN
 <213> Secuencia artificial

 <220>
 <223> Sintética

 25 <400> 75
 ttaggaggcg aggtttgcbc 20

 <210> 76
 <211> 28
 <212> ADN
 30 <213> Secuencia artificial

 <220>
 <223> Sintética

 <400> 76
 gacgaaaccg taacgaaat aaaaacga 28

 35 <210> 77
 <211> 24
 <212> ADN
 <213> Secuencia artificial

 <220>
 40 <223> Sintética

 <400> 77
 cgaactatcc gaaaaaacga cgaa 24

 <210> 78
 <211> 22
 45 <212> ADN
 <213> Secuencia artificial

 <220>
 <223> Sintética

 <400> 78
 50 gcgacgcgag cgtaatttt tc 22

<210> 79
 <211> 23
 <212> ADN
 <213> Secuencia artificial
 5 <220>
 <223> Sintética
 <400> 79
 ttccggtata tattcgtcga gtc 23
 10 <210> 80
 <211> 20
 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Sintética
 15 <400> 80
 caccgacct atcaccgacga 20
 <210> 81
 <211> 22
 <212> ADN
 20 <213> Secuencia artificial
 <220>
 <223> Sintética
 <400> 81
 gtaccgctggt ttagttcgtg gc 22
 25 <210> 82
 <211> 21
 <212> ADN
 <213> Secuencia artificial
 <220>
 30 <223> Sintética
 <400> 82
 ccgaaacgcg atatcaaccg a 21
 <210> 83
 <211> 20
 <212> ADN
 35 <213> Secuencia artificial
 <220>
 <223> Sintética
 <400> 83
 40 cgggacggtta gagggttgtc 20
 <210> 84
 <211> 26
 <212> ADN
 <213> Secuencia artificial
 45 <220>
 <223> Sintética
 <400> 84
 ctcgaaaatt cgtaaaaacc ctccga 26
 50 <210> 85
 <211> 29
 <212> ADN

<213> Secuencia artificial
 <220>
 <223> Sintética
 5 <400> 85
 cgagtagttt ttttttat cgttagac 29
 <210> 86
 <211> 24
 <212> ADN
 <213> Secuencia artificial
 10 <220>
 <223> Sintética
 <400> 86
 caaaaaacga cacgtaaacg atcg 24
 <210> 87
 <211> 24
 <212> ADN
 <213> Secuencia artificial
 15 <220>
 <223> Sintética
 20 <400> 87
 gtttcgtttt gcgtttttt gcgc 24
 <210> 88
 <211> 19
 <212> ADN
 25 <213> Secuencia artificial
 <220>
 <223> Sintética
 <400> 88
 tccgaatcg ctactocga 19
 30 <210> 89
 <211> 17
 <212> ADN
 <213> Secuencia artificial
 35 <220>
 <223> Sintética
 <400> 89
 gcggttaggc ggggtgc 17
 <210> 90
 <211> 25
 <212> ADN
 <213> Secuencia artificial
 40 <220>
 <223> Sintética
 <400> 90
 attatatcaa tcccaaaaac acgcg 25
 <210> 91
 <211> 22
 <212> ADN
 <213> Secuencia artificial
 45 <220>
 50 <220>

<223> Sintética
 <400> 91
 tatttttcga attcgagttc gc 22
 5
 <210> 92
 <211> 22
 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Sintética
 10
 <400> 92
 tcacccgata aaaacgaaaa cg 22
 <210> 93
 <211> 21
 <212> ADN
 15
 <213> Secuencia artificial
 <220>
 <223> Sintética
 <400> 93
 gcgtcgttag tagtacgaag c 21
 20
 <210> 94
 <211> 21
 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Sintética
 25
 <400> 94
 gcacctcaac gaaaacaccg a 21
 <210> 95
 <211> 23
 <212> ADN
 30
 <213> Secuencia artificial
 <220>
 <223> Sintética
 <400> 95
 tcgaggcggg taattttatt cgc 23
 35
 <210> 96
 <211> 23
 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Sintética
 40
 <400> 96
 gctctaacc aaatacgcga 23
 <210> 97
 <211> 22
 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Sintética
 50
 <400> 97

