

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 734 143**

51 Int. Cl.:

C12Q 1/6886 (2008.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **03.03.2015 PCT/EP2015/054451**

87 Fecha y número de publicación internacional: **11.09.2015 WO15132273**

96 Fecha de presentación y número de la solicitud europea: **03.03.2015 E 15708187 (8)**

97 Fecha y número de publicación de la concesión europea: **24.04.2019 EP 3114234**

54 Título: **Método para diagnosticar cáncer colorrectal a partir de una muestra de heces humanas mediante PCR cuantitativa**

30 Prioridad:

03.03.2014 EP 14382074

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

04.12.2019

73 Titular/es:

**FUNDACIO INSTITUT D'INVESTIGACIÓ
BIOMÈDICA DE GIRONA DR. JOSEP TRUETA
(50.0%)**

**C/ Dr. Castany, s/n, Edifici M2 - Parc Hospitalari
Martí i Julià**

17190 Salt (Girona) , ES y

UNIVERSITAT DE GIRONA (50.0%)

72 Inventor/es:

**SERRA PAGÈS, MARIONA;
GARCÍA-GIL, JESÚS;
MAS DE XAXARS, TERESA y
ALDEGUER, XAVIER**

74 Agente/Representante:

FÚSTER OLAGUIBEL, Gustavo Nicolás

Observaciones:

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

ES 2 734 143 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

Método para diagnosticar cáncer colorrectal a partir de una muestra de heces humanas mediante PCR cuantitativa

5 **Campo de invención**

La presente invención se refiere al campo de detección de cáncer colorrectal (CCR). Específicamente, se refiere a métodos para la detección temprana, cribado del riesgo y monitorización de CCR en un sujeto humano basándose en la cuantificación de una o más secuencias bacterianas de ADNr 16S en las heces. Se refiere además al uso de dichas secuencias bacterianas como biomarcadores de cáncer colorrectal.

Antecedentes de la invención

El cáncer colorrectal (CCR) es la segunda causa principal de muerte debida a cáncer en Europa y en Estados Unidos, y es el cáncer que se diagnostica más frecuentemente en Europa, con más de 400.000 nuevos casos y 200.000 muertes en 2008. Estos datos reflejan que es necesario tomar medidas preventivas.

La medida más eficaz y económica para reducir la mortalidad e incidencia de CCR son pruebas de cribado y monitorización del riesgo de CCR. Las pruebas de cribado se agrupan en aquellas que detectan principalmente cáncer temprano; y aquellas que pueden detectar cáncer temprano y también pueden detectar pólipos adenomatosos, proporcionando por tanto un mayor potencial para la prevención a través de polipectomía (es decir, extirpación de pólipos), véanse las pautas de cribado y vigilancia de la Sociedad Americana contra el Cáncer (ACS) de 2008, Levin *et al.* (CA Cancer J Clinicians, 2008;58(3)130-160). Se aplican diferentes pautas de cribado en personas que tienen un riesgo promedio y en personas con un riesgo aumentado o alto (<http://www.cancer.org/cancer/colonandrectumcancer/moreinformation/colonandrectumcancerearlydetection/colorectal-cancer-early-detection-acsc-recommendations>).

Actualmente, las pruebas que detectan principalmente CCR son pruebas fecales, que incluyen: i) pruebas de sangre fecal: la prueba de sangre oculta en las heces ("FOBT") y la prueba inmunoquímica fecal ("FIT"); y ii) la prueba de ADN en heces ("ADNh") (véase, <http://www.cancer.org/cancer/colonandrectumcancer/moreinformation/colonandrectumcancerearlydetection/colorectal-cancer-early-detection-screening-tests-used>).

35 i) Pruebas de sangre fecal

Tanto la FOBT como la FIT criban CCR detectando la cantidad de sangre en una muestra de heces. Las pruebas se basan en la premisa de que el tejido neoplásico, particularmente tejido maligno, sangra más que la mucosa típica, aumentando la cantidad de sangrado con el tamaño de pólipo y el estadio de cáncer. Se recomiendan múltiples pruebas debido al sangrado intermitente. Aunque las pruebas de sangre fecal pueden detectar algunos tumores de estadio temprano en la parte inferior del colon, son incapaces de detectar (i) CCR en la parte superior del colon ya que cualquier sangre se metabolizará y/o (ii) pólipos adenomatosos más pequeños, creando por tanto falsos negativos. Cualquier sangrado gastrointestinal debido a hemorroides, fisuras, trastornos inflamatorios (colitis ulcerosa, enfermedad de Crohn), enfermedades infecciosas, incluso carreras de largas distancias, crearán falsos positivos (Beg *et al.*, J Indian AcadClin Med, 2002;3(2)153-158). Las actuales pautas de la ACS (Levin *et al.*, CA Cancer J Clinicians, 2008;58(3)130-160) recomiendan el cribado anual de adultos con riesgo promedio con edades de 50 años y más usando la prueba de sangre oculta en las heces basada en Guaiac (gFOBT). La FOBT es la única prueba recomendada en la Unión Europea (Segnan N. *et al.*, 2010, European guidelines for quality assurance in colorectal cancer screening and diagnosis. Luxemburgo: Oficina para Publicaciones Oficiales de las Comunidades Europeas). Ambas pautas recomiendan que cualquier prueba FOBT positiva debe continuarse con una colonoscopia.

ii) Prueba de ADN en heces ("ADNh")

La prueba de ADNh mide una variedad de marcadores de ADN a partir de una muestra de heces. La prueba de ADNh disponible actualmente Cologuard™ de Exact Sciences Corp. (Madison, WI), mide un panel de marcadores múltiples que incluye mutaciones puntuales diferenciadas en genes K-ras, APC y P53; una sonda para BAT-26; un marcador para la integridad de ADN (DIA); y metilación del gen vimentina (Levin *et al.*, CA Cancer J Clinicians, 2008;58(3)130-160). Aunque algunas pautas recomiendan pruebas de ADNh otras pautas son más conservadoras y no lo recomiendan. En un estudio, una versión de la prueba de ADNh fue superior a FOBT, pero sólo detectó el 15% de los adenomas avanzados (Imperiale *et al.*, N Engl J Med, 2004;351:2704-2714).

Las pruebas indicadas en las pautas de la ACS como que son capaces de detectar pólipos adenomatosos y cáncer son pruebas estructurales que incluyen: (i) colonoscopia, (ii) sigmoidoscopia flexible ("FSIG"), (iii) enema de bario con contraste doble ("DCBE"), y (iv) colonografía CT ("CTC", colonoscopia virtual). Todos estos métodos requieren tanto purgado de los intestinos de un paciente como bombeo de aire en el colon para ayudar a la visualización. Las pautas actuales apoyan el uso de pruebas diseñadas para detectar tanto cáncer temprano como pólipos

adenomatosos si están disponibles los recursos y los pacientes están dispuestos a someterse a una prueba invasiva.

5 Una colonoscopia es una técnica invasiva que permite la inspección directa de todas las biopsias de extracción de mucosa y la resección del colon y pólipos del recto en una sesión. Esto no sólo mejora el pronóstico de cáncer colorrectal sino que previene la enfermedad. Sin embargo, las principales desventajas asociadas con esta técnica son: alto coste debido a que requiere personal cualificado, el riesgo de muerte asociado con la anestesia y el riesgo de perforación intestinal (Garborg K. *et al.*, *Ann Oncol.*, 2013;24(8):1963-72).

10 La colonoscopia en manos de un profesional experimentado es 100% sensible y 100% específica, pero debido al riesgo y coste de la colonoscopia, antes de la colonoscopia se requiere llevar a cabo un cribado de población de riesgo usando pruebas no invasivas tales como las pruebas fecales mencionadas anteriormente para la detección de sangre en heces, es decir, gFOBT o FIT, que tal como se mencionó anteriormente conducen a un número importante de falsos resultados positivos, por tanto, a un número igual de colonoscopias innecesarias. Esto se refleja en la necesidad de una herramienta de cribado de CCR no invasiva eficaz. La actual carencia de una herramienta de cribado de CCR eficaz da como resultado largas listas de espera (> 6 semanas) para una colonoscopia y una gran cantidad de recursos económicos asignados a colonoscopias que fueron indicadas incorrectamente. Esta situación representa tanto para la salud pública como privada, un gasto significativo y da como resultado la saturación del sistema (Allison J. *et al.*, *Practical Gastroenterology*, 2007, 3: 21-32).

20 Por tanto, existe la necesidad de una mejora de la sensibilidad, especificidad, invasividad y/o rentabilidad de los métodos actuales para la detección temprana y monitorización de CCR y/o pólipos adenomatosos.

25 La base genética e historia natural de CCR están bien definidas, considerando que menos del 5% de los CCR son hereditarios, y la mayoría son esporádicos, diagnosticados en pacientes sin antecedentes personales o familiares de neoplasia colónica. La etiología de esta enfermedad aún se desconoce, aunque se sospecha un origen multifactorial en el que factores endógenos y exógenos están implicados activamente en el desarrollo tumoral. Estos factores son: edad, tabaco, antecedentes personales de enfermedad intestinal inflamatoria, dieta, estilo de vida y microbiota.

30 Últimamente, se ha demostrado que las comunidades bacterianas en la mucosa colónica de pacientes con CCR difieren de individuos sanos y la microbiota intestinal se ha propuesto como agente determinante en el desarrollo y la progresión de CCR a lo largo de sus estadios (Chen W *et al.*, *PLoS One*, 2012;7(6):e39743; Zhu Q *et al.*, *Tumour Biol.*, 2013;34(3):1285-300; Ahn J *et al.*, *J Natl Cancer Inst.*, 2013;105(24):1907-11; Na wu *et al.*, *Microbial ecology*, 2013;66(2):462-470).

35 El documento WO2012/170478 se refiere a métodos para detectar adenomas y cáncer colorrectal usando una firma bacteriana. Se llevó a cabo secuenciación génica de ARNr16 para caracterizar comunidades bacterianas adherentes a partir de biopsias de mucosa. Se calcularon las abundancias relativas entre sujetos con adenoma y sin adenoma, concluyendo que el desarrollo de adenomas está asociado con cambios en la abundancia relativa de diversos taxones bacterianos presentes en la mucosa del intestino. Específicamente, se determinó la abundancia de *Fusobacterium nucleatum* en la mucosa normal del recto de sujetos con y sin adenomas mediante cuantificación por qPCR del gen ARNr 16S (ADN 16S), concluyendo que la abundancia es mayor en casos de adenoma en comparación con los controles. Según dichos datos, se sugiere el uso de *F. nucleatum* como biomarcador de carcinogénesis colorrectal.

45 La importancia de la microbiota del intestino como agente en el desarrollo y la evolución de CCR a lo largo de sus estadios también se ha confirmado en un estudio llevado a cabo por uno de los inventores en muestras de biopsia de mucosa, lo que muestra la asociación de la abundancia de Firmicutes, Actinobacteria y *Escherichia coli* en pacientes con cáncer colorrectal y de *Faecalibacterium prausnitzii* en sujetos sanos (Mas de Xaxars T., 2012, Dipòsit legal: Gl. 1664-2012, <http://hdl.handle.net/10803/94513>).

50 Se conoce bien en la técnica, sin embargo, que existen diferencias entre la composición de microbiota en biopsias de mucosa y muestras de heces. Lepage *et al.* (*Inflamm Bowel Dis.*, 2005;11(5):473-80) muestran que en un individuo dado las especies dominantes difieren entre microbiota fecal y asociada a la mucosa. De manera similar, Eckburg *et al.* (*Science*, 2005;308(5728): 1635-1638) notificaron diferencias estadísticamente significativas entre los linajes filogenéticos encontrados en muestras de la mucosa y heces, postulando que la microbiota fecal representa una combinación de bacterias de la mucosa desprendidas y población luminal no adherente diferenciada.

60 Nechvatal *et al.* (*J Microbiol Methods.*, 2008;72(2):124-32) mostraron que puede usarse ADN fecal como fuente para estudios epidemiológicos de bacterias intestinales y marcadores de cáncer humano. En particular, describe la cuantificación mediante PCR cuantitativa en tiempo real de secuencias bacterianas y humanas en extractos de ADN a partir de heces humanas. Balamurugan *et al.* (*J GastroenterolHepatol.*, 2008;23:1298-303) describen la cuantificación mediante PCR cuantitativa en tiempo real de bacterias específicas con productos del metabolismo que se conoce que tienen un efecto protector (es decir, *Eubacterium rectale* y *Faecalibacterium prausnitzii*, que son bacterias productoras de butirato) o que son tóxicos o carcinógenos (es decir, *Desulfovibrio* (bacterias reductoras de sulfato) y *Enterococcus faecalis* (que produce superóxido extracelular) en las heces de pacientes con cáncer

colorrectal. Sin embargo, ambas son silenciosas ante la determinación del valor de diagnóstico o predictivo de tales secuencias bacterianas como biomarcadores para CCR.

5 En el documento US2014/0024036, Wang *et al.* describen un método para la detección de CCR que comprende la cuantificación del gen *nusG* de *F. nucleatum* (FNN) mediante qPCR a partir de heces humanas de pacientes con CCR y controles, detectando CCR con una sensibilidad del 57% y especificidad del 89,5%.

10 Hassan Brim *et al.* (PLoS ONE, 8(12), e81352) se refieren a la obtención del perfil taxonómico de secuencias bacterianas en muestras de heces de individuos sanos y pacientes con pólipos en el colon usando tecnología de microalineamiento de ARNr 16S y pirosecuenciación. La presencia de pólipos en el colon se dice que está asociada con un 10% de riesgo de progresión maligna. No se comunican diferencias en la composición de la microbiota al nivel de género entre los dos grupos. Al nivel de subgénero, se encontró que las bacterias del grupo Bacteroides eran predominantes en las muestras de pacientes con pólipos. No obstante, el análisis metagenómico no reveló grandes diferencias en prevalencia/abundancia de genes bacterianos entre los dos grupos incluso cuando el análisis y las comparaciones se limitaron a los genomas de Bacteroides disponibles.

15 Aunque se han descrito métodos para la detección y/o cuantificación de biomarcadores de ADN para CCR en heces, aún sigue habiendo la necesidad de métodos fiables de diagnóstico temprano, cribado del riesgo o monitorización de CCR en pacientes en un sujeto humano a partir de muestras de heces. La disponibilidad de un método de este tipo permitiría prevenir eficazmente CCR mediante detección temprana y reducir el número de pruebas de examen estructural innecesarias, tales como colonoscopias, que se realizan hoy en día en individuos que se sospecha que tienen CCR más a menudo de lo necesario debido a la baja exactitud de las pruebas FOBT y FIT que son los métodos de cribado convencionales actuales para la detección de CCR.

25 **Sumario de la invención**

El primer aspecto de la invención se refiere a un método para determinar el riesgo de desarrollar cáncer colorrectal (CCR) en un sujeto humano que comprende:

- 30 i. cuantificar al menos una secuencia bacteriana de ADNr 16S seleccionada de la lista que consiste en SEQ ID NO: 1, SEQ ID NO: 7 y SEQ ID NO: 10, a partir de una muestra de heces de dicho sujeto;
- ii. comparar los niveles de cuantificación de la muestra del sujeto en (i) con los niveles de cuantificación en una muestra de control; y
- 35 iii. determinar el riesgo de desarrollar CCR en dicho sujeto humano, cuando existe una desviación de los valores de muestra del sujeto a partir de los valores en dicha muestra de control.