tcggttcga ggtaagtta gc 22
 <210> 98
 <211> 23
 <212> ADN
 <213> Secuencia artificial
 5
 <220>
 <223> Sintética
 <400> 98
 cacttcgaaa caaaattacg cga 23
 10
 <210> 99
 <211> 20
 <212> ADN
 <213> Secuencia artificial
 15
 <220>
 <223> Sintética
 <400> 99
 ggaagcggac gtttcgttc 20
 20
 <210> 100
 <211> 22
 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Sintética
 <400> 100
 acccaaatc cgaaaacgac ga 22
 25
 <210> 101
 <211> 19
 <212> ADN
 <213> Secuencia artificial
 30
 <220>
 <223> Sintética
 <400> 101
 aggttcgccc cgtgatttc 19
 35
 <210> 102
 <211> 20
 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Sintética
 <400> 102
 ccaaaaccac gcgaacacga 20
 40
 <210> 103
 <211> 20
 <212> ADN
 <213> Secuencia artificial
 45
 <220>
 <223> Sintética
 <400> 103
 gttcggagtg tcgtagtcgc 20
 50
 <210> 104

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

5

<220>
 <223> Sintética

<400> 104
 aatctcgct acgaaacgac g 21

10

<210> 105
 <211> 22
 <212> ADN
 <213> Secuencia artificial

<220>
 <223> Sintética

15

<400> 105
 gtttagggac gtttcgttt tc 22

20

<210> 106
 <211> 20
 <212> ADN
 <213> Secuencia artificial

<220>
 <223> Sintética

25

<400> 106
 aacgaacgct cgataaccga 20

<210> 107
 <211> 20
 <212> ADN
 <213> Secuencia artificial

<220>
 <223> Sintética

30

<400> 107
 ttgggtcgg gttaggtcgc 20

35

<210> 108
 <211> 25
 <212> ADN
 <213> Secuencia artificial

<220>
 <223> Sintética

40

<400> 108
 gaaacaaaa aaacgctaac togta 25

<210> 109
 <211> 20
 <212> ADN
 <213> Secuencia artificial

45

<220>
 <223> Sintética

<400> 109
 cgtgttgga gtttgcgctc 20

50

<210> 110
 <211> 21
 <212> ADN
 <213> Secuencia artificial

<220>
 <223> Sintética

 <400> 110
 taccggaacc gcgataaaac g 21

 5 <210> 111
 <211> 25
 <212> ADN
 <213> Secuencia artificial

 <220>
 <223> Sintética

 10 <400> 111
 ggtatagttc ggttttagt cgttc 25

 <210> 112
 <211> 24
 15 <212> ADN
 <213> Secuencia artificial

 <220>
 <223> Sintética

 20 <400> 112
 tcttttctc cgaaaaccga aacg 24

 <210> 113
 <211> 23
 <212> ADN
 <213> Secuencia artificial

 25 <220>
 <223> Sintética

 <400> 113
 ggttacggtt agtattcgga ttc 23

 30 <210> 114
 <211> 20
 <212> ADN
 <213> Secuencia artificial

 <220>
 <223> Sintética

 35 <400> 114
 atatcaaccg cctaccgcg 20

 <210> 115
 <211> 24
 <212> ADN
 40 <213> Secuencia artificial

 <220>
 <223> Sintética

 <400> 115
 tttgttta atgcggcggg tggc 24

 45 <210> 116
 <211> 22
 <212> ADN
 <213> Secuencia artificial

 <220>
 <223> Sintética

 50 <220>
 <223> Sintética

	<400> 116 tatccgaact atccgctacc ga	22
5	<210> 117 <211> 19 <212> ADN <213> Secuencia artificial	
	<220> <223> Sintética	
10	<400> 117 ggtcgcgtcg ttttcgatc	19
	<210> 118 <211> 17 <212> ADN <213> Secuencia artificial	
15	<220> <223> Sintética	
	<400> 118 gccgcaaacg ccgacga	17
20	<210> 119 <211> 21 <212> ADN <213> Secuencia artificial	
	<220> <223> Sintética	
25	<400> 119 aatcggcgggt agtacgagta c	21
	<210> 120 <211> 26 <212> ADN <213> Secuencia artificial	
30	<220> <223> Sintética	
	<400> 120 aaaccaaacg cgtaacgata ataacg	26
35	<210> 121 <211> 26 <212> ADN <213> Secuencia artificial	
40	<220> <223> Sintética	
	<400> 121 gagtttgggt ttcgaaatta ttcgc	26
45	<210> 122 <211> 18 <212> ADN <213> Secuencia artificial	
	<220> <223> Sintética	
50	<400> 122 cccgaattac cgacgacg	18

<210> 123
 <211> 21
 <212> ADN
 <213> Secuencia artificial
 5 <220>
 <223> Sintética
 <400> 123
 aggttcgggt tcgacgattt c 21

10 <210> 124
 <211> 21
 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Sintética
 15 <400> 124
 aactctacaa cgccgaaacc g 21

20 <210> 125
 <211> 23
 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Sintética
 <400> 125
 ttagttcgtt tagcgatggc gtc 23

25 <210> 126
 <211> 20
 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Sintética
 30 <400> 126
 ccgaaactat cccgcaacga 20

35 <210> 127
 <211> 21
 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Sintética
 <400> 127
 ttcgtcgttt gggttatcgg c 21

40 <210> 128
 <211> 23
 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Sintética
 <400> 128
 gccctaaaac taaacaacc gcg 23

50 <210> 129
 <211> 22
 <212> ADN

<213> Secuencia artificial
 <220>
 <223> Sintética
 <400> 129
 5 ggttgctggt ttagttcgtc gc 22
 <210> 130
 <211> 19
 <212> ADN
 <213> Secuencia artificial
 10 <220>
 <223> Sintética
 <400> 130
 gcgaaaacgc ccgaaccga 19
 <210> 131
 15 <211> 22
 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Sintética
 20 <400> 131
 tcgtttggga gacgtattcg tc 22
 <210> 132
 <211> 25
 25 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Sintética
 <400> 132
 30 actcgaaaaa ttccgaact aacga 25
 <210> 133
 <211> 20
 <212> ADN
 <213> Secuencia artificial
 35 <220>
 <223> Sintética
 <400> 133
 tcggcggttt tcgttatcgc 20
 <210> 134
 <211> 21
 40 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Sintética
 <400> 134
 45 ccacgaaact cgcaactacg a 21
 <210> 135
 <211> 25
 <212> ADN
 <213> Secuencia artificial
 50 <220>

<223> Sintética
 <400> 135
 cgtttatta gcgtaatcgt ttgc 25
 <210> 136
 <211> 24
 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Sintética
 <400> 136
 gaataacgaa cgttcgacta ccga 24
 <210> 137
 <211> 24
 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Sintética
 <400> 137
 cggacgagat tagtttcgt tagc 24
 <210> 138
 <211> 24
 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Sintética
 <400> 138
 tcgtcaatca ctcgacgaaa acga 24
 <210> 139
 <211> 22
 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Sintética
 <400> 139
 tcggttcgta ggtatacgtg tc 22
 <210> 140
 <211> 22
 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Sintética
 <400> 140
 gctactacca atactccgc ga 22
 <210> 141
 <211> 21
 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Sintética
 <400> 141

gttagacggt ttagtttcg c 21
 <210> 142
 <211> 20
 <212> ADN
 <213> Secuencia artificial
 5
 <220>
 <223> Sintética
 <400> 142
 aaaaaccga cgaagattcg 20
 10
 <210> 143
 <211> 21
 <212> ADN
 <213> Secuencia artificial
 15
 <220>
 <223> Sintética
 <400> 143
 gttagagttc gcgtagcgta c 21
 20
 <210> 144
 <211> 25
 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Sintética
 <400> 144
 gaaaaaacca accgaacgaa aacga 25
 25
 <210> 145
 <211> 19
 <212> ADN
 <213> Secuencia artificial
 30
 <220>
 <223> Sintética
 <400> 145
 cggggcgttt cgatgtcgc 19
 35
 <210> 146
 <211> 24
 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Sintética
 <400> 146
 ccgaacgacc aaataaaacc aacg 24
 40
 <210> 147
 <211> 22
 <212> ADN
 <213> Secuencia artificial
 45
 <220>
 <223> Sintética
 <400> 147
 cgttttatgt tgggagcgtt cg 22
 50
 <210> 148