40 En un segundo aspecto, la invención se refiere a un método de cribado de cáncer colorrectal (CCR) en un sujeto humano que comprende:

- i. cuantificar al menos una secuencia bacteriana de ADNr 16S seleccionada de la lista que consiste en SEQ ID NO: 1, SEQ ID NO: 4, SEQ ID NO: 7 y SEQ ID NO: 10 a partir de una muestra de heces de dicho sujeto;
- 45 ii. comparar los niveles de cuantificación de la muestra del sujeto en (i) con los niveles de cuantificación en una muestra de control; y
- iii. diagnosticar, detectar de manera temprana o determinar recaída de CCR en dicho sujeto humano, o determinar que debe realizarse una colonoscopia en dicho sujeto humano, cuando existe una desviación de los valores de muestra del sujeto a partir de los valores en dicha muestra de control.
- 50

En un tercer aspecto, la invención se refiere al uso de un kit para diagnosticar, detectar de manera temprana o determinar recaída de CCR, determinar el riesgo de desarrollar CCR o predecir CCR o determinar si debe realizarse una colonoscopia, a partir de una muestra de heces de un sujeto humano, según el método de cualquiera de las reivindicaciones 1 a 11, comprendiendo dicho kit:

- a. un reactivo seleccionado del grupo que consiste en:
- 60 i. sondas de ácido nucleico capaces de hibridarse específicamente con al menos una secuencia de ADNr 16S seleccionada de la lista que consiste en SEQ ID NO: 1, SEQ ID NO: 4, SEQ ID NO: 7 y SEQ ID NO: 10; o
- ii. un par de cebadores de ácido nucleico capaces de amplificar específicamente al menos una secuencia de ADNr 16S seleccionada de la lista que consiste en SEQ ID NO: 1, SEQ ID NO: 4, SEQ ID NO: 7 y SEQ ID NO: 10; y
- 65 b. instrucciones para cuantificar los niveles de una o más de dichas secuencias a partir de una muestra de heces humanas.

Breve descripción de los dibujos

Los siguientes símbolos se usan en las figuras desde 1 hasta 7 con el siguiente significado:

- 5 - * significa que la significación estadística es $p= 0,05$
- ** significa que la significación estadística es $p= 0,01$
- 10 - *** significa que la significación estadística es $p= 0,005$
- La significación estadística de las figuras desde 8 hasta 26 está representada por * y el valor de p aparece en las tablas correspondientes.
- 15 **Figura 1.** Razón de valores de Ct para 2 secuencias: SEQ ID NO: 10/SEQ ID NO: 4. CCR representa los valores obtenidos a partir de muestras de heces de pacientes con cáncer colorrectal, y C representa los valores de control obtenidos de paciente sanos.
- 20 **Figura 2.** Razón de valores de Ct para 2 secuencias: SEQ ID NO: 7/SEQ ID NO: 4. CCR representa los valores obtenidos a partir de muestras de heces de pacientes con cáncer colorrectal, y C representa los valores de control obtenidos de paciente sanos.
- 25 **Figura 3.** Razón de valores de Ct para 2 secuencias: SEQ ID NO: 4/SEQ ID NO: 1. CCR representa los valores obtenidos a partir de muestras de heces de pacientes con cáncer colorrectal, y C representa los valores de control obtenidos de paciente sanos.
- 30 **Figura 4.** Valores de Ct absolutos para la secuencia SEQ ID NO: 4. CCR representa los valores obtenidos a partir de muestras de heces de pacientes con cáncer colorrectal, y C representa los valores de control obtenidos de pacientes sanos.
- 35 **Figura 5.** Valores de Ct absolutos para la secuencia SEQ ID NO: 7. CCR representa los valores obtenidos a partir de muestras de heces de pacientes con cáncer colorrectal, y C representa los valores de control obtenidos de pacientes sanos.
- 40 **Figura 6.** Razón de valores de Ct para 2 secuencias: SEQ ID NO: 4/ SEQ ID NO: 1. CCR representa los valores obtenidos a partir de muestras de heces de cáncer colorrectal, alto riesgo representa los valores obtenidos a partir de individuos con síndrome de Lynch que tenían pólipos en su última colonoscopia y tienen un riesgo aumentado de desarrollar cáncer colorrectal, bajo riesgo representa los valores obtenidos a partir de individuos con síndrome de Lynch que no tenían pólipos en su última colonoscopia.
- 45 **Figura 7.** Valores de Ct absolutos para la secuencia SEQ ID NO: 4. CCR representa los valores obtenidos a partir de muestras de heces de cáncer colorrectal, alto riesgo representa los valores obtenidos a partir de individuos con síndrome de Lynch que tenían pólipos en su última colonoscopia y tienen un riesgo aumentado de desarrollar cáncer colorrectal, bajo riesgo representa los valores obtenidos a partir de individuos con síndrome de Lynch que no tenían pólipos en su última colonoscopia.
- 50 **Figura 8.** Validación de cebadores para amplificación de B3: gráficos de amplificación (8a) y curva de disociación (8b).
- Figura 9.** Validación de cebadores para amplificación de B10: gráficos de amplificación (9a) y curva de disociación (9b).
- 55 **Figura 10.** Validación de cebadores para amplificación de B41: gráficos de amplificación (10a) y curva de disociación (10b).
- Figura 11.** Validación de cebadores para amplificación de B46: gráficos de amplificación (11a) y curva de disociación (11b).
- 60 **Figura 12.** Validación de cebadores para amplificación de B48: gráficos de amplificación (12a) y curva de disociación (12b).
- Figura 13.** Validación de cebadores para amplificación de B50: gráficos de amplificación (13a) y curva de disociación (13b).
- 65 **Figura 14.** Valores de Ct absolutos para B3 en grupos de CCR y C.

Figura 15. Valores de Ct absolutos para B48 en grupos de CCR y C.

Figura 16. Valores de Ct absolutos para B10 en grupos de CCR y C.

5 Figura 17. Valores de Ct absolutos para B46 en grupos de CCR y C.

Figura 18. Valores de Ct absolutos para B3 en grupos de CCR, L y C.

Figura 19. Valores de Ct absolutos para B48 en grupos de CCR, L y C.

10 Figura 20. Valores de Ct absolutos para B10 en grupos de CCR, L y C.

Figura 21. Valores de Ct absolutos para B46 en grupos de CCR, L y C.

15 Figura 22. Valores de Ct absolutos para B3, B10, B46 y B48 en CCR+L frente a C.

Figura 23. Razón B48/B10 de valores de Ct absolutos en grupos de CCR, alto riesgo L, bajo riesgo L y C.

Figura 24. Razón B3/B10 de valores de Ct absolutos en grupos de CCR, alto riesgo L, bajo riesgo L y C.

20 Figura 25. Razón B46/B10 de valores de Ct absolutos en grupos de CCR, alto riesgo L, bajo riesgo L y C.

Figura 26. Razón B3/B48 de valores de Ct absolutos en grupos de CCR, alto riesgo L, bajo riesgo L y C.

25 Figura 27. Curvas ROC para B3, B10, B46 y B48 en análisis de sanos frente a CCR.

Figura 28. Curvas ROC para B3, B10, B46 y B48 en análisis de sanos frente a Lynch.

Figura 29. Curvas ROC para B3, B10, B46 y B48 en análisis de sanos frente a CCR + Lynch.

30 Figura 30. Curvas ROC para B3, B10, B46 y B48 en análisis de CCR frente a Lynch.

Descripción detallada

35 En la presente divulgación, se han identificado secuencias bacterianas específicas asociadas con diagnóstico, detección temprana, cribado del riesgo y monitorización de CCR y/o pólipos adenomatosos.

El primer aspecto de la divulgación, se refiere a un método de cribado de cáncer colorrectal (CCR) y/o pólipos adenomatosos en un sujeto humano que comprende:

40 i. cuantificar al menos una secuencia bacteriana de ADNr 16S seleccionada de la lista que consiste en SEQ ID NO: 1, SEQ ID NO: 4, SEQ ID NO: 7 y SEQ ID NO: 10 o la secuencia de ARNr complementaria de la misma, a partir de una muestra de heces de dicho sujeto; y

45 ii. diagnosticar, detectar de manera temprana, determinar la recaída, determinar el riesgo de desarrollar, o predecir CCR y/o pólipos en un sujeto humano; o determinar si debe realizarse una colonoscopia en dicho sujeto humano; o determinar el pronóstico en el que dicho sujeto humano es un paciente al que se le ha diagnosticado CCR y/o pólipos; o guiar una terapia en un paciente con CCR y/o pólipos a partir de los niveles de cuantificación de al menos una de dichas secuencias.

50 En la presente solicitud, la secuencia identificada por SEQ ID NO: 1 también se identifica como "B3". De manera similar, SEQ ID NO: 4 también se identifica como "B10", SEQ ID NO: 7 también se identifica como "B46" y SEQ ID NO: 10 también se identifica como "B48", y estas designaciones se usan de manera intercambiable. B3, B10, B46 y B48 fueron los nombres usados por los inventores de la presente solicitud para identificar las secuencias durante el rendimiento de todos los experimentos llevados a cabo para llegar a la presente invención. Las secuencias identificadas por SEQ ID NO: 1, 4, 7 y 10 están disponibles en la base de datos de secuencias pública EMBL-EBI con los siguientes números de registro: GQ411111.1 (SEQ ID NO: 1), GQ411118.1 (SEQ ID NO: 4), GQ411150.1 (SEQ ID NO: 7), GQ411152.1 (SEQ ID NO: 10).

60 Las secuencias B3, B10, B46 y B48 se corresponden con bandas de gel de electroforesis en gel con gradiente desnaturizante (DGGE) aisladas previamente de aislados de bacterias no cultivadas. Se llevaron a cabo análisis BLAST con el fin de identificar las especies bacterianas correspondientes:

Mejor coincidencia BLAST para ADNr 16S de B3: *Collinsella aerofaciens*;

65 Mejor coincidencia BLAST para ADNr 16S de B10: *Faecalibacterium prausnitzii*;

Mejor coincidencia BLAST para ADNr 16S de B46: *Faecalibacterium prausnitzii*, *Subdoligranulum variabil*;

Mejor coincidencia BLAST para ADNr 16S de B48: *Ruminococcus*, *Roseburia*, *Coprococcus*.

Tal como se usa en el presente documento, el término “cáncer colorrectal” o “CCR” se usa para cáncer que se inicia en el colon o el recto. Estos cánceres también pueden denominarse por separado como cáncer de colon o cáncer colorrectal, según dónde comiencen. El cáncer de colon y cáncer colorrectal tienen muchas características en común.

Según la ACS, varios tipos de cáncer pueden iniciarse en el colon o recto. Más del 95% de los cánceres colorrectales son un tipo de cáncer conocido como adenocarcinomas. Estos cánceres se inician en células que forman glándulas que elaboran moco para lubricar el interior del colon y recto. Otros tipos de tumores menos frecuentes también pueden iniciarse en el colon y recto. Estos incluyen: tumores carcinoides, tumores estromales gastrointestinales (GIST), linfomas y sarcomas.

El objetivo del cribado de cáncer es reducir la mortalidad mediante una reducción en la incidencia de enfermedad avanzada. Para este fin, el cribado de CCR moderno puede lograr este objetivo mediante la detección de adenocarcinomas de estadio temprano y la detección y extirpación de pólipos adenomatosos, el último aceptado generalmente como lesiones precursoras no obligadas.

Un “pólipo adenomatoso” es un crecimiento no canceroso de células glandulares anómalas sobre el revestimiento interno de un órgano tal como el colon. Por ejemplo, tres tipos de pólipos adenomatosos que pueden crecer en el colon son adenomas tubulares, vellosos y tubulovellosos. Los pólipos adenomatosos son comunes en adultos de más de 50 años de edad, pero la mayoría de pólipos no se convertirá en un adenocarcinoma; la histología y el tamaño determinan su importancia clínica.

Según la ACS, se prefieren pruebas que tienen la mejor posibilidad de encontrar ambos pólipos y cáncer, ya que encontrar y extirpar sus pólipos previene que algunas personas desarrollen cáncer colorrectal. Preferiblemente, el método de la divulgación es un método para el cribado de CCR y pólipos adenomatosos.

En una realización particular, el método de la divulgación es un método de cribado para el diagnóstico o la detección de CCR y/o pólipos adenomatosos, preferiblemente para la detección temprana de CCR y/o pólipos adenomatosos. La expresión “detección temprana” se refiere a detección antes de la presencia de signos clínicos. Los términos “detección” y “diagnóstico” se usan de manera intercambiable en la presente solicitud. Se encuentra que el método de la divulgación es una potente herramienta para prevenir cáncer colorrectal mediante un diagnóstico de enfermedad en estadio temprano.

En una realización particular, el método de cribado de la divulgación se lleva a cabo en sujetos con 50 años de edad y más. Se recomienda que las mujeres y hombres de 50 años y más con riesgo promedio sigan pruebas de cribado de cáncer colorrectal (véase la tabla 2 de Levin *et al.*, CA Cancer J Clinicians, 2008;58(3)130-160).

En otra realización, el método de cribado de la divulgación se lleva a cabo en sujetos de alto riesgo y/o de riesgo aumentado. Las pautas específicas son de aplicación a pacientes de alto riesgo o riesgo aumentado.

Según la ACS (véase la tabla 3 de Levin *et al.*, CA Cancer J Clinicians, 2008;58(3)130-160), los sujetos de alto riesgo y riesgo aumentado incluyen los siguientes:

Riesgo aumentado

- sujetos con antecedentes de pólipos en colonoscopia previa;
- sujetos con cáncer colorrectal; y
- sujetos con antecedentes familiares de cáncer colorrectal o pólipos adenomatosos.

Alto riesgo

- sujetos con poliposis adenomatosa familiar (FAP) diagnosticada mediante pruebas genéticas, o FAP sospechada sin pruebas genéticas;
- sujetos con cáncer de colon sin poliposis hereditario (HNPCC o síndrome de Lynch), o en riesgo aumentado de HNPCC basándose en los antecedentes familiares sin pruebas genéticas; y
- sujetos con enfermedad intestinal inflamatoria, tal como colitis ulcerosa crónica o enfermedad de Crohn.

Un objetivo de la presente divulgación es proporcionar una herramienta de diagnóstico previo. Específicamente, la presente invención proporciona un método para diagnosticar cáncer colorrectal antes de presentar signos clínicos identificando sujetos con predisposición o riesgo de desarrollar CCR y/o pólipos adenomatosos.

5 En una realización particular, el método de cribado de la presente divulgación es un método para determinar el riesgo de desarrollar CCR y/o pólipos adenomatosos en un sujeto humano que comprende:

10 i. cuantificar al menos una secuencia bacteriana de ADNr 16S seleccionada de la lista que consiste en SEQ ID NO: 1, SEQ ID NO: 4, SEQ ID NO: 7 y SEQ ID NO: 10 o la secuencia de ARNr 16S complementaria de la misma, a partir de una muestra de heces de dicho sujeto; y

ii. determinar el riesgo de desarrollar CCR y/o pólipos adenomatosos a partir de los niveles de cuantificación de al menos una de dichas secuencias.

15 En una realización preferida, el método de cribado de la presente divulgación es un método para determinar el riesgo de desarrollar CCR y/o pólipos adenomatosos en sujetos de alto riesgo y/o de riesgo aumentado, preferiblemente en portadores del síndrome de Lynch.

20 El término "síndrome de Lynch" se refiere a cáncer colorrectal sin poliposis hereditario o HNPCC, que es un estado hereditario que aumenta enormemente el riesgo de una persona de padecer cáncer colorrectal, así como cáncer del revestimiento del útero (cáncer de endometrio), cáncer de ovario, y algunos otros cánceres. Las personas con este estado tienden a desarrollar cáncer a una edad temprana, a menudo sin tener en primer lugar muchos pólipos.

25 El método de cribado de la divulgación ha mostrado en el ejemplo 4 (tablas 3 y 4) la capacidad de discriminar de manera estadísticamente significativa entre pacientes con síndrome de Lynch con pólipos (población de alto riesgo) y sin pólipos (población de bajo riesgo). La cuantificación de SEQ ID NO: 4 (B10) y determinación de la razón de SEQ ID NO: 7 (B46)/SEQ ID NO: 4 (B10) parecen ser buenos indicadores según los resultados en el ejemplo 4. Por consiguiente, las razones de SEQ ID NO: 4 (B10) y SEQ ID NO: 7 (B46)/SEQ ID NO: 4 (B10) son biomarcadores preferidos para su uso en el método de la divulgación para cribado de la presencia de pólipos.

30 En sujetos con antecedentes de pólipos en la colonoscopia previa o con antecedentes de cáncer colorrectal tras resección del cáncer, se llevan a cabo pruebas de cribado con propósitos de monitorización o vigilancia, es decir, para detectar recaída de cáncer y/o pólipos adenomatosos en dichos sujetos humanos.

35 En una realización particular adicional, el método de cribado de la presente divulgación es un método de monitorización para la detección de cáncer colorrectal (CCR) y/o pólipos adenomatosos en un sujeto humano con antecedentes previos de pólipos y/o cáncer colorrectal que comprende:

40 i. cuantificar al menos una secuencia bacteriana de ADNr 16S seleccionada de la lista que consiste en SEQ ID NO: 1, SEQ ID NO: 4, SEQ ID NO: 7 y SEQ ID NO: 10 o la secuencia de ARNr 16S complementaria de la misma, a partir de una muestra de heces de dicho sujeto; y

45 ii. determinar la recaída de CCR y/o pólipos adenomatosos a partir de los niveles de cuantificación de al menos una de dichas secuencias.

50 El método de cribado de la divulgación ha mostrado en el ejemplo 8 (tabla 8) la capacidad de discriminar de manera estadísticamente significativa entre pacientes con síndrome de Lynch con antecedentes previos de CCR y pólipos en su última colonoscopia (población de alto riesgo) y pacientes con síndrome de Lynch sin antecedentes previos de CCR y sin pólipos en su última colonoscopia (población de bajo riesgo). Se obtienen resultados particularmente notables con la cuantificación de las razones de SEQ ID NO: 1/SEQ ID NO: 4 (B3/B10), SEQ ID NO: 4/SEQ ID NO: 7 (B10/B46), SEQ ID NO: 1/SEQ ID NO: 10 (B3/B48) y SEQ ID NO: 10/SEQ ID NO: 4 (B48/B10).

55 Por consiguiente, las razones de SEQ ID NO: 1/SEQ ID NO: 4 (B3/B10), SEQ ID NO: 4/SEQ ID NO: 7 (B10/B46), SEQ ID NO: 1/SEQ ID NO: 10 (B3/B48) y SEQ ID NO: 10/SEQ ID NO: 4 (B48/B10) son biomarcadores preferidos para su uso en el método de la divulgación para cribado/monitorización de la presencia de pólipos, preferiblemente en pacientes de alto riesgo o riesgo aumentado, tales como pacientes con síndrome de Lynch con antecedentes de CCR.

60 En una realización particular adicional, el método de cribado de la presente divulgación es un método para determinar si debe realizarse una colonoscopia en un sujeto humano que comprende:

65 i. cuantificar al menos una secuencia bacteriana de ADNr 16S seleccionada de la lista que consiste en SEQ ID NO: 1, SEQ ID NO: 4, SEQ ID NO: 7 y SEQ ID NO: 10 o la secuencia de ARNr 16S complementaria de la misma, a partir de una muestra de heces de dicho sujeto; y

ii. determinar si debe realizarse una colonoscopia en dicho sujeto humano a partir de los niveles de cuantificación de

al menos una de dichas secuencias.

Un experto habitual en la técnica conoce varios métodos y dispositivos para la cuantificación y análisis de los marcadores bacterianos de la divulgación. El término “cuantificar” se refiere a la capacidad para cuantificar la cantidad de una secuencia específica de ácido nucleico en una muestra. En la técnica se conocen bien métodos de biología molecular para medir cantidades de secuencias diana de ácido nucleico. Estos métodos incluyen pero no se limitan a PCR de punto final, PCR competitiva, PCR con transcriptasa inversa (RT-PCR), PCR cuantitativa (qPCR), qPCR con transcriptasa inversa (RT-qPCR), PCR-ELISA, microalineamientos de ADN, ensayos de hibridación *in situ* tales como ensayo de transferencia puntual o de hibridación *in situ* de fluorescencia (FISH), ADN ramificado (Nolte, Adv. Clin. Chem., 1998,33:201-235) y múltiples versiones de dichos métodos (véase por ejemplo, Andoh *et al.*, Current Pharmaceutical Design, 2009; 15,2066-2073). Los cebadores y/o sondas preferidos reaccionan de manera predecible, normalmente ofreciendo una respuesta directa y lineal a cantidades crecientes de secuencias bacterianas de ácido nucleico. Mediante la preparación de y comparación con patrones apropiados, puede cuantificarse fácilmente la cantidad de una secuencia de ácido nucleico dada en una muestra.

Preferiblemente, dicha secuencia de ácido nucleico es una secuencia de ADNr 16S seleccionada de la lista que consiste en SEQ ID NO: 1, SEQ ID NO: 4, SEQ ID NO: 7 y SEQ ID NO: 10,

Un método de cuantificación particularmente preferido es FISH, que combina hibridación de sonda con microscopía óptica fluorescente, microscopía láser confocal o citometría de flujo para cuantificación directa de secuencias bacterianas individuales. Para revisiones de la metodología FISH, véase, por ejemplo, Harmsen *et al.*, Appl Environ Microbiol, 2002;68 2982-2990, Kalliomaki *et al.*, J Allerg Clin Immunol, 2001;107 129-134; Tkachuk *et al.*, Genet. Anal. Tech. Appl., 1991;8:67-74; Trask *et al.*, Trends Genet., 1991;7 (5): 149-154; y Weier *et al.*, Expert Rev. Mol. Diagn., 2002,2(2):109-119; y patente estadounidense n.º 6.174.681.

Otro método de cuantificación particularmente preferido es PCR cuantitativa (qPCR), también conocido como PCR en tiempo real. El proceso de PCR se conoce bien en la técnica y, por tanto, no se describe en detalle en el presente documento. Están disponibles un resumen y protocolos de la tecnología de Q-PCR de vendedores tales como Sigma-Aldrich en aplicaciones de qPCR SYBR Green qPCR, véase por ejemplo <http://www.sigmaaldrich.com/technical-documents/protocols/biology/sybr-green-qpcr.html> o <http://www.sigmaaldrich.com/life-science/molecular-biology/pcr/quantitative-pcr/qpcr-technical-guide.html>. Para una revisión de métodos de qPCR para cuantificar la abundancia y expresión de marcadores de genes bacterianos véase, por ejemplo, Smith CJ y Osborn AM., FEMS Microbiol Ecol., 2009;67(1):6-20.

El término “niveles de cuantificación” puede ser la concentración (cantidad de ADN por unidad de volumen), la cantidad de ADN por número de células, el valor umbral de ciclo (valor de Ct) o cualquier transformación matemática de los mismos. En una realización preferida, la cuantificación de dichas secuencias bacterianas se realiza mediante qPCR. En una realización más preferida, la cuantificación de dichas secuencias bacterianas se realiza mediante qPCR y los niveles de cuantificación son el valor de Ct. El valor de Ct (umbral de ciclo) se define como el número de ciclos de qPCR requeridos para que la señal fluorescente cruce el umbral. Los niveles de Ct son inversamente proporcionales a la cantidad de ácido nucleico diana en la muestra (es decir, cuanto menor sea el nivel de Ct mayor será la cantidad de ácido nucleico diana en la muestra).

La cuantificación de la abundancia de una secuencia de ácido nucleico diana (por ejemplo SEQ ID NO: 1) en una muestra fecal puede ser absoluta o relativa. La cuantificación relativa se basa en uno o más genes de referencia interna, es decir, genes de ARNr 16S de cepas de referencia, tales como determinación de bacterias totales (Eubacteria) usando cebadores universales y expresando la abundancia de la secuencia de ácido nucleico diana como porcentaje de Eubacteria (por ejemplo razón de SEQ ID NO: 1/Eubacteria). La cuantificación absoluta proporciona el número exacto de moléculas diana mediante comparación con patrones de ADN.

En una realización particular, el método de la divulgación comprende además después de la etapa i) comparar los niveles de muestra del sujeto con los niveles en una muestra de control, en el que cuando existe una desviación de los niveles de una o más de las secuencias seleccionadas de la lista que consiste en SEQ ID NO: 1, SEQ ID NO: 4, SEQ ID NO: 7 y SEQ ID NO: 10 (preferiblemente, una desviación estadísticamente significativa) de los niveles de la(s) respectiva(s) secuencia(s) en la muestra de control es indicativa de CCR y/o pólipos adenomatosos. Por ejemplo, cuando los niveles de cuantificación presentan una desviación de la mediana del valor (punto de corte) más(+)/menos(-) la desviación estándar de/de los nivel(es) respectivo(s) medido(s) en una muestra de control es indicativo de CCR y/o pólipos adenomatosos.

En una realización preferida, en la que los niveles de concentración de una o más de dichas secuencias en dicha muestra del sujeto son menores que la mediana del valor (punto de corte) menos la desviación estándar de los niveles de la(s) respectiva(s) secuencia(s) medida en una muestra de control es indicativo de CCR y/o pólipos adenomatosos. En otra realización preferida, en la que los niveles de Ct de una o más de dichas secuencias en dicho sujeto muestra son mayores que la mediana del valor (punto de corte) más la desviación estándar de los niveles de la(s) respectiva(s) secuencia(s) medida en una muestra de control es indicativa de CCR y/o pólipos adenomatosos.

El término “muestra de control” puede referirse a una colección de muestras de control de la población de referencia, por ejemplo las muestras de control pueden ser muestras de sujetos sanos, de sujetos sin pólipos adenomatosos, de

5 La clasificación de sujetos humanos bajo una de las poblaciones de referencia mencionadas anteriormente se lleva a cabo mediante una exploración de colonoscopia. Por ejemplo, “sujetos sanos” son aquellos que en una colonoscopia previa no presentaron ni pólipos adenomatosos ni cáncer colorrectal.

10 Preferiblemente, el método de la divulgación comprende cuantificar 2, 3 ó 4 secuencias seleccionadas de la lista que consiste en SEQ ID NO: 1, SEQ ID NO: 4, SEQ ID NO: 7 y SEQ ID NO: 10, preferiblemente las 4 secuencias.

15 En una realización particular, el método de la divulgación comprende la cuantificación de al menos dos secuencias seleccionadas de la lista que consiste en SEQ ID NO: 1, SEQ ID NO: 4, SEQ ID NO: 7 y SEQ ID NO: 10 y la determinación de la relación entre al menos dos de dichas secuencias. Preferiblemente, se determina la relación entre dos, tres o cuatro de dichas secuencias (por ejemplo razón, análisis multivariante, etc.).

20 En una realización preferida, la razón entre estas al menos dos secuencias se obtiene dividiendo los niveles de cuantificación de una primera secuencia entre los niveles de cuantificación de una segunda secuencia. Por ejemplo, la razón de concentración de las secuencias SEQ ID NO: 10/SEQ ID NO: 4 se obtiene dividiendo la concentración de la secuencia SEQ ID NO: 10 entre la concentración de la secuencia SEQ ID NO: 4. La razón de SEQ ID NO: 10/SEQ ID NO: 4 también pudo obtenerse dividiendo el valor de Ct de la secuencia identificada por SEQ ID NO: 10 entre el valor de Ct de la secuencia identificada por SEQ ID NO: 4.

25 En una realización adicional del método de la divulgación, al menos dos de dichas secuencias bacterianas se cuantifican en la etapa i) y se determina al menos una de las razones de los niveles de dichas secuencias en la muestra del sujeto, comprendiendo además comparar al menos una de dichas razones en dicho sujeto con la(s) razón/razones respectiva(s) en una muestra de control, en la que una desviación de la razón en dicha muestra de control (preferiblemente, una desviación estadísticamente significativa) es indicativa de CCR y/o pólipos adenomatosos. Por ejemplo, en la que una o más de dichas razones presentan una desviación de la mediana del valor (punto de corte) más(+)/menos(-) la desviación estándar de la(s) razón/razones respectiva(s) medida en una muestra de control es indicativa de CCR y/o pólipos adenomatosos.

35 El término “estadísticamente significativo” cuando hace referencia a diferencias entre la muestra de prueba y la muestra de control, se refiere a la condición de que cuando se usa el análisis estadístico apropiado, la probabilidad de que los grupos sean los mismos es menor del 5%, por ejemplo $p < 0,05$. En otras palabras, la probabilidad de obtener los mismos resultados en una base completamente aleatoria es de menos de 5 de 100 intentos. Un experto en la técnica sabrá cómo escoger el análisis estadístico apropiado. Normalmente, el análisis estadístico apropiado se determina basándose en si la variable bajo el estudio tiene una distribución normal, por ejemplo, usando la prueba de Kolmogorov-Smirnov. Preferiblemente, en aquellos casos en los que hay una distribución normal, puede usarse un modelo paramétrico tal como prueba de la t o prueba de ANOVA; y cuando no hay una distribución normal entonces puede usarse un modelo no paramétrico tal como prueba de la U de Mann-Whitney U o prueba de Kruskal-Wallis.

45 En una realización particular, se usan los resultados de análisis de ROC, exactitud, especificidad, sensibilidad o una combinación de los mismos para describir el método de cribado de la divulgación. En particular, se usan para cuantificar cómo de bueno y fiable es el método.

50 Hay varios términos que se usan comúnmente junto con la descripción de sensibilidad, especificidad y exactitud. Son verdadero positivo (TP), verdadero negativo (TN), falso negativo (FN) y falso positivo (FP). Si se demuestra que una enfermedad está presente en un paciente, la prueba de cribado dada también indica la presencia de enfermedad, el resultado de la prueba se considera un verdadero positivo. De manera similar, si se demuestra que una enfermedad está ausente en un paciente, la prueba de cribado sugiere que la enfermedad también está ausente, el resultado de la prueba es un verdadero negativo (TN). Tanto el verdadero positivo como el verdadero negativo sugieren un resultado uniforme entre la prueba de cribado y el estado demostrado (también denominado patrón de verdad). Sin embargo, ninguna prueba médica es perfecta. Si la prueba de cribado indica la presencia de enfermedad en un paciente que realmente no tiene tal enfermedad, el resultado de la prueba es un falso positivo (FP). De manera similar, si el resultado de la prueba de cribado sugiere que la enfermedad está ausente en un paciente con enfermedad, el resultado de la prueba es un falso negativo (FN). Tanto el falso positivo como el falso negativo indican que los resultados de la prueba son los contrarios al estado real.

Sensibilidad, especificidad y exactitud se describen en cuanto a TP, TN, FN y FP.

65
$$\text{Sensibilidad} = \frac{\text{TP}}{\text{TP} + \text{FN}} = \frac{(\text{Número de evaluación de verdaderos positivos})}{(\text{número de toda la evaluación de positivos})}$$

Especificidad = $TN/(TN + FP)$ = (Número de evaluación de verdaderos negativos)/(número de toda la evaluación de negativos)

5 Exactitud = $(TN + TP)/(TN + TP + FN + FP)$ = (Número de evaluaciones correctas)/número de todas las evaluaciones).

10 En una realización preferida, el método de cribado de la presente divulgación diagnóstica, detecta de manera temprana, determina la recaída, determina el riesgo de desarrollar, o predice CCR y/o pólipos en un sujeto humano; o determina si debe realizarse una colonoscopia en dicho sujeto humano; o determina el pronóstico cuando dicho sujeto humano es un paciente al que se le ha diagnosticado CCR y/o pólipos; o guía una terapia en un paciente con CCR y/o pólipos de manera estadísticamente significativa con una sensibilidad y/o especificidad de al menos el 60%, al menos el 65%, al menos el 70%, al menos el 75%, al menos el 80%, al menos el 85%, al menos el 90%, al menos el 95% o preferiblemente el 100%.

15 Tal como sugieren las ecuaciones anteriores, el término “sensibilidad” tal como se usa en el presente documento se refiere a la proporción de verdaderos positivos que se identifican correctamente mediante una prueba de cribado. Muestra cómo de buena es la prueba en la detección de una enfermedad. La sensibilidad (“sens”) puede estar dentro del intervalo de 0 (0%) $< \text{sens} < 1$ (100%) y de manera ideal, el número de falsos negativos iguales a cero o próximos a ser iguales a cero y la sensibilidad igual a uno (100%) o próxima a ser igual a uno (100%).
20 Preferiblemente, tiene una sensibilidad del 70% al 90%, del 75% al 95%, del 80% al 95%, del 85% al 100%, o del 90% al 100%. Más preferiblemente, el método de la divulgación tiene valores de sensibilidad de al menos el 85%, tal como aproximadamente el 86%, el 87%, el 88%, el 89%, el 90%, el 91%, el 92%, el 93%, el 94%, el 95%, el 96%, el 97%, el 98%, el 99% o el 100%.