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

5
 <220>
 <223> Sintética

<400> 148
 gaccgaaccg cgtctaaacg 20

10
 <210> 149
 <211> 21
 <212> ADN
 <213> Secuencia artificial

<220>
 <223> Sintética

15
 <400> 149
 tacgtatcga ggtgcgctcg c 21

<210> 150
 <211> 25
 <212> ADN
 <213> Secuencia artificial

20
 <220>
 <223> Sintética

<400> 150
 aaactctaaa acgaacgaaa ctgca 25

25
 <210> 151
 <211> 21
 <212> ADN
 <213> Secuencia artificial

<220>
 <223> Sintética

30
 <400> 151
 tcgagaccg tttttgcgt c 21

<210> 152
 <211> 20
 <212> ADN
 <213> Secuencia artificial

35
 <220>
 <223> Sintética

<400> 152
 aacgatcccg aaccgccgta 20

40
 <210> 153
 <211> 21
 <212> ADN
 <213> Secuencia artificial

<220>
 <223> Sintética

45
 <400> 153
 cgagtagtgc gttttcggc c 21

<210> 154
 <211> 22
 <212> ADN
 <213> Secuencia artificial

50

<220>
 <223> Sintética

 <400> 154
 gacaacaacg ataacgacga cg 22
 5 <210> 155
 <211> 22
 <212> ADN
 <213> Secuencia artificial

 <220>
 <223> Sintética
 10 <400> 155
 agcgtgcggt attcggttt gc 22

 <210> 156
 <211> 23
 15 <212> ADN
 <213> Secuencia artificial

 <220>
 <223> Sintética
 20 <400> 156
 acctacgatt cgtaaaccga acg 23

 <210> 157
 <211> 23
 <212> ADN
 <213> Secuencia artificial
 25 <220>
 <223> Sintética

 <400> 157
 agttcgcggt ttttcggtc gtc 23
 30 <210> 158
 <211> 21
 <212> ADN
 <213> Secuencia artificial

 <220>
 <223> Sintética
 35 <400> 158
 aaccgacgca ccgactaacg a 21

 <210> 159
 <211> 22
 <212> ADN
 <213> Secuencia artificial
 40 <220>
 <223> Sintética

 <400> 159
 ttgcgtcggt tgcgttttc gc 22
 45 <210> 160
 <211> 20
 <212> ADN
 <213> Secuencia artificial

 <220>
 <223> Sintética
 50

<400> 160
 caactcgcca atctcgccga 20

 <210> 161
 <211> 18
 5 <212> ADN
 <213> Secuencia artificial

 <220>
 <223> Sintética

 <400> 161
 10 tattgcgggg aggtgttc 18

 <210> 162
 <211> 24
 <212> ADN
 <213> Secuencia artificial

 15 <220>
 <223> Sintética

 <400> 162
 tcaaaaaata attaaccgaa ccga 24

 20 <210> 163
 <211> 24
 <212> ADN
 <213> Secuencia artificial

 <220>
 <223> Sintética

 25 <400> 163
 ttagtttcg aagtttcgt tcgc 24

 <210> 164
 <211> 19
 30 <212> ADN
 <213> Secuencia artificial

 <220>
 <223> Sintética

 <400> 164
 tccgacccta tccgacga 19

 35 <210> 165
 <211> 27
 <212> ADN
 <213> Secuencia artificial

 <220>
 40 <223> Sintética

 <400> 165
 gattagtaa ttaacgataa agttcgc 27

 <210> 166
 <211> 23
 45 <212> ADN
 <213> Secuencia artificial

 <220>
 <223> Sintética

 <400> 166
 50 ccgaaaacgc ataatacct cga 23

<210> 167
 <211> 26
 <212> ADN
 <213> Secuencia artificial
 5 <220>
 <223> Sintética
 <400> 167
 ttggagagtt tttcgaattt ttctgc 26
 10 <210> 168
 <211> 19
 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Sintética
 15 <400> 168
 gaaaaccacc ctaacgccg 19
 <210> 169
 <211> 17
 <212> ADN
 20 <213> Secuencia artificial
 <220>
 <223> Sintética
 <400> 169
 cgcggggttc gtaggtc 17
 25 <210> 170
 <211> 23
 <212> ADN
 <213> Secuencia artificial
 <220>
 30 <223> Sintética
 <400> 170
 cgacaacaa caacgaaatc gaa 23
 <210> 171
 <211> 22
 <212> ADN
 35 <213> Secuencia artificial
 <220>
 <223> Sintética
 <400> 171
 agtagcgggtt atagtggcgt tc 22
 <210> 172
 <211> 22
 <212> ADN
 <213> Secuencia artificial
 45 <220>
 <223> Sintética
 <400> 172
 gcattcgcga cgaaaacaaa cg 22
 <210> 173
 <211> 20
 <212> ADN
 50

<213> Secuencia artificial
 <220>
 <223> Sintética
 <400> 173
 5 gtattgaggt cggcgtgtgc 20
 <210> 174
 <211> 19
 <212> ADN
 <213> Secuencia artificial
 10 <220>
 <223> Sintética
 <400> 174
 cgcgccgaat aaaccgcga 19
 <210> 175
 15 <211> 20
 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Sintética
 20 <400> 175
 cgtagttcgg cgtagttcgc 20
 <210> 176
 <211> 21
 <212> ADN
 25 <213> Secuencia artificial
 <220>
 <223> Sintética
 <400> 176
 aaccgccccg acgacaatac g 21
 30 <210> 177
 <211> 20
 <212> ADN
 <213> Secuencia artificial
 <220>
 35 <223> Sintética
 <400> 177
 gtcgagcgcg ttcgtgtac 20
 <210> 178
 <211> 22
 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Sintética
 <400> 178
 45 gaccgaaaa ataaatccccg aa 22
 <210> 179
 <211> 24
 <212> ADN
 <213> Secuencia artificial
 50 <220>

<223> Sintética

<400> 179
gatttttta gttgttcga cggc 24

5 <210> 180
<211> 25
<212> ADN
<213> Secuencia artificial

<220>
<223> Sintética

10 <400> 180
aaaattacta aacgcgaaat cgacg 25

<210> 181
<211> 26
<212> ADN
15 <213> Secuencia artificial

<220>
<223> Sintética

<400> 181
taatgggatg ataaatgtat tcgcfg 26

20 <210> 182
<211> 26
<212> ADN
<213> Secuencia artificial

<220>
25 <223> Sintética

<400> 182
accgcctaata ccaactcgaa ctcgta 26

<210> 183
<211> 22
30 <212> ADN
<213> Secuencia artificial

<220>
<223> Sintética

<400> 183
35 ggtgtttta aagggtcgtc gt 22

<210> 184
<211> 19
<212> ADN
<213> Secuencia artificial

40 <220>
<223> Sintética

<400> 184
gaccgcgactc ctccacgta 19

<210> 185
45 <211> 30
<212> ADN
<213> Secuencia artificial

<220>
<223> Sintética

50 <400> 185

ggaagtttat agtggttcg gcgggtaggc 30
 <210> 186
 <211> 22
 <212> ADN
 5 <213> Secuencia artificial
 <220>
 <223> Sintética
 <400> 186
 gcgaaaaacg ttcgaacccg cg 22
 10 <210> 187
 <211> 28
 <212> ADN
 <213> Secuencia artificial
 <220>
 15 <223> Sintética
 <400> 187
 tgtcgtcgtc gcgtatfff agttgttc 28
 <210> 188
 <211> 22
 20 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Sintética
 <400> 188
 aaccgccgtc caaacatcg ta 22
 25 <210> 189
 <211> 25
 <212> ADN
 <213> Secuencia artificial
 30 <220>
 <223> Sintética
 <400> 189
 gaagagttag ggttcgggac gcgag 25
 35 <210> 190
 <211> 25
 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Sintética
 40 <400> 190
 aacgacaaa taaacgccga accga 25
 <210> 191
 <211> 25
 45 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Sintética
 <400> 191
 cgtaggagcg attaggtggg cgtcg 25
 50 <210> 192