25 El término “especificidad” tal como se usa en el presente documento se refiere a la proporción de los verdaderos negativos identificados correctamente mediante una prueba de cribado. Sugiere cómo de buena es la prueba en la identificación de un estado normal (negativo). La especificidad (“spec”) puede estar dentro del intervalo de 0 (0%) $< \text{spec} < 1$ (100%) y de manera ideal, el número de falsos positivos es igual a cero o próximo a ser igual a cero y la especificidad es igual a uno (100%) o es próxima a ser igual a uno (100%). Preferiblemente, tiene una especificidad del 70% al 90%, del 75% al 95%, del 80% al 95%, del 85% al 100%, o del 90% al 100%. Más preferiblemente, el método de la divulgación tiene valores de especificidad de al menos el 85%, tal como aproximadamente el 86%, el 87%, el 88%, el 89%, el 90%, el 91%, el 92%, el 93%, el 94%, el 95%, el 96%, el 97%, el 98%, el 99% o el 100%.

35 El término “exactitud” tal como se hace referencia en el presente documento es la proporción de resultados verdaderos, o bien verdadero positivo o bien verdadero negativo, en una población. Mide el grado de veracidad de una prueba de cribado en un estado, es decir, cómo de correcta es la determinación y exclusión de un estado dado. La exactitud (“acc”) puede estar en el intervalo de 0 (0%) $< \text{acc} < 1$ (100%) y de manera ideal, el número de falsos positivos es igual a cero o próximo a ser igual a cero y la exactitud es igual a uno (100%) o es próxima a ser igual a uno (100%). Preferiblemente, la exactitud del método de la divulgación es de al menos el 60%, al menos el 65%, al menos el 70%, al menos el 75%, al menos el 80%, al menos el 85%, al menos el 90%, al menos el 95% o preferiblemente el 100%. En una realización preferida, tiene una exactitud del 70% al 90%, del 75% al 95%, del 80% al 95%, del 85% al 100%, o del 90% al 100%. Preferiblemente, el método de la divulgación tiene valores de exactitud de al menos el 85%, tal como aproximadamente el 86%, el 87%, el 88%, el 89%, el 90%, el 91%, el 92%, el 93%, el 94%, el 95%, el 96%, el 97%, el 98%, el 99% o el 100%.

45 La “curva ROC” es una presentación gráfica de la relación entre tanto la sensibilidad como especificidad y ayuda a decidir el modelo óptimo mediante la determinación del mejor umbral (punto de corte óptimo) para la prueba de cribado. El área bajo la curva ROC (AUC) proporciona un modo de medir la exactitud de una prueba de cribado. Preferiblemente, los valores de intervalo de AUC del método de la divulgación son desde 0,6 hasta 1, más preferiblemente de 0,7 a 1, estando los valores más preferidos en el intervalo de 0,8 a 1, más preferiblemente de 0,9 a 1. En una realización particular, AUC es desde 0,6 hasta 0,8, tal como desde 0,6 hasta 0,75 o de 0,65 a 0,75.

50 Además, la sensibilidad, especificidad y exactitud son proporciones, por tanto pueden calcularse los intervalos de confianza conformes usando métodos convencionales para proporciones bien conocidos en la técnica. Dos tipos de intervalos de confianza del 95% se definen generalmente alrededor de la proporciones. El intervalo de confianza exacto se define usando distribución binomial para alcanzar una estimación exacta. El intervalo de confianza asintótico se calcula suponiendo una aproximación normal de la distribución de la muestra. Un experto en la técnica sabrá cómo definir el intervalo de confianza apropiado. La elección de un tipo u otro de intervalo de confianza dependerá normalmente de si la proporción de muestra es una buena aproximación a una distribución normal.

60 En una realización particular del método de la divulgación, antes de la cuantificación de dichas secuencias bacterianas, se extrae ADN de la muestra de heces. Se conocen varios métodos de extracción de ADN a partir de muestras fecales, todos estos métodos se basan en la alteración química o mecánica, lisis usando detergentes, o una combinación de estos enfoques (Kennedy A. *et al.*, PLoS One, 2014;9(2):e88982). Se conocen métodos para la extracción de ADN bacteriano en muestras fecales por ejemplo de M Corist *et al.*, Journal of Microbiological Methods, 2002; 50(2):131-139, Whitney D *et al.*, Journal of Molecular Diagnostics, American Society for Investigative

Pathology, 2004;6(4):386-395 y documento WO2003/068788.

Los métodos preferidos usan una combinación de alteración mecánica, tal como extracción de golpeo de perlas a alta velocidad, lisis química y una etapa de purificación final, preferiblemente usando una columna de membrana de sílice tal como las incluidas en los kits de extracción de ADN disponibles comercialmente “procedimiento de extracción de ADN MobioPowerSoil®” (Mo-Bio Laboratories Inc.), kit FastDNA® SPIN para procedimiento de suelo (MP biomedical) y NucleoSpin® Soil (Macherey-Nagel GmbH & Co. KG). La presencia de inhibidores de PCR en los extractos de ADN de muestras fecales tales como bilirrubinas, sales biliares e hidratos de carbono complejos es una de las dificultades a las que se enfrenta la determinación de biomarcadores de ADN en extractos de ADN de heces (Fleckna *et al.*, Mol Cell Probes, 2007;21(4):282-7). Métodos de extracción de ADN preferidos son aquellos que proporcionan extractos fecales con una baja cantidad de inhibidores de PCR, tal como menos del 5%, preferiblemente menos del 2%, más preferiblemente menos del 1%, incluso más preferiblemente menos del 0,5%, tal como menos del 0,25%, el 0,1%, el 0,05% o el 0,01%.

En una realización particular, el método de la divulgación comprende:

(a) determinar mediante PCR cuantitativa a partir de dicha muestra de heces humanas los niveles de al menos una secuencia bacteriana de ADNr 16S seleccionada del grupo que consiste en SEQ ID NO: 1, SEQ ID NO: 4, SEQ ID NO: 7 y SEQ ID NO: 10, en la que los niveles de SEQ ID NO: 1 se determinan usando cebadores con al menos una identidad del 90% con respecto a SEQ ID NO: 2-13, los niveles de SEQ ID NO: 4 se determinan usando cebadores con al menos una identidad del 90% con respecto a SEQ ID NO: 5-6, los niveles de SEQ ID NO: 7 se determinan usando cebadores con al menos una identidad del 90% con respecto a SEQ ID NO: 8-9 y los niveles de SEQ ID NO: 10 se determinan usando cebadores con al menos una identidad del 90% con respecto a SEQ ID NO: 11-12; y/o

(b) determinar mediante PCR cuantitativa a partir de dicha muestra de heces humanas los niveles de al menos dos secuencias bacterianas de ADNr 16S seleccionadas del grupo que consiste en SEQ ID NO: 1, SEQ ID NO: 4, SEQ ID NO: 7 y SEQ ID NO: 10 tal como se menciona en la etapa a), y determinar al menos una de las razones seleccionadas del grupo que consiste en SEQ ID NO: 10/SEQ ID NO: 4, SEQ ID NO: 7/SEQ ID NO: 4, SEQ ID NO: 4/SEQ ID NO: 1, SEQ ID NO: 7/SEQ ID NO: 1; SEQ ID NO: 7/SEQ ID NO: 10 y SEQ ID NO: 10/SEQ ID NO: 1.

En una realización preferida, el método de cribado para CCR y/o pólipos adenomatosos a partir de una muestra de heces humanas de la divulgación, comprende:

(a) determinar mediante PCR cuantitativa a partir de dicha muestra de heces humanas la concentración de al menos una secuencia bacteriana de ADNr 16S seleccionadas del grupo que consiste en SEQ ID NO: 1, SEQ ID NO: 4, SEQ ID NO: 7 y SEQ ID NO: 10, en la que la concentración de SEQ ID NO: 1 se determina usando cebadores con al menos una identidad del 90% con respecto a SEQ ID NO: 2-13, la concentración de SEQ ID NO: 4 se determina usando cebadores con al menos una identidad del 90% con respecto a SEQ ID NO: 5-6, la concentración de SEQ ID NO: 7 se determina usando cebadores con al menos una identidad del 90% con respecto a SEQ ID NO: 8-9 y la concentración de SEQ ID NO: 10 se determina usando cebadores con al menos una identidad del 90% con respecto a SEQ ID NO: 11-12,

(b) opcionalmente, determinar mediante PCR cuantitativa a partir de dicha muestra de heces humanas los niveles de al menos dos secuencias bacterianas de ADNr 16S seleccionadas del grupo que consiste en SEQ ID NO: 1, SEQ ID NO: 4, SEQ ID NO: 7 y SEQ ID NO: 10 tal como se menciona en la etapa a), y determinar al menos una de las razones de concentraciones seleccionadas del grupo que consiste en SEQ ID NO: 10/SEQ ID NO: 4, SEQ ID NO: 7/SEQ ID NO: 4, SEQ ID NO: 4/SEQ ID NO: 1, SEQ ID NO: 7/SEQ ID NO: 1; SEQ ID NO: 7/SEQ ID NO: 10 y SEQ ID NO: 10/SEQ ID NO: 1 y

(c) diagnosticar, determinar el riesgo o determinar la recaída de CCR y/o pólipos adenomatosos a partir de la concentración de al menos una de las secuencias de la etapa (a) y/o el valor de al menos una razón obtenida en la etapa (b).

En otra realización del método de la divulgación, dichos cebadores con al menos una identidad del 90% con respecto a SEQ ID NO: 2-13 son cebadores con al menos una identidad del 95% con respecto a SEQ ID NO: 2-13, dichos cebadores con al menos una identidad del 90% con respecto a SEQ ID NO: 5-6 son cebadores con al menos una identidad del 95% con respecto a SEQ ID NO: 5-6, dichos cebadores con al menos una identidad del 90% con respecto a SEQ ID NO: 8-9 son cebadores con al menos una identidad del 95% con respecto a SEQ ID NO: 8-9 y dichos cebadores con al menos una identidad del 90% con respecto a SEQ ID NO: 11-12 son cebadores con al menos una identidad del 95% con respecto a SEQ ID NO: 11-12.

En una realización particularmente preferida, dichos cebadores con al menos una identidad del 95% con respecto a SEQ ID NO: 2-13 son cebadores identificados por SEQ ID NO: 2-13, dichos cebadores con al menos una identidad del 95% con respecto a SEQ ID NO: 5-6 son cebadores identificados por SEQ ID NO: 5-6, dichos cebadores con al menos una identidad del 95% con respecto a SEQ ID NO: 8-9 son cebadores identificados por SEQ ID NO: 8-9 y dichos cebadores con al menos una identidad del 95% con respecto a SEQ ID NO: 11-12 son cebadores

identificados por SEQ ID NO: 11-12.

En la presente solicitud, una secuencia problemática tiene una identidad del 95% con respecto a una secuencia determinada si el 95% de los residuos de la secuencia problemática son idénticos a los residuos de la secuencia determinada.

En una realización preferida, el método de la presente divulgación es un método para el diagnóstico de cáncer colorrectal y se diagnostica cáncer colorrectal en la etapa (c) si el valor de Ct de SEQ ID NO: 1 está por encima del valor de punto de corte de Ct de 37,2, o si el valor de Ct de SEQ ID NO: 4 está por encima del valor de punto de corte de Ct de 16,56, o si el valor de Ct de SEQ ID NO: 7 está por encima del valor de punto de corte de Ct de 22,97, o si el valor de Ct de SEQ ID NO: 10 está por encima del valor de punto de corte de Ct de 19,24, o si el valor de la razón SEQ ID NO: 10/SEQ ID NO: 4 está por debajo del valor de punto de corte de 1,16, o si el valor de la razón SEQ ID NO: 7/SEQ ID NO: 4 está por debajo del valor de punto de corte de 1,40, o si el valor de la razón SEQ ID NO: 4/SEQ ID NO: 1 está por encima del valor de punto de corte de 0,44, o si el valor de la razón SEQ ID NO: 7/SEQ ID NO: 1 está por encima del valor de punto de corte de 0,62, o si el valor de la razón y SEQ ID NO: 10/SEQ ID: 1 está por encima del valor de punto de corte de 0,52, o si la SEQ ID NO: 10/SEQ ID: 7 está por encima del valor de punto de corte de 1,2.

En otra realización preferida, el método de la presente divulgación es un método para diagnosticar el riesgo de desarrollar cáncer colorrectal, en el que se diagnostica un riesgo de desarrollar cáncer colorrectal en la etapa (c) si el valor de Ct de SEQ ID NO: 1 está por encima del valor de punto de corte de Ct de 36,45, o si el valor de Ct de SEQ ID NO: 4 está por encima del valor de punto de corte de Ct de 14,02, o si el valor de Ct de SEQ ID NO: 7 está por encima del valor de punto de corte de Ct de 24,76, o si el valor de Ct de SEQ ID NO: 10 está por encima del valor de punto de corte de Ct de 21,42, o si el valor de la razón SEQ ID NO: 10/SEQ ID NO: 4 está por debajo del valor de punto de corte de 1,56, o si el valor de la razón SEQ ID NO: 7/SEQ ID NO: 4 está por debajo del valor de punto de corte de 1,78, o si el valor de la razón SEQ ID NO: 4/SEQ ID NO: 1 está por encima del valor de punto de corte de 0,38, o si el valor de la razón SEQ ID NO: 7/SEQ ID NO: 1 está por encima del valor de punto de corte de 0,68, o si el valor de la razón SEQ ID NO: 10/SEQ ID 1 está por encima del valor de punto de corte de 0,59, o si la SEQ ID NO: 10/SEQ ID: 7 está por encima del valor de punto de corte de 1,16.

En otra realización, el método de la presente divulgación comprende además detectar y/o cuantificar uno o más biomarcadores moleculares que se sabe que son indicadores de CCR. Preferiblemente, se usa un panel de marcadores múltiples. Dicho panel de marcadores múltiples puede comprender la determinación de mutaciones puntuales presentes en uno o más de los siguientes genes: APC, K-ras, BRAF, p53, NRAS, PIK3CA, CDK8, CMYC, CCNE1, CTNNB1, NEU (HER2), MYB, FBXW7, PTEN, SMAD4, SMAD2, SMAD3, TGFβIIIR, TCF7L2, ACVR2 y BAX (Fearon, E. R., Annu. Rev. Pathol. Mech. Dis., 2011;6:479-507). Otros marcadores de CCR conocidos que pueden determinarse son metilación de isla de CpG, la cantidad de ADN humano por unidad de peso de las heces, y la cantidad de ADN total por unidad de peso de las heces.

En una realización adicional, el método de la presente divulgación comprende además almacenar los resultados de la determinación en un medio legible por ordenador. Tal como se usa en el presente documento, "un medio legible por ordenador" puede ser cualquier aparato que puede incluir, almacenar, comunicar, propagar o transportar los resultados de la determinación del método. El medio puede ser un sistema (o aparato o dispositivo) electrónico, magnético, óptico, electromagnético, infrarrojo o semiconductor o un medio de propagación.

Un segundo aspecto, la divulgación se refiere al uso de al menos una secuencia bacteriana de ADNr 16S seleccionada de la lista que consiste en SEQ ID NO: 1, SEQ ID NO: 4, SEQ ID NO: 7 y SEQ ID NO: 10 o la secuencia de ARNr complementaria de la misma como biomarcador para diagnosticar, detectar de manera temprana, determinar la recaída, determinar el riesgo de desarrollar, o predecir CCR y/o pólipos en un sujeto humano; o determinar si debe realizarse una colonoscopia en dicho sujeto humano; o determinar el pronóstico en el que dicho sujeto humano es un paciente al que se le ha diagnosticado CCR y/o pólipos; o guiar una terapia en un paciente con CCR y/o pólipos en el que dichas secuencias bacterianas se cuantifican a partir de una muestra de heces de dicho sujeto humano.

Preferiblemente, dicha secuencia de ácido nucleico es una secuencia de ADNr 16S seleccionada de la lista que consiste en SEQ ID NO: 1, SEQ ID NO: 4, SEQ ID NO: 7 y SEQ ID NO: 10.

El término "biomarcador" tal como se usa en el presente documento se refiere a marcadores de enfermedad que son normalmente sustancias encontradas en una muestra corporal que pueden medirse fácilmente. La cantidad medida puede correlacionarse con una fisiopatología de la enfermedad subyacente, tal como presencia o ausencia de CCR y/o pólipos, o probabilidad de CCR y/o pólipos en el futuro. En pacientes que reciben tratamiento para su estado, la cantidad medida también se correlacionará con el grado de respuesta a la terapia.