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

5

<220>
 <223> Sintética

<400> 192
 aaaccaaac cgaaacgcg aaa 23

10

<210> 193
 <211> 26
 <212> ADN
 <213> Secuencia artificial

<220>
 <223> Sintética

15

<400> 193
 cgattcgggg gatggattag cgttgt 26

20

<210> 194
 <211> 30
 <212> ADN
 <213> Secuencia artificial

<220>
 <223> Sintética

<400> 194
 cgaaatcccc ctaacgaaaa tctccgaaaa 30

25

<210> 195
 <211> 25
 <212> ADN
 <213> Secuencia artificial

<220>
 <223> Sintética

30

<400> 195
 cggggtttt ttagcggggg tttc 25

35

<210> 196
 <211> 29
 <212> ADN
 <213> Secuencia artificial

<220>
 <223> Sintética

<400> 196
 cgcgatccga aaaataaatt aacgtact 29

40

<210> 197
 <211> 20
 <212> ADN
 <213> Secuencia artificial

45

<220>
 <223> Sintética

<400> 197
 agcggttcga gttgggacgg 20

50

<210> 198
 <211> 24
 <212> ADN
 <213> Secuencia artificial

<220>
 <223> Sintética

 <400> 198
 gaaaaacgcg atgccgaaa acgc 24
 5 <210> 199
 <211> 28
 <212> ADN
 <213> Secuencia artificial

 10 <220>
 <223> Sintética

 <400> 199
 gaaggttatt aathtaag tcgcgaa 28

 15 <210> 200
 <211> 25
 <212> ADN
 <213> Secuencia artificial

 20 <220>
 <223> Sintética

 <400> 200
 aaaaaaacg ttcccgacga cgcg 25

 25 <210> 201
 <211> 25
 <212> ADN
 <213> Secuencia artificial

 30 <220>
 <223> Sintética

 <400> 201
 agtgttata tatatcgcg tcgg 25

 35 <210> 202
 <211> 23
 <212> ADN
 <213> Secuencia artificial

 <220>
 <223> Sintética

 <400> 202
 gactatacac gcttaaccgc gaa 23

REIVINDICACIONES

1. procedimiento para cribar una neoplasia en una muestra obtenida de un sujeto, comprendiendo el procedimiento:
 - 5 1) analizar el estado de metilación de un marcador en una muestra obtenida de un sujeto; y
 - 2) identificar el sujeto que tiene una neoplasia cuando el estado de metilación del marcador es diferente de un estado de metilación del marcador analizado en un sujeto que no tiene una neoplasia,

en el que una región cromosómica que tiene una anotación CLEC11A comprende el marcador, en el que el marcador comprende una base en una región metilada de manera diferencial (DMR) seleccionada entre DMR 29 o 30 de la Tabla 1, o DMR 64 de la Tabla 10,

10 en el que la neoplasia es una neoplasia gastrointestinal, una neoplasia del páncreas, una neoplasia colorrectal, una neoplasia del estómago, o una neoplasia del esófago.
2. El procedimiento de la reivindicación 1, que comprende además localizar el sitio de la neoplasia en el sujeto, en el que el estado de metilación del marcador indica el sitio de la neoplasia en el sujeto.
3. El procedimiento de la reivindicación 1, en el que la neoplasia está presente en la región gastrointestinal superior del paciente o en la región gastrointestinal inferior del paciente.
- 15 4. procedimiento de la reivindicación 1 que comprende analizar una pluralidad de marcadores.
5. procedimiento de la reivindicación 1 en el que analizar el estado de metilación del marcador en la muestra comprende determinar el estado de metilación de una o más bases.
6. procedimiento de la reivindicación 1 en el que el estado de metilación del marcador comprende una metilación aumentada o disminuida del marcador con respecto a un estado de metilación del marcador normal o un patrón de metilación diferente del marcador con respecto a un estado de metilación normal del marcador.
- 20 7. procedimiento de la reivindicación 1, en el que la muestra es una muestra de heces, una muestra de tejido, una muestra de jugo pancreático, una muestra de fluido de un quiste pancreático, una muestra de sangre, o una muestra de orina.
8. procedimiento de la reivindicación 1, en el que el ensayo comprende usar la reacción en cadena de la polimerasa específica de metilación, secuenciación de ácidos nucleicos, espectrometría de masas, nucleasa específica de metilación, separación por masas, o captura de la diana.
9. procedimiento de la reivindicación 1, en el que el ensayo comprende usar un oligonucleótido específico de la metilación seleccionado de la SEQ ID NO: 47 o la SEQ ID NO: 48.
- 30 10. procedimiento de la reivindicación 1 en el que la DMR es la DMR N.º 29 o 30 de la Tabla 1.
11. procedimiento de la reivindicación 1 en el que al menos una región cromosómica que tiene una anotación de CLEC11A comprende el marcador.
12. procedimiento de la reivindicación 7 en el que el tejido es tejido esofágico o tejido pancreático.

Bosque aleatorio

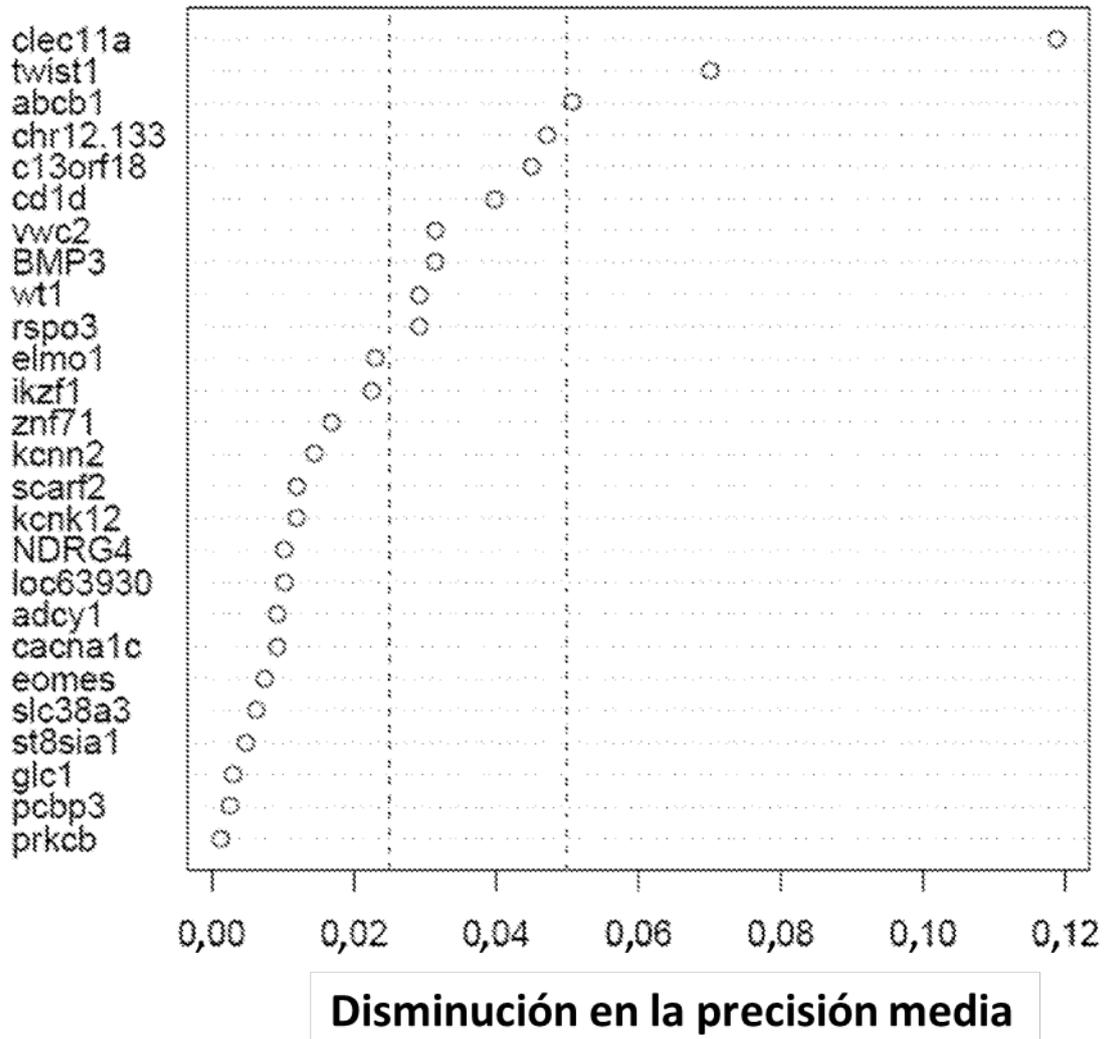


FIG. 1

Fig. 2.