En un tercer aspecto, la divulgación se refiere a un kit que comprende:

a. un reactivo seleccionado del grupo que consiste en:

i) sondas de ácido nucleico capaces de hibridarse específicamente con al menos una secuencia de ADNr 16S seleccionada de la lista que consiste en SEQ ID NO: 1, SEQ ID NO: 4, SEQ ID NO: 7 y SEQ ID NO: 10 o la secuencia de ARNr complementaria de la misma; o

ii) un par de cebadores de ácido nucleico capaces de amplificar específicamente al menos una secuencia de ADNr 16S seleccionada de la lista que consiste en SEQ ID NO: 1, SEQ ID NO: 4, SEQ ID NO: 7 y SEQ ID NO: 10; y

b. instrucciones para cuantificar los niveles de una o más de dichas secuencias a partir de una muestra de heces humanas según un método de la presente divulgación.

El término “sonda” tal como se usa en el presente documento se refiere a ácidos nucleicos sintéticos o producidos de manera biológica, entre 10 y 285 pares de bases de longitud que contienen secuencias de nucleótidos específicas que permiten hibridación específica y preferencial en condiciones predeterminadas con secuencias diana de ácido nucleico, y opcionalmente contienen un resto para detección o para potenciar el rendimiento de un ensayo. Generalmente es necesario un mínimo de diez nucleótidos con el fin de obtener estadísticamente especificidad y para formar productos de hibridación estables, y un máximo de 285 nucleótidos generalmente representa un límite superior para la longitud en el que los parámetros de reacción pueden ajustarse fácilmente para determinar secuencias de apareamiento erróneo e hibridación preferencial. Las sondas pueden contener opcionalmente determinados constituyentes que contribuyen a su funcionamiento adecuado u óptimo en determinadas condiciones de ensayo. Por ejemplo, las sondas pueden modificarse para mejorar su resistencia a la degradación de nucleasas (por ejemplo, mediante ocupación de los centros activos), para llevar ligandos de detección (por ejemplo, fluoresceína) o para facilitar su captura sobre un soporte sólido (por ejemplo, “colas” de polidesoxiadenosina).

El término “cebadores” tal como se usa en el presente documento se refiere a oligonucleótidos que pueden usarse en un método de amplificación, tal como una reacción en cadena de la polimerasa (“PCR”), para amplificar una secuencia de nucleótidos. Los cebadores se diseñan basándose en la secuencia de polinucleótido de una secuencia diana particular, por ejemplo, una secuencia de ADNr 16S específica.

El diseño y la validación de cebadores y sondas se conocen bien en la técnica. Para métodos de PCR en tiempo real cuantitativa, véase por ejemplo Rodríguez A *et al.* (Methods Mol Biol., 2015, 1275:31-56).

“Específico” tal como se usa en el presente documento significa que una secuencia de nucleótidos se hibridará con/amplificará una secuencia diana predeterminada y no se hibridará con/amplificará sustancialmente una secuencia no diana en las condiciones del ensayo, generalmente se usan condiciones rigurosas.

“Hibridación” tal como se usa en el presente documento se refiere a un proceso mediante el cual, en condiciones de reacción predeterminadas, se permite que dos hebras parcial o completamente complementarias de ácido nucleico se junten de manera antiparalela para formar un ácido nucleico bicatenario con enlaces de hidrógeno específicos y estables, siguiendo las reglas explícitas relativas a las que las bases de ácido nucleico pueden emparejarse entre sí.

“Hibridación sustancial” significa que la cantidad de hibridación observada será tal que al observar los resultados se consideraría que el resultado es positivo con respecto a datos de hibridación en controles positivos y negativos. Los datos que se consideran “ruido de fondo” no son una hibridación sustancial.

“Condiciones de hibridación rigurosas” significa de aproximadamente 35°C a 65°C en una disolución salina de aproximadamente NaCl 0,9 molar. La rigurosidad también puede regirse por tales parámetros de reacción como la concentración y el tipo de especies iónicas presente en la disolución de hibridación, los tipos y concentraciones de agentes desnaturalizantes presentes, y la temperatura de hibridación. Generalmente, a medida que las condiciones de hibridación se vuelven más rigurosas, se prefieren sondas más largas si van a formarse híbridos estables. Como norma, la rigurosidad de las condiciones en las que tiene lugar la hibridación dictará determinadas características de las sondas preferidas que van a emplearse.

Preferiblemente, dicha secuencia de ácido nucleico es una secuencia de ADNr 16S seleccionada de la lista que consiste en SEQ ID NO: 1, SEQ ID NO: 4, SEQ ID NO: 7 y SEQ ID NO: 10.

En una realización particular, dicho kit se usa para diagnosticar, detectar de manera temprana, determinar la recaída, determinar el riesgo de desarrollar, o predecir CCR y/o pólipos en un sujeto humano; o para determinar si debe realizarse una colonoscopia en dicho sujeto humano; o para determinar el pronóstico en un paciente al que se le ha diagnosticado CCR y/o pólipos; o para guiar una terapia en un paciente con CCR y/o pólipos. En una realización preferida, el kit puede comprender además medios de extracción de ADN, medios para llevar a cabo la hibridación y/o amplificación, medios de detección, y/o uno o más recipientes para recoger y/o mantener la muestra biológica.

En una realización particular, la presente divulgación se refiere a un kit para diagnosticar cáncer colorrectal a partir

de una muestra de heces humanas que comprende al menos un par de cebadores de PCR seleccionados del grupo de pares de cebadores que consiste en SEQ ID NO: 2-13, SEQ ID NO: 5-6, SEQ ID NO: 8-9 y SEQ ID NO: 11-12.

5 En un cuarto aspecto, la presente divulgación se refiere a una secuencia de ácido nucleico seleccionada del grupo que consiste en SEQ ID NO: 2, SEQ ID NO: 13, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 11, y SEQ ID NO: 12; o con una identidad de al menos el 90% de los mismos, preferiblemente con una identidad del 95% de los mismos, más preferiblemente con una identidad del 96%, el 97%, el 98%, el 99% o el 100%. Preferiblemente, dichas secuencias se usan como cebadores en un ensayo de amplificación (tal como qPCR) o como sondas de ácido nucleico en un ensayo de hibridación para la cuantificación de al menos una secuencia de ADNr 16S seleccionada de la lista que consiste en SEQ ID NO: 1, SEQ ID NO: 4, SEQ ID NO: 7 y SEQ ID NO: 10.

15 En una realización particular, dicha secuencia de ácido nucleico es un par de cebadores de PCR seleccionados del grupo que consiste en SEQ ID NO: 2-13, SEQ ID NO: 5-6, SEQ ID NO: 8-9 y SEQ ID NO: 11-12. En otra realización particular, dicha secuencia de ácido nucleico es una sonda de ácido nucleico seleccionada del grupo que consiste en SEQ ID NO: 2, SEQ ID NO: 13, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 11, y SEQ ID NO: 12.

20 En un quinto aspecto, la divulgación se refiere al uso de una secuencia de ácido nucleico seleccionada del grupo que consiste en SEQ ID NO: 2, SEQ ID NO: 13, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 11, y SEQ ID NO: 12; o con una identidad de al menos el 90% de los mismos en un método de cribado de la divulgación, en el que se usan SEQ ID NO: 2 y/o SEQ ID NO: 13 para la cuantificación de SEQ ID NO: 1; se usan SEQ ID NO: 5 y/o SEQ ID NO: 6 para la cuantificación de SEQ ID NO: 4; se usan SEQ ID NO: 8 y/o SEQ ID NO: 9 para la cuantificación de SEQ ID NO: 7; y se usan SEQ ID NO: 11 y/o SEQ ID NO: 12 para la cuantificación de SEQ ID NO: 10.

25 Se entenderá que las realizaciones particulares descritas en los ejemplos se muestran a modo de ilustración y no como limitaciones de la invención.

30 Todas las publicaciones y solicitudes de patente mencionadas en la memoria descriptiva son indicativas del nivel de destreza de los expertos en la técnica a la que pertenece esta invención.

El uso de la palabra “un” o “una” puede significar “un/una”, pero también es compatible con el significado de “uno o más”, “al menos uno”, y “uno o más de uno.”

35 En la totalidad de esta solicitud, el término “aproximadamente” significa el valor indicado \pm 5% de su valor, preferiblemente el valor indicado \pm 2% de su valor, lo más preferiblemente el término “aproximadamente” significa exactamente el valor indicado (\pm 0%).

40 Ejemplos

Ejemplo 1.- Material y métodos

1. Muestras biológicas

45 Se obtuvieron muestras fecales de 27 pacientes a los que se les diagnosticó recientemente cáncer colorrectal (CCR) (fase 0-I), 24 pacientes portadores del síndrome de Lynch (L) y 19 individuos sanos (C). Todos los pacientes tuvieron una colonoscopia al menos 15 días antes de la recogida de muestras.

50 Todos ellos firmaron el correspondiente consentimiento informado. Los criterios de exclusión incluyeron tratamiento con antibióticos dentro de 1 mes antes del estudio y edad < 18 años.

2. Secuencias bacterianas de ADNr 16S:

55 Las secuencias B3, B10, B41, B46, B48 y B50 se corresponden con bandas de gel de electroforesis en gel de gradiente desnaturante (DGGE) aisladas previamente de aislados de bacterias no cultivadas.

- Secuencia bacteriana de ADNr 16S SEQ ID NO: 1 (B3), publicada en la base de datos del Archivo de Nucleótidos Europeo (ENA) EMBL-EBI como GQ411111.1;

60 - secuencia bacteriana de ADNr 16S SEQ ID NO: 4 (B10), publicada en la base de datos del Archivo de Nucleótidos Europeo (ENA) EMBL-EBI como GQ411118.1;

- secuencia bacteriana de ADNr 16S SEQ ID NO: 7 (B46), publicada en la base de datos del Archivo de Nucleótidos Europeo (ENA) EMBL-EBI como GQ411150.1;

65 - secuencia bacteriana de ADNr 16S SEQ ID NO: 10 (B48), publicada en la base de datos del Archivo de

Nucleótidos Europeo (ENA) EMBL-EBI como GQ411152.1;

- secuencia bacteriana de ADNr 16S B41, publicada en la base de datos del Archivo de Nucleótidos Europeo (ENA) EMBL-EBI como GQ411145.1; y

- secuencia bacteriana de ADNr 16S B50, publicada en la base de datos del Archivo de Nucleótidos Europeo (ENA) EMBL-EBI como GQ411154.1.

3. Recogida, conservación y almacenamiento de muestras

Se recogieron muestras fecales de 24 h en un recipiente para heces y se almacenaron a -20°C durante un mes y se transportaron a un congelador a -80°C en condiciones de la norma ISO 9001.

4. Procesamiento de muestras y extracción de ADN

Para la extracción de ADN 16S fecal se usó NucleoSpin® Soil (Macherey-Nagel).

En condiciones de seguridad biológicas clase II, se colocaron 200 mg-500 mg de muestra fecal en un tubo de perlas Nucleospin. Se añadieron 700 µl de SL1 y 150 µl de potenciador SX a un tubo de perlas.

5. Después de eso, se extrajo ADN y se purificó siguiendo las instrucciones del fabricante, y se eluyó con 50 UI de tampón de elución o Tris HCl 10 mM pH = 7,4.

6. Cálculo de la concentración de ADN

Se determinó la concentración de ADN del extracto mediante análisis fluorimétricos con el n.º de catálogo de fluorímetro Qubit® 2.0 Q32866 usando el kit de ensayo de amplio intervalo de ADNbc Quant-IT. Este método es muy selectivo para ADNbc con respecto a ARN y proporciona una cuantificación exacta en un intervalo desde 2 ng hasta 1.000 ng. Con en este intervalo, la fluorescencia está relacionada linealmente con la concentración de ADN. Se realizaron mediciones de ADN en 5 ul del extracto.

7. PCR en tiempo real cuantitativa (qPCR) de ADN extraído de muestras fecales

Se analizó el ADN de muestras fecales mediante PCR en tiempo real cuantitativa. En particular, se evaluó la cantidad de cuatro poblaciones bacterianas por muestra: B3, B10, B48 y B46. Las secuencias bacterianas se cuantificaron usando una PCR en tiempo real cuantitativa con un tinte de ADNbc fluorescente. En este trabajo, se ha usado la mezcla previa BRYT® de Promega (Madison, EE.UU.). Se amplificaron Eubacteria siguiendo el procedimiento descrito por Furet *et al.* (FEMS Microbiol Ecol., 2009;68(3):351-62).

Se expresaron las abundancias bacterianas para cada muestra como Ct normalizado frente a la concentración de ADN total. El Ct (umbral de ciclo) se define como el número de ciclos de q-PCR requeridos para que la señal fluorescente cruce el umbral. Los niveles de Ct son inversamente proporcionales al logaritmo de la concentración de ácido nucleico diana en la muestra (es decir, cuanto menor sea el nivel de Ct mayor será la cantidad de ácido nucleico diana en la muestra). Los ensayos en tiempo real se someten a 40 ciclos de amplificación.

8. Métodos de análisis estadístico

Se analizó la distribución estadística normal de los datos mediante la prueba de Kolmogorov-Smirnov. Según si hubo una distribución estadística normal de los datos o no, se usó una prueba estadística adecuada para comparar los siguientes grupos. Los grupos analizados fueron: CCR frente a sano; CCR frente a Lynch frente a sano; CCR + Lynch frente a sano; y CCR frente a Lynch-alto riesgo frente a Lynch-bajo riesgo frente a sano. Para estos grupos, las variables analizadas fueron:

- cuantificación de cuatro secuencias bacterianas expresadas en Ct;

- razón de la cuantificación de cuatro secuencias bacterianas;

- peso; y

- concentración de ADN.

Usando una regresión binaria y análisis de ROC, se determinaron la correlación de la cuantificación de secuencia(s) bacteriana(s) entre el estado sano y con CCR y el riesgo (en conjunto y separado en dos condiciones de riesgo) y la sensibilidad y especificidad de la prueba.

No se observó diferencia entre grupos en ADN y peso. Se encontró que la cantidad de Eubacteria (Eub) era

diferente entre Lynch y CCR + Lynch y donantes sanos (C), pero se consideró normal porque el estado de la enfermedad implica un cambio del núcleo filogenético, que incluye bacterias con menor número de copias de ADN. No implica una disminución de bacterias totales (véase por ejemplo, Sobhani *et al.*, PLoS One, 2011,27;6(1):e16393). Por este motivo, no se usó cuantificación de Eub como referencia interna con propósitos de normalización.

Ejemplo 2. Diseño y validación de cebadores

Se diseñó el cebador para la cuantificación de los marcadores bacterianos B3, B10, B46, B48, B41 y B50 en biopsia a partir de los análisis comparativos con secuencias previamente obtenidas de grupos filogenéticos usando herramientas bioinformáticas: ClustalX del European Bioinformatics Institute (www.ebi.ac.uk), Netprimer (Premier Biosoft) y PrimerExpress (Life Technologies-Thermo Fischer).

El sistema de detección escogido fue SybrGreen®. Se descartó un conjunto de cebadores con menos de 3 posiciones diferentes con respecto a grupos próximos. También se descartó el conjunto de cebadores con valores de Tm en curvas de disociación diferentes de los esperados. La validación de cebadores se realizó en una muestra de biopsia y muestra fecal. Se determinó la curva de disociación con el fin de analizar el rendimiento del cebador.

En las figuras 8a a 13b se muestran los gráficos de amplificación y curvas de disociación para la validación de los cebadores diseñados para la amplificación de B3, B10, B46, B48, B41 y B50, respectivamente. Se seleccionaron cebadores que muestren una curva de disociación adecuada (pico único para cada conjunto de cebadores) para la cuantificación bacteriana de ADNr 16S en muestras fecales. En particular, se seleccionaron cebadores que amplifican B3, B10, B46 y B48, estos se especifican en la tabla 9 a continuación. Se descartaron B51 y B40 debido a la falta de especificidad (picos múltiples para cada conjunto de cebadores en curvas de fusión).

B3 directo (SEQ ID NO: 2)	GGAGGCCTTCGGGTCGTAA
B3 inverso (SEQ ID NO: 13)	AGGTTCCGGGGGCTTCGG
B10 directo (SEQ ID NO: 5)	CAACAAGGTAAGTGACGGC
B10 inverso (SEQ ID NO: 6)	CGCCTACCTGTGCACTACTC
B46 directo (SEQ ID NO: 8)	TCCACGTAAGTCACAAGCG
B46 inverso (SEQ ID NO: 9)	CGCCTACCTGTGCACTACTC
B48 directo (SEQ ID NO: 11)	GTACGGGGAGCAGCAGTG
B48 inverso (SEQ ID NO: 12)	GCACTCTAGATGCACAGTTCC

Tabla 9. Pares de cebadores seleccionados para la amplificación de qPCR de secuencias bacterianas B3, B10, B46 y B48 de ADNr 16S en muestras fecales.

Ejemplo 3: Análisis de biomarcadores de ADNr 16S en pacientes con cáncer colorrectal frente a sujetos sanos

Se ha analizado un total de 16 muestras de heces en 7 controles y 9 pacientes con cáncer colorrectal. La cuantificación de cada uno de los marcadores de bacterias citados en el ejemplo 1 en muestras de heces analizadas se expresa en valores de Ct. El Ct (umbral de ciclo) se define como el número de ciclos de q-PCR requeridos para que la señal fluorescente cruce el umbral. Los niveles de Ct son inversamente proporcionales al logaritmo de la concentración de ácido nucleico diana en la muestra (es decir, cuanto menor sea el nivel de Ct mayor será la cantidad de ácido nucleico diana en la muestra). Los ensayos en tiempo real se someten a 40 ciclos de amplificación. Los resultados obtenidos se muestran a continuación, en las tablas 1 y 2. La tabla 1 representa valores absolutos de Ct y la tabla 2 representa razones de Ct. Se aplicó prueba de la t de dos muestras.

Tabla 1. Valores absolutos de Ct

ID	Grupo	Estado	B48	B46	B10	B3
F10	CCR	1	18,59	22,78	16,33	33,64
F11	CCR	1	20,33	26,45	20,37	41,57
F3	CCR	1	23,52	26,2	19,8	38,45
F4	CCR	1	16,41	20,87	14,88	
F5	CCR	1	19,04	24	17,61	35,68
F6	CCR	1	23,22	27,67	21,82	
F16	CCR	1		20,75	14,35	
F7	CCR	1	23,19	28,2	22,34	
F8	CCR	1	18,82	22,04	15,61	

ES 2 734 143 T3

F12	C	0	17,55	20,9	14,95	
F13	C	0	19,36	24,78	19,22	
F14	C	0	19,46	21,44	15,07	37,34
F15	C	0	17,55	19,85	12,84	36,98
F9	C	0	17,61	21,8	15,41	35,66
F1	C	0	18,09	21,97	14,79	38,3
F2	C	0	17,07	20,57	12,76	
Media	CCR		20,39	24,329	18,12	37,34
	C		18,099	21,616	15,01	37,07
Desviación estándar	CCR		2,65	2,89	3,04	3,44
	C		0,94	1,58	2,15	1,09
Valor de p	CCR frente a C		0,025	0,021	0,019	0,444
Punto medio (valores de punto de corte)			19,244	22,972	16,56	37,2
% de sensibilidad			66,67	83,33	83,33	50
% de especificidad			55,56	60	60	50
Exactitud*			60	68,75	68,75	50

* La exactitud se determina como $(TN + TP)/(TN + TP + FN + FP)$

Tabla 2. Razones de Ct

5

ID	Grupo	Estado	B48/B10	B10/B3	B46/B10	B46/B48	B46/B3	B48/B3
F10	CCR	1	1,14	0,49	1,39	1,23	0,68	0,55
F11	CCR	1	1	0,49	1,3	1,3	0,64	0,49
F3	CCR	1	1,19	0,51	1,32	1,11	0,68	0,61
F4	CCR	1	1,1		1,4	1,27		
F5	CCR	1	1,08	0,49	1,36	1,26	0,67	0,53
F6	CCR	1	1,06		1,27	1,19		
F16	CCR	1			1,45			
F7	CCR	1	1,04		1,26	1,22		
F8	CCR	1	1,21		1,41	1,17		
F12	C	0	1,17		1,4	1,19		
F13	C	0	1,01		1,29	1,28		
F14	C	0	1,29	0,4	1,42	1,1	0,57	0,52
F15	C	0	1,37	0,35	1,55	1,13	0,54	0,47
F9	C	0	1,14	0,43	1,41	1,24	0,61	0,49
F1	C	0	1,22	0,39	1,49	1,21	0,57	0,47
F2	C	0	1,34		1,61	1,21		
Media	CCR		1,1	0,5	1,35	1,22	0,67	0,55
	C		1,22	0,39	1,45	1,19	0,57	0,49
Desviación estándar	CCR		0,07	0,01	0,07	0,06	0,02	0,05
	C		0,13	0,04	0,11	0,06	0,03	0,02
Valor de p	CCR		0,02	0,001	0,018	0,224	0,001	0,045

	frente a C						
Punto medio (valores de punto de corte)		1,16123	0,4441	1,40243	1,20669	0,62	0,52
% de sensibilidad		75	100	77,78	55,56	100,00	75,00
% de especificidad		71,43	100	71,43	42,86	100,00	75,00
Exactitud*		73,33	100	75	50	100,00	75,00

* La exactitud se determina como $(TN + TP)/(TN + TP + FN + FP)$

5 La sensibilidad y especificidad del diagnóstico de cáncer colorrectal es del 75 y el 71% para la razón SEQ ID NO: 10/SEQ ID NO: 4; el 77% y el 71%, respectivamente, para la razón SEQ ID NO: 7/SEQ ID NO: 4; el 100% y el 100%, respectivamente, para la razón SEQ ID NO: 4/SEQ ID NO: 1; el 100% y el 100%, respectivamente, para la razón SEQ ID NO: 7/SEQ ID NO: 1; y el 75% y el 75%, respectivamente, para la razón SEQ ID NO: 10/SEQ ID NO: 1.

10 Las figuras 1, 2 y 3 son representaciones gráficas de razones de valores de Ct (SEQ ID NO: 10/SEQ ID NO: 4, SEQ ID NO: 7/SEQ ID NO: 4 y SEQ ID NO: 4/SEQ ID NO: 1).

Las figuras 4 y 5 son representaciones gráficas de los absolutos de Ct (SEQ ID NO: 4, SEQ ID NO: 7).

15 Ejemplo 4. Análisis de biomarcadores de ADNr 16S en pacientes con síndrome de Lynch: CCR frente a alto riesgo (con pólipos) frente a bajo riesgo (sin pólipos)

El objetivo de este análisis es determinar el valor predictivo de los biomarcadores para detectar el riesgo de cáncer colorrectal antes de los signos clínicos.

20 Se analizó un total de 8 individuos con síndrome de Lynch (con riesgo genético aumentado de desarrollar cáncer colorrectal). Todos los individuos tuvieron una colonoscopia, en el periodo máximo de un año, antes de la recogida de la muestra fecal. Según un examen endoscópico 4 de ellos tuvieron un pólipo maligno (denominados alto riesgo) y 4 no tuvieron ningún pólipo (denominados bajo riesgo).

25 La cuantificación de cada uno de los marcadores de bacterias citados en el ejemplo 1 en muestras de heces humanas analizadas se determinó usando los cebadores descritos en el ejemplo 2 y se cuantificaron en valores de Ct.

30 Los resultados obtenidos se muestran a continuación, en las tablas 3 y 4. La tabla 3 representa valores absolutos de Ct y la tabla 4 representa razones de Ct. Se aplicó una prueba de la t de dos muestras.

Tabla 3. Valores absolutos de Ct

ID	Grupo	Estado	B48	B46	B10	B3
F10	CCR	1	18,59	22,78	16,33	33,64
F11	CCR	1	20,33	26,45	20,37	41,57
F3	CCR	1	23,52	26,20	19,80	38,45
F4	CCR	1	16,41	20,87	14,88	
F5	CCR	1	19,04	24,00	17,61	35,68
F6	CCR	1	23,22	27,67	21,82	
F16	CCR	1		20,75	14,35	
F7	CCR	1	23,19	28,20	22,34	
F8	CCR	1	18,82	22,04	15,61	
MIL1	Alto riesgo	2	22,58	25,87	14,56	35,56
MIL2	Alto riesgo	2	24,53	25,79	14,73	37,06
MIL5	Alto riesgo	2	22,56	24,57	14,16	38,44
MIL4	Alto riesgo	2	20,00	29,90	18,41	37,99
MIL6	Bajo riesgo	3	20,15	24,90	15,20	37,58

ES 2 734 143 T3

MIL3	Bajo riesgo	3	18,11	20,45	10,34	31,16
MIL7	Bajo riesgo	3	22,81	24,70	13,35	41,40
MIL8	Bajo riesgo	3	20,60	21,88	11,38	32,42
Media	CCR		20,39	24,33	18,12	37,34
	Alto riesgo		22,42	26,53	15,46	37,26
	Bajo riesgo		20,42	22,98	12,57	35,64
Desviación estándar	CCR		2,65	2,89	3,04	3,44
	Alto riesgo		1,86	2,32	1,98	1,27
	Bajo riesgo		1,93	2,18	2,15	4,74
Valor de p	CCR frente a alto riesgo		0,10	0,10	0,070	0,48
	CCR frente a bajo riesgo		0,49	0,21	0,004	0,29
	Alto riesgo frente a bajo riesgo		0,09	0,03	0,047	0,27
Punto medio (valores de punto de corte) alto riesgo frente a bajo riesgo			21,42	24,76	14,02	36,45
% de sensibilidad alto riesgo frente a bajo riesgo			75,00	75,00	80,00	60,00
% de especificidad% alto riesgo frente a bajo riesgo			75,00	75,00	100,00	66,67
Exactitud alto riesgo frente a bajo riesgo			75,00	75,00	87,50	62,50

Tabla 4. Razones de Ct

ID	Grupo	Estado	B48/B10	B10/B3	B46/B10	B46/B48	B46/B3	B48/B3
F10	CCR	1	1,14	0,49	1,39	1,23	0,68	0,55
F11	CCR	1	1,00	0,49	1,30	1,30	0,64	0,49
F3	CCR	1	1,19	0,51	1,32	1,11	0,68	0,61
F4	CCR	1	1,10		1,40	1,27		
F5	CCR	1	1,08	0,49	1,36	1,26	0,67	0,53
F6	CCR	1	1,06		1,27	1,19		
F 16	CCR	1			1,45			
F7	CCR	1	1,04		1,26	1,22		
F8	CCR	1	1,21		1,41	1,17		
MIL1	Alto riesgo	2	1,55	0,41	1,78	1,15	0,73	0,63
MIL2	Alto riesgo	2	1,67	0,40	1,75	1,05	0,70	0,66
MIL5	Alto riesgo	2	1,59	0,37	1,74	1,09	0,64	0,59
MIL4	Alto riesgo	2	1,09	0,48	1,62	1,49	0,79	0,53
MIL6	Bajo	3	1,33	0,40	1,64	1,24	0,66	0,54

	riesgo							
MIL3	Bajo riesgo	3	1,75	0,33	1,98	1,13	0,66	0,58
MIL7	Bajo riesgo	3	1,71	0,32	1,85	1,08	0,60	0,55
MIL8	Bajo riesgo	3	1,81	0,35	1,92	1,06	0,67	0,64
Media	CCR		1,10	0,50	1,35	1,22	0,67	0,55
	Alto riesgo		1,47	0,41	1,72	1,20	0,71	0,60
	Bajo riesgo		1,65	0,35	1,85	1,13	0,65	0,58
Desviación estándar	CCR		0,07	0,01	0,07	0,06	0,02	0,05
	Alto riesgo		0,26	0,05	0,07	0,20	0,06	0,06
	Bajo riesgo		0,22	0,04	0,15	0,08	0,03	0,04
Valor de p	CCR frente a alto riesgo		0,001491	0,009755	0,000001	0,38	0,11	0,10
	CCR frente a bajo riesgo		0,000029	0,000159	0,000002	0,02	0,19	0,21
	Alto riesgo frente a bajo riesgo		0,172991	0,044382	0,087799	0,28	0,06	0,25
Punto medio (valores de punto de corte) alto riesgo frente a bajo riesgo			1,56	0,38	1,78	1,16	0,68	0,59
% de sensibilidad alto riesgo frente a bajo riesgo			66,67	75,00	80,00	50,00	75,00	75,00
Especificidad% alto riesgo frente a bajo riesgo			60,00	75,00	100,00	50,00	75,00	75,00
Precisión alto riesgo frente a bajo riesgo			62,50	75,00	87,50	50,00	75,00	75,00

5 Los marcadores de razones microbiológicas fueron capaces de discernir el grupo de individuos con alto riesgo de desarrollar cáncer colorrectal (pacientes con síndrome de Lynch y pólipos). La sensibilidad y especificidad en la detección de la población de alto riesgo de cáncer colorrectal fueron del 80% y el 100% para la SEQ ID NO: 4; el 75% y el 75%, respectivamente, para la razón SEQ ID NO: 4/SEQ ID NO: 1; el 80% y el 100%, respectivamente, para la razón SEQ ID NO: 7/SEQ ID NO: 4; el 75% y el 75%, respectivamente, para la razón SEQ ID NO: 7/SEQ ID NO: 1; y el 75% y el 75% para la razón SEQ ID NO: 10/SEQ ID NO: 1.

10 La figura 6 es una representación gráfica de la razón de valores de Ct (SEQ ID NO: 4/SEQ ID NO: 1) y la figura 7 es una representación gráfica de los absolutos de Ct (SEQ ID NO: 4).

Ejemplo 5: Análisis de biomarcadores de ADNr 16S en pacientes con cáncer colorrectal frente a sujetos sanos

15 Se obtuvieron muestras fecales de 27 pacientes a los que se les diagnosticó recientemente cáncer colorrectal (CCR) (fase 0-I), e individuos sanos (C). Todos los pacientes tuvieron una colonoscopia al menos 15 días antes de la recogida de muestras. Todos ellos firmaron el correspondiente consentimiento informado. Los criterios de exclusión incluyeron tratamiento con antibióticos dentro de 1 mes antes del estudio y edad < 18 años. En la tabla 5 (véase a continuación) se muestra que la cuantificación de las cuatro secuencias bacterianas presentó un aumento significativo en Ct en las heces de pacientes con CCR. Este hecho significa que hay una menor carga de las cuatro bacterias en este grupo.

20 La disminución significativa de la carga de estas cuatro bacterias podría ser una potente herramienta para cribar

pacientes con CCR. Si se analiza el rendimiento de los marcadores bacterianos por separado, se aprecia que se obtuvieron los mejores resultados con B3 que identificó pacientes con CCR con un 94% de especificidad y un 48% de sensibilidad y 0,7 de exactitud; y en segundo lugar, B46 con un 84% de especificidad y un 61,5% de sensibilidad y 0,698 de exactitud.

5 B48 y B10 también muestran una exactitud de 0,69 pero un 100% de especificidad y un 36% de sensibilidad, y un 57% de especificidad y un 89,5% de sensibilidad, respectivamente.

10 Se calculó la razón entre Ct de la cuantificación de secuencias bacterianas. Aunque no se observan diferencias estadísticas, ha de considerarse que aplicar esta razón en una muestra aumentada podría ser un algoritmo útil para completar el cribado de CCR. En esta línea, la combinación de cuatro secuencias también debe considerarse que aumenta la sensibilidad y especificidad de la prueba.

15 Tabla 5. Cuantificación de las 4 secuencias bacterianas en individuos sanos y pacientes con CCR (expresados en Ct), análisis estadístico a través de dos pruebas t de muestra y análisis ROC.