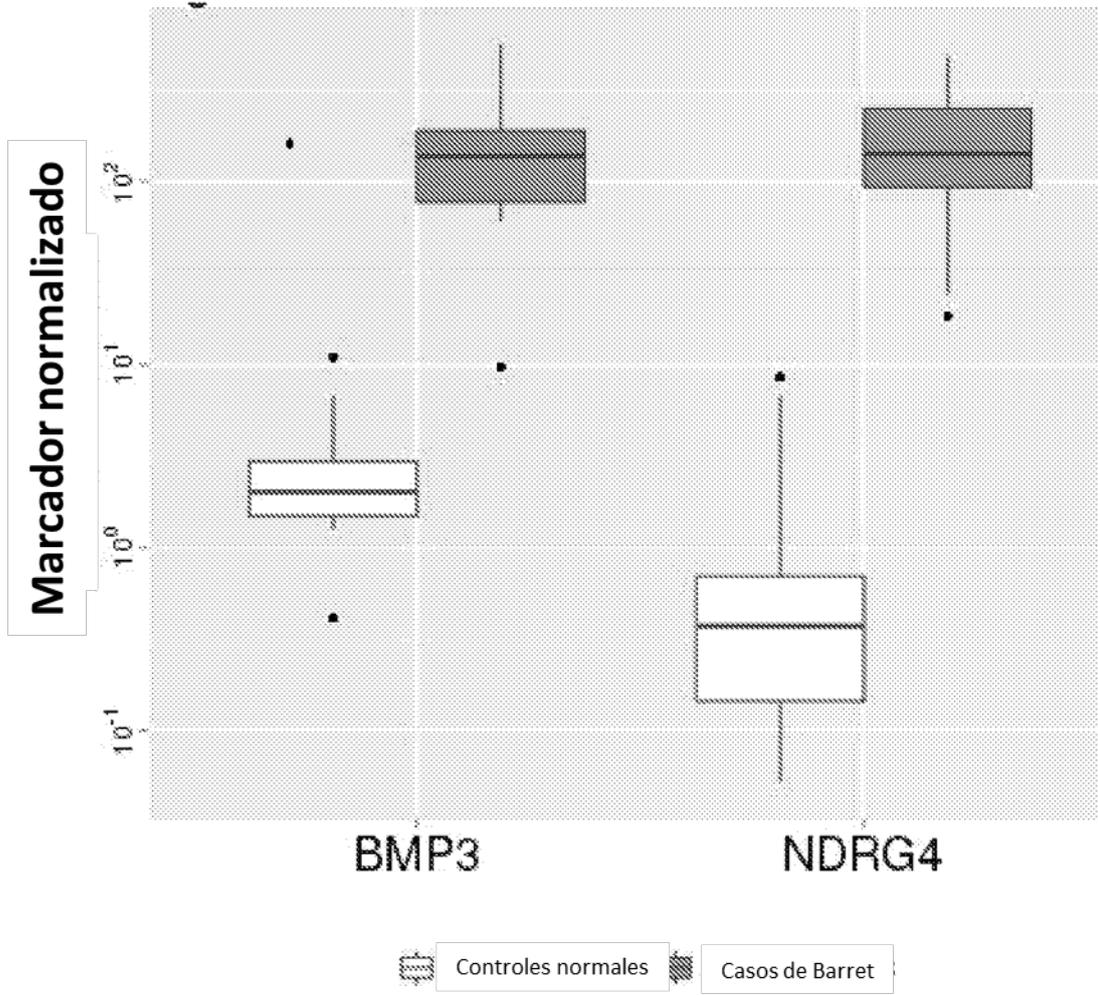


FIG. 3.