ID	Grupo	Estado	B3	B48	B10	B46
4	CCR	1	29,86	20,84	14,18	21,34
5	CCR	1	24,3	15,88	10,34	17,53
6	CCR	1	27,33	20,66	15,77	23,13
7	CCR	1	24,88	19,06	13,64	20,69
8	CCR	1	20,61	17,78	12,92	19,7
11	CCR	1	27,28	25,69	17,05	24,35
12	CCR	1	29,04	22,58	16,49	23,17
13	CCR	1	21,98	17,52	10,97	17,34
14	CCR	1	27,07	19,94	13,52	19,9
15	CCR	1	28,57	26,60	19,77	27,53
24	CCR	1	26,77	22,64	17,84	23,83
25	CCR	1	21,64	17,01	11,54	18,7
27	CCR	1	26,52	20,97	15,49	22,63
31	CCR	1	29,91	18,92	16,44	23,8
35	CCR	1	23,37	18,81	12,24	18,77
36	CCR	1	22,87	17,28	12,74	19,76
38	CCR	1		16,36	24,43	30,7
F10	CCR	1	20,49	18,59	16,33	22,78
F16	CCR	1	18,24		14,35	20,75
F11	CCR	1	29,19	20,33	20,37	26,45
F3	CCR	1	32,48	23,52	19,80	26,20
F4	CCR	1	21,76	16,41	14,88	20,87
F5	CCR	1	20,56	19,04	17,61	24,00
F6	CCR	1	31,6	23,22	21,82	27,67
F7	CCR	1	27,14	23,19	22,34	28,20
F8	CCR	1	22,37	18,82	15,61	22,04
43	CCR	1	30,9	22,46	17,12	25,57
16	C	0	22,97	20,21	15,54	23,12
17	C	0	20,4	15,84	10,71	17,54
18	C	0	22,05	15,72	12,06	17,67
23	C	0	21,79	17,16	12,59	18,89

28	C	0	20	18,86	15,29	21,33
30	C	0	26,61	20,47	12,78	20,26
32	C	0	22,76	15,07	15,95	23,09
33	C	0	23,66	18,30	13,54	19,57
34	C	0	22,01	17,73	14,29	19,07
37	C	0	24,06	19,60	14,47	19,65
F1	C	0	24,71	18,09	14,79	21,97
F12	C	0	20,79	17,55	14,95	20,90
F13	C	0	25,1	19,36	19,22	24,78
F14	C	0	21,63	19,46	15,07	21,44
F2	C	0	20,42	17,07	12,76	20,57
F15	C	0	19,69	17,55	12,84	19,85
F9	C	0	23,29	17,61	15,41	21,80
39	C	0		18,03	11,95	18,97
Media	CCR		25,64346154	20,15846154	16,13222222	22,86666667
	C		22,46705882	17,98222222	14,12111111	20,58166667
Desviación estándar	CCR		3,96679613	2,887129637	3,575465632	3,448853767
	C		1,948072011	1,514799756	1,957886432	1,927739456
Valor de p (prueba de la t de dos muestras)	CCR frente a C		0,038	0,019	0,049	0,021
CCR frente a C	Punto medio		26,69	20,90	15,75	22,0
	AUC		0,70	0,69	0,69	0,698
	Sensibilidad		48%	36%	57%	61,5%
	Especificidad		94%	100%	89,5%	84%

Las figuras 14 a 17 proporcionan la representación gráfica de valores de Ct absolutos para B3, B10, B46 y B48, respectivamente, en grupos de CCR y sanos (C).

- 5 La figura 27 proporciona las curvas ROC para B3, B10, B46 y B48 en análisis de sanos frente a CCR.

Ejemplo 6: Análisis de biomarcadores de ADNr 16S de CCR frente Lynch (sin signos clínicos) frente a individuos sanos

- 10 El objetivo de este análisis es determinar el valor predictivo de los biomarcadores para detectar el riesgo de cáncer colorrectal antes de los signos clínicos. Se obtuvieron muestras fecales de 27 pacientes a los que se les diagnosticó recientemente cáncer colorrectal (CCR) (fase 0-I), 24 pacientes portadores del síndrome de Lynch (L) y 19 individuos sanos (C). Todos los pacientes tuvieron una colonoscopia al menos 15 días antes de la recogida de muestras. Todos ellos firmaron el correspondiente consentimiento informado. Los criterios de exclusión incluyeron
- 15 tratamiento con antibióticos dentro de 1 mes antes del estudio y edad < 18 años.

- Se cuantificaron las cuatro secuencias bacterianas en muestras fecales de individuos sanos, pacientes a los que se les diagnosticó recientemente CCR y portadores del síndrome de Lynch. Los portadores del síndrome de Lynch están predispuestos genéticamente a desarrollar cáncer, específicamente cáncer colorrectal. Este grupo está
- 20 compuesto por individuos sin neoplasia de cáncer colorrectal y sin signos clínicos, y se considera como grupo de riesgo de cáncer colorrectal.

Se observó que el grupo con síndrome de Lynch presentó un aumento significativo del valor de Ct de B3 y B48 en

comparación con grupos de pacientes sanos. Por tanto, en el grupo de Lynch hubo una carga menor de bacterias B3 y B48 en muestra de heces.

- 5 Efectivamente, el grupo de Lynch no muestra ninguna diferencia estadística de cuantificación de secuencias bacterianas en comparación con CCR excepto para B10. Lo que implica que el perfil microbiológico del grupo de Lynch es similar a CCR aunque no se observan signos clínicos y neoplasia en este grupo. Este hecho significa que la detección de uno o más de estas cuatro secuencias bacterianas puede ser una potente herramienta para cribar personas con riesgo de CCR y que pueden necesitar exploración de colonoscopia.
- 10 Tabla 6. Cuantificación de las 4 secuencias bacterianas en individuos sanos, pacientes con CCR y portadores del síndrome de Lynch (expresado en Ct), análisis estadístico mediante análisis ROC y de ANOVA de un factor.

ID	Grupo	Estado	B3	B48	B10	B46
4	CCR	1	29,86	20,84	14,18	21,34
5	CCR	1	24,3	15,88	10,34	17,53
6	CCR	1	27,33	20,66	15,77	23,13
7	CCR	1	24,88	19,06	13,64	20,69
8	CCR	1	20,61	17,78	12,92	19,7
11	CCR	1	27,28	25,69	17,05	24,35
12	CCR	1	29,04	22,58	16,49	23,17
13	CCR	1	21,98	17,52	10,97	17,34
14	CCR	1	27,07	19,94	13,52	19,9
15	CCR	1	28,57	26,60	19,77	27,53
24	CCR	1	26,77	22,64	17,84	23,83
25	CCR	1	21,64	17,01	11,54	18,7
27	CCR	1	26,52	20,97	15,49	22,63
31	CCR	1	29,91	18,92	16,44	23,8
35	CCR	1	23,37	18,81	12,24	18,77
36	CCR	1	22,87	17,28	12,74	19,76
38	CCR	1		16,36	24,43	30,7
F10	CCR	1	20,49	18,59	16,33	22,78
F16	CCR	1	18,24		14,35	20,75
F11	CCR	1	29,19	20,33	20,37	26,45
F3	CCR	1	32,48	23,52	19,80	26,20
F4	CCR	1	21,76	16,41	14,88	20,87
F5	CCR	1	20,56	19,04	17,61	24,00
F6	CCR	1	31,6	23,22	21,82	27,67
F7	CCR	1	27,14	23,19	22,34	28,20
F8	CCR	1	22,37	18,82	15,61	22,04
43	CCR	1	30,9	22,46	17,12	25,57
16	C	0	22,97	20,21	15,54	23,12
17	C	0	20,4	15,84	10,71	17,54
18	C	0	22,05	15,72	12,06	17,67
23	C	0	21,79	17,16	12,59	18,89
28	C	0	20	18,86	15,29	21,33
30	C	0	26,61	20,47	12,78	20,26
32	C	0	22,76	15,07	15,95	23,09

ES 2 734 143 T3

33	C	0	23,66	18,30	13,54	19,57
34	C	0	22,01	17,73	14,29	19,07
37	C	0	24,06	19,60	14,47	19,65
F1	C	0	24,71	18,09	14,79	21,97
F12	C	0	20,79	17,55	14,95	20,90
F13	C	0	25,1	19,36	19,22	24,78
F14	C	0	21,63	19,46	15,07	21,44
F2	C	0	20,42	17,07	12,76	20,57
F15	C	0	19,69	17,55	12,84	19,85
F9	C	0	23,29	17,61	15,41	21,80
39	C	0		18,03	11,95	18,97
1	L	2	23,08	18,76	12,03	19,21
2	L	2	26,58	18,47	13,60	21,02
3	L	2	23,51	19,36	10,85	18,33
9	L	2		16,94	12,89	19,77
10	L	2	19,91	17,68	10,44	17,84
19	L	2	32,81	21,60	15,72	21,67
20	L	2	19,9	19,77	17,81	24,2
21	L	2	19,66	17,05	13,55	19,7
22	L	2	21,28	16,04	10,22	17,48
26	L	2	19,57	20,73	11,79	19,36
29	L	2	23,67	17,92	11,30	18,2
40	L	2	29,14	23,47	19,42	26,3
41	L	2	32,14	20,18	17,27	23,72
42	L	2	35,19	18,34	19,63	27,07
44	L	2	28,83	23,16	16,15	23,42
45	L	2	24,79	16,85	10,84	19,02
MIL1	L	2	37,1	22,60	14,69	25,52
MIL2	L	2	32,74	24,60	14,63	25,53
MIL3	L	2	27,80	18,00	10,34	20,28
MIL4	L	2	29,09	20,00	18,17	29,66
MIL5	L	2	33,45	22,60		24,81
MIL6	L	2	31,69	20,20	14,66	24,54
MIL7	L	2	35,88	22,90	13,05	24,59
MIL8	L	2	26,2	20,60	10,85	21,87
Media	CCR		25,64346154	20,15846154	16,13222222	22,86666667
	C		22,46705882	17,98222222	14,12111111	20,58166667
	Lynch		27,56565217	19,90916667	13,9076087	22,21291667
Desviación estándar	CCR		3,96679613	2,887129637	3,575465632	3,448853767
	C		1,948072011	1,514799756	1,957886432	1,927739456
	Lynch		5,628873475	2,403976627	3,02383331	3,369239985
Valor de p (ANOVA de un factor)	CCR frente a C		0,038	0,019	0,049	0,021

	CCR frente a L	0,381	0,988	0,057	0,849
	L frente a C	0,004	0,018	0,989	0,163
C frente a L	Punto medio	25,65	19,69	11,87	23,27
	AUC	0,73	0,746	0,55	0,629
	Sensibilidad	60,9	54,2	34,8	45,8
L frente a CCR	Especificidad	88,9	94,1	94,7	94,7
	Punto medio	31,65	19,21	14,78	21,95
	AUC	0,60	0,487	0,69	0,55
	Sensibilidad	96	52	65,4	61,5
	Especificidad	34,8	58,3	69,6	54,2
	Punto medio	26,69	20,90	15,75	22,0
CCR frente a C	AUC	0,70	0,69	0,69	0,698
	Sensibilidad	48%	36%	57%	61,5%
	Especificidad	94%	100%	89,5%	84%

Las figuras 18 a 21 proporcionan la representación gráfica de valores de Ct absolutos en grupos de CCR, L y C de secuencias SEQ ID NO: 1, SEQ ID NO: 4, SEQ ID NO: 7 y SEQ ID NO: 10, respectivamente.

5 La figura 28 proporciona las curvas ROC para B3, B10, B46 y B48 en análisis de sanos (C) frente a Lynch (L).

La figura 30 proporciona las curvas ROC para B3, B10, B46 y B48 en análisis de CCR frente a Lynch.

Ejemplo 7: Análisis de biomarcadores de ADNr 16S de CCR + Lynch frente a individuos sanos

10 Según el análisis del ejemplo 6, los portadores del síndrome de Lynch tienen un perfil microbiológico en heces similar a pacientes con CCR. Por tanto, en este experimento, se pretende someter a prueba el potencial de usar detección de cuantificación de las cuatro secuencias bacterianas para cribar individuos con riesgo de cáncer colorrectal en individuos sanos e individuos con riesgo de cáncer colorrectal (CCR + L), antes de la exploración de colonoscopia.

15 En la tabla 7, se han representado la cuantificación de secuencias bacterianas expresadas en Ct.

20 Se observa que B3, B48 y B46 aumentan significativamente el valor de Ct en el grupo de riesgo de cáncer colorrectal (CCR + L). Por tanto, el nivel de estas bacterias en heces disminuye.

B10 en CCR + L también presenta un nivel más bajo aunque no se observan diferencias estadísticas ($p=0,137$).

25 Al evaluar el rendimiento de cuantificación bacteriana en la detección de cáncer colorrectal, se muestra que B3 y B48 tienen la mejor exactitud (0,7) y, respectivamente, el 56,2% de sensibilidad-el 88,9% de especificidad y el 51% de sensibilidad-el 84% de especificidad.

30 En segundo lugar, B46 tienen una exactitud de 0,665 y el 94,7% de sensibilidad-el 23,12% de especificidad. Finalmente, B10 tiene una exactitud de 0,58 y el 44,9% de sensibilidad-el 89,5% de especificidad.

35 Se calculó la razón entre Ct de cuantificación de secuencias bacterianas. Aunque no se observan diferencias estadísticas (B3/B46 $p = 0,061$ con prueba de la U de Mann-Whitney; B10/B3 $p = 0,132$ con análisis de prueba de la t de dos muestras; B46/B10 con análisis de prueba de la t de dos muestras $p = 0,127$); ha de considerarse que aplicar esta razón en una muestra aumentada puede ser un algoritmo útil para completar el cribado de CCR.

En esta línea, la combinación de cuatro secuencias también debe considerarse que aumenta la sensibilidad y especificidad de la prueba.

40 Tabla 7. Cuantificación de las 4 secuencias bacterianas expresadas en Ct de individuos sanos (C) e individuos con riesgo de cáncer colorrectal (CCR + L), análisis estadístico mediante análisis de ROC y prueba de la t de dos muestras.

ID	Grupo	Estado	B3	B48	B10	B46
----	-------	--------	----	-----	-----	-----

ES 2 734 143 T3

4	CCR + L	3	29,86	20,84	14,18	21,34
5	CCR + L	3	24,3	15,88	10,34	17,53
6	CCR + L	3	27,33	20,66	15,77	23,13
7	CCR + L	3	24,88	19,06	13,64	20,69
8	CCR + L	3	20,61	17,78	12,92	19,7
11	CCR + L	3	27,28	25,69	17,05	24,35
12	CCR + L	3	29,04	22,58	16,49	23,17
13	CCR + L	3	21,98	17,52	10,97	17,34
14	CCR + L	3	27,07	19,94	13,52	19,9
15	CCR + L	3	28,57	26,60	19,77	27,53
24	CCR + L	3	26,77	22,64	17,84	23,83
25	CCR + L	3	21,64	17,01	11,54	18,7
27	CCR + L	3	26,52	20,97	15,49	22,63
31	CCR + L	3	29,91	18,92	16,44	23,8
35	CCR + L	3	23,37	18,81	12,24	18,77
36	CCR + L	3	22,87	17,28	12,74	19,76
38	CCR + L	3		16,36	24,43	30,7
F10	CCR + L	3	20,49	18,59	16,33	22,78
F16	CCR + L	3	18,24		14,35	20,75
F11	CCR + L	3	29,19	20,33	20,37	26,45
F3	CCR + L	3	32,48	23,52	19,80	26,20
F4	CCR + L	3	21,76	16,41	14,88	20,87
F5	CCR + L	3	20,56	19,04	17,61	24,00
F6	CCR + L	3	31,6	23,22	21,82	27,67
F7	CCR + L	3	27,14	23,19	22,34	28,20
F8	CCR + L	3	22,37	18,82	15,61	22,04
43	CCR + L	3	30,9	22,46	17,12	25,57
1	CCR + L	3	23,08	18,76	12,03	19,21
2	CCR + L	3	26,58	18,47	13,60	21,02
3	CCR + L	3	23,51	19,36	10,85	18,33
9	CCR + L	3		16,94	12,89	19,77
10	CCR + L	3	19,91	17,68	10,44	17,84
19	CCR + L	3	32,81	21,60	15,72	21,67
20	CCR + L	3	19,9	19,77	17,81	24,2
21	CCR + L	3	19,66	17,05	13,55	19,7
22	CCR + L	3	21,28	16,04	10,22	17,48
26	CCR + L	3	19,57	20,73	11,79	19,36
29	CCR + L	3	23,67	17,92	11,30	18,2
40	CCR + L	3	29,14	23,47	19,42	26,3
41	CCR + L	3	32,14	20,18	17,27	23,72
42	CCR + L	3	35,19	18,34	19,63	27,07
44	CCR + L	3	28,83	23,16	16,15	23,42
45	CCR + L	3	24,79	16,85	10,84	19,02

ES 2 734 143 T3

MIL1	CCR + L	3	37,1	22,60	14,69	25,52
MIL2	CCR + L	3	32,74	24,60	14,63	25,53
MIL3	CCR + L	3	27,80	18,00	10,34	20,28
MIL4	CCR + L	3	29,09	20,00	18,17	29,66
MIL5	CCR + L	3	33,45	22,60		24,81
MIL6	CCR + L	3	31,69	20,20	14,66	24,54
MIL7	CCR + L	3	35,88	22,90	13,05	24,59
MIL8	CCR + L	3	26,2	20,60	10,85	21,87
16	C	0	22,97	20,21	15,54	23,12
17	C	0	20,4	15,84	10,71	17,54
18	C	0	22,05	15,72	12,06	17,67
23	C	0	21,79	17,16	12,59	18,89
28	C	0	20	18,86	15,29	21,33
30	C	0	26,61	20,47	12,78	20,26
32	C	0	22,76	15,07	15,95	23,09
33	C	0	23,66	18,30	13,54	19,57
34	C	0	22,01	17,73	14,29	19,07
37	C	0	24,06	19,60	14,47	19,65
F1	C	0	24,71	18,09	14,79	21,97
F12	C	0	20,79	17,55	14,95	20,90
F13	C	0	25,1	19,36	19,22	24,78
F14	C	0	21,63	19,46	15,07	21,44
F2	C	0	20,42	17,07	12,76	20,57
F15	C	0	19,69	17,55	12,84	19,85
F9	C	0	23,29	17,61	15,41	21,80
39	C	0		18,03	11,95	18,97
Media	CCR + L		26,54	20,03	15,10	22,56
	C		22,46	17,98	14,12	20,58
Desviación estándar	CCR + L		4,86	2,64	3,48	3,39
	C		1,94	1,51	1,96	1,93
Valor de p (prueba la t de dos muestras)	CCR + L frente a C		0,0001	0,001	0,137	0,004
	Punto medio		25,65	19,69	15,58	23,13
	AUC		0,71	0,70	0,580	0,665
	Sensibilidad		56,2	51	44,9	94,7
	Especificidad		88,9	84	89,5	23,12

La figura 22 proporciona la representación gráfica de valores de Ct absolutos de secuencias SEQ ID NO: 1, SEQ ID NO: 4, SEQ ID NO: 7 y SEQ ID NO: 10, respectivamente, en grupos de CCR+ Lynch (L) y sanos (C).

- 5 La figura 29 proporciona las curvas ROC para B3, B10, B46 y B48 en análisis de sanos frente a CCR + con síndrome de Lynch.

Ejemplo 8: Análisis de biomarcadores de ADNr 16S de CCR frente a L-alto riesgo frente a L-bajo riesgo frente a individuos sanos

En este ejemplo, se pretende definir diferentes grados de riesgo de cáncer colorrectal usando cuantificación de secuencias bacterianas en portadores del síndrome de Lynch.

5 Se clasificaron los portadores del síndrome de Lynch en cáncer colorrectal de alto riesgo (L-alto riesgo) y bajo riesgo (L-bajo riesgo) según sus antecedentes de neoplasia colorrectal y la presencia de adenomas en su última colonoscopia. Se tienen 6 individuos de alto riesgo y 18 de bajo.

10 Se realizó la cuantificación de las cuatro secuencias bacterianas y no se observaron diferencias estadísticas entre los diferentes grupos. Sólo para los niveles de B3, hubo una tendencia de un Ct disminuido en comparación con C frente a L-bajo riesgo ($p = 0,088$, prueba de Kruskal-Wallis) y C frente a L-alto riesgo ($p = 0,143$, prueba de Kruskal-Wallis).

Se calculó la razón entre Ct de cuantificación de secuencias bacterianas.

15 Se aplicó la prueba de Kruskal-Wallis y las razones B3/B10 y B10/B46 presentan una diferencia significativa entre C y L-bajo riesgo. La razón B3/B48 muestra una diferencia significativa al comparar C frente a L-alto riesgo. Las razones B3/B10 y B48/B10 fueron significativamente diferentes en CCR frente a L-alto riesgo.

20 No se realizaron los análisis de ROC debido al tamaño de la muestra, pero considerando los análisis estadísticos, se podría predecir que la mejor razón podría ser B3/B10 porque es capaz de distinguir grupo de L-bajo riesgo.

En esta línea, la combinación de cuatro secuencias también debe considerarse que aumenta la potencia de la prueba, y debe realizarse un análisis de regresión logística multivariable pero se requiere un tamaño de muestra aumentado.

25 Tabla 8. Cuantificación de las 4 secuencias bacterianas expresadas en Ct de individuos sanos (C), L-bajo riesgo, L-alto riesgo y CRC

ID	Grupo	Estado	B3	B48	B10	B46	B3/B10	B48/B10	B3/B48	B10/B46
4	CCR	1	29,86	20,84	14,18	21,34	2,11	1,47	1,43	0,66
5	CRC	1	24,3	15,88	10,34	17,53	2,35	1,54	1,53	0,59
6	CCR	1	27,33	20,66	15,77	23,13	1,73	1,31	1,32	0,68
7	CCR	1	24,88	19,06	13,64	20,69	1,82	1,40	1,31	0,66
8	CCR	1	20,61	17,78	12,92	19,7	1,60	1,38	1,16	0,66
11	CCR	1	27,28	25,69	17,05	24,35	1,60	1,51	1,06	0,70
12	CCR	1	29,04	22,58	16,49	23,17	1,76	1,37	1,29	0,71
13	CCR	1	21,98	17,52	10,97	17,34	2,00	1,60	1,25	0,63
14	CCR	1	27,07	19,94	13,52	19,9	2,00	1,47	1,36	0,68
15	CCR	1	28,57	26,60	19,77	27,53	1,45	1,35	1,07	0,72
24	CCR	1	26,77	22,64	17,84	23,83	1,50	1,27	1,18	0,75
25	CCR	1	21,64	17,01	11,54	18,7	1,88	1,47	1,27	0,62
27	CCR	1	26,52	20,97	15,49	22,63	1,71	1,35	1,26	0,68
31	CCR	1	29,91	18,92	16,44	23,8	1,82	1,15	1,58	0,69
35	CCR	1	23,37	18,81	12,24	18,77	1,91	1,54	1,24	0,65
36	CCR	1	22,87	17,28	12,74	19,76	1,80	1,36	1,32	0,64
38	CCR	1		16,36	24,43	30,7		0,67		0,80
F10	CCR	1	20,49	18,59	16,33	22,78	1,25	1,14	1,10	0,72
F16	CCR	1	18,24		14,35	20,75	1,27			0,69

ES 2 734 143 T3

F11	CCR	1	29,19	20,33	20,37	26,45	1,43	1,00	1,44	0,77
F3	CCR	1	32,48	23,52	19,80	26,20	1,64	1,19	1,38	0,76
F4	CCR	1	21,76	16,41	14,88	20,87	1,46	1,10	1,33	0,71
F5	CCR	1	20,56	19,04	17,61	24,00	1,17	1,08	1,08	0,73
F6	CCR	1	31,6	23,22	21,82	27,67	1,45	1,06	1,36	0,79
F7	CCR	1	27,14	23,19	22,34	28,20	1,21	1,04	1,17	0,79
F8	CCR	1	22,37	18,82	15,61	22,04	1,43	1,21	1,19	0,71
43	CCR	1	30,9	22,46	17,12	25,57	1,80	1,31	1,38	0,67
16	C	0	22,97	20,21	15,54	23,12	1,48	1,30	1,14	0,67
17	C	0	20,4	15,84	10,71	17,54	1,90	1,48	1,29	0,61
18	C	0	22,05	15,72	12,06	17,67	1,83	1,30	1,40	0,68
23	C	0	21,79	17,16	12,59	18,89	1,73	1,36	1,27	0,67
28	C	0	20	18,86	15,29	21,33	1,31	1,23	1,06	0,72
30	C	0	26,61	20,47	12,78	20,26	2,08	1,60	1,30	0,63
32	C	0	22,76	15,07	15,95	23,09	1,43	0,94	1,51	0,69
33	C	0	23,66	18,30	13,54	19,57	1,75	1,35	1,29	0,69
34	C	0	22,01	17,73	14,29	19,07	1,54	1,24	1,24	0,75
37	C	0	24,06	19,60	14,47	19,65	1,66	1,35	1,23	0,74
F1	C	0	24,71	18,09	14,79	21,97	1,67	1,22	1,37	0,67
F12	C	0	20,79	17,55	14,95	20,90	1,39	1,17	1,18	0,72
F13	C	0	25,1	19,36	19,22	24,78	1,31	1,01	1,30	0,78
F14	C	0	21,63	19,46	15,07	21,44	1,44	1,29	1,11	0,70
F2	C	0	20,42	17,07	12,76	20,57	1,60	1,34	1,20	0,62
F15	C	0	19,69	17,55	12,84	19,85	1,53	1,37	1,12	0,65
F9	C	0	23,29	17,61	15,41	21,80	1,51	1,14	1,32	0,71
39	C	0		18,03	11,95	18,97		1,51		0,63
1	L-bajo riesgo	4	23,08	18,76	12,03	19,21	1,92	1,56	1,23	0,63
3	L-bajo riesgo	4	23,51	19,36	10,85	18,33	2,17	1,78	1,21	0,59
9	L-bajo riesgo	4		16,94	12,89	19,77		1,31		0,65

ES 2 734 143 T3

10	L-bajo riesgo	4	19,91	17,68	10,44	17,84	1,91	1,69	1,13	0,58
19	L-bajo riesgo	4	32,81	21,60	15,72	21,67	2,09	1,37	1,52	0,73
20	L-bajo riesgo	4	19,9	19,77	17,81	24,2	1,12	1,11	1,01	0,74
21	L-bajo riesgo	4	19,66	17,05	13,55	19,7	1,45	1,26	1,15	0,69
26	L-bajo riesgo	4	19,57	20,73	11,79	19,36	1,66	1,76	0,94	0,61
29	L-bajo riesgo	4	23,67	17,92	11,30	18,2	2,09	1,59	1,32	0,62
40	L-bajo riesgo	4	29,14	23,47	19,42	26,3	1,50	1,21	1,24	0,74
42	L-bajo riesgo	4	35,19	18,34	19,63	27,07	1,79	0,93	1,92	0,73
44	L-bajo riesgo	4	28,83	23,16	16,15	23,42	1,79	1,43	1,24	0,69
MIL1	L-bajo riesgo	4	37,1	22,60	14,69	25,52	2,53	1,54	1,64	0,58
MIL2	L-bajo riesgo	4	32,74	24,60	14,63	25,53	2,24	1,68	1,33	0,57
MIL3	L-bajo riesgo	4	27,80	18,00	10,34	20,28	2,69	1,74	1,54	0,51
MIL6	L-bajo riesgo	4	31,69	20,20	14,66	24,54	2,16	1,38	1,57	0,60
MIL7	L-bajo riesgo	4	35,88	22,90	13,05	24,59	2,75	1,75	1,57	0,53
MIL8	L-bajo riesgo	4	26,2	20,60	10,85	21,87	2,41	1,90	1,27	0,50
2	L-alto riesgo	5	26,58	18,47	13,60	21,02	1,95	1,36	1,44	0,65
22	L-alto riesgo	5	21,28	16,04	10,22	17,48	2,08	1,57	1,33	0,58
41	L-alto riesgo	5	32,14	20,18	17,27	23,72	1,86	1,17	1,59	0,73
45	L-alto riesgo	5	24,79	16,85	10,84	19,02	2,29	1,55	1,47	0,57
MIL4	L-alto riesgo	5	29,09	20,00	18,17	29,66	1,60	1,10	1,45	0,61
MIL5	L-alto riesgo	5	33,45	22,60		24,81			1,48	
Media		CCR	25,64	20,16	16,13	22,87	1,66	1,28	1,28	0,70
		C	22,47	17,98	14,12	20,58	1,60	1,29	1,25	0,68
		L-bajo riesgo	27,45	20,20	13,88	22,08	2,02	1,50	1,34	0,63
		L-alto riesgo	27,89	19,02	14,02	22,62	1,96	1,35	1,46	0,63
Desviación estándar		CCR	3,97	2,89	3,58	3,45	0,29	0,21	0,14	0,05
		C	1,95	1,51	1,96	1,93	0,21	0,16	0,11	0,05
		L-bajo riesgo	6,08	2,40	2,96	3,09	0,44	0,27	0,25	0,08

	L-alto riesgo	4,59	2,41	3,62	4,42	0,26	0,22	0,09	0,06
Valor de p Prueba de Kruskal-Wallis	CCR frente a C	0.32							
	CCR frente a LL	1				0.017	0.022		
	CCR frente a LH	1							
	C frente a LL	0.088				0.016			0.01
	C frente a LH	0.143						0.031	
	LL frente a LH	1							

Las figuras 23 a 26 proporcionan la representación gráfica de las siguientes razones de valores de Ct absolutos B48/B10, B3/B10, B46/B10 y B3/B48, respectivamente en grupos de CCR, alto riesgo L, bajo riesgo L y C.

5

Lista de secuencias

<110> Institut d'Investigació Biomèdica de Girona Dr. Josep Trueta (IDIBGI)
Universitat de Girona (UdG)

10

<120> BIOMARCADORES PARA DETECCIÓN TEMPRANA, CRIBADO DEL RIESGO Y MONITORIZACIÓN DE CÁNCER COLORRECTAL Y PÓLIPOS ADENOMATOSOS

<130> 900 663

15

<150> Documento EP14382074
<151> 03-03-2014

<160> 13

20

<170> PatentIn versión 3.5

<210> 1

<211> 489

25

<212> ADN
<213> Secuencia artificial

<220>

<223> /mol_type="ADN no asignado"

30

/note="secuencia de ADN ribosómico 16S de B3, secuencia parcial (número de registro GQ411111.1)"

<400> 1

```

acggaggcct tcgggtcgta accgctttca gcaggaaga gtcaagactg tacctgcaga      60
agaagccccg gctaactacg tgccagcagc cgcggtaata cgtagggggc gagcgttatac    120
cggattcatt gggcgtaaag cgcgcgtagg cggcccggca ggccgggggt cgaagcgggg      180
ggctcaacc  cccgaagccc ccggaacctc cgcggcttgg gtccggtagg ggagggtgga      240
acaccgggtg tagcgggtgga atgcgcagat atcgggtgga acaccgggtg cgaagcgggc      300
cctctggggc gagaccgacg ctgaggcgcg aaagctgggg gagcgaacag gattagatac      360
cctggtagtc ccagccgtaa acgatggacg ctaggtgtgg ggggacgatc cccccgtgcc      420
gcagccaacg cattaagcgt cccgcctggg gagtacggcc gcaaggctaa aactcaaagg      480
aattgacgg                                     489
    
```

ES 2 734 143 T3

<210> 2
 <211> 19
 <212> ADN
 5 <213> Secuencia artificial

 <220>
 <223> /mol_type="ADN no asignado"
 /note="cebador directo de secuencia de ADN ribosómico 16S de B3"
 10
 <400> 2

 ggaggccttc gggtcgtaa 19

 15 <210> 3
 <211> 18
 <212> ADN
 <213> Secuencia artificial

 20 <220>
 <223> /mol_type="ADN no asignado"
 /note="cebador inverso de secuencia de ADN ribosómico 16S de B3"

 <400> 3
 25
 ccgaagcccc cggaacct 18

 <210> 4
 <211> 485
 30 <212> ADN
 <213> Secuencia artificial

 <220>
 <223> /mol_type="ADN no asignado"
 35 /note="secuencia de ADN ribosómico 16S de B10, secuencia parcial (número de registro GQ411118.1)"
 <400> 4

 cttcggattg taaactcctg ttgttgagga gataatgacg gtactcaaca aggtaagtga 60
 cggctaacta cgtgccagca gccgcggtaa aacgtaggtc acaagcgttg tccggaatta 120
 ctgggtgtaa agggagcgca ggcgggaaga caagttgga gtgaaatcca tgggctcaac 180
 ccatgaactg ctttcaaaac tgtttttctt gagtagtgca caggtaggcg gaattcccgg 240
 tgtagcggtg gaatgcgtag atatcgggag gaacaccagt ggcgaaggcg gcctactggg 300
 caccaactga cgctgaggct cgaaagtgtg ggtagcaaac aggattagat accctggtag 360
 tccacactgt aaacgatgat tactaggtgt tggaggattg accccttcag tgccgcagtt 420
 aacacaataa gtaatccacc tggggagtac gaccgcaagg ttgaaactca aaggaattga 480

 cggac 485
 40
 <210> 5
 <211> 19
 <212> ADN
 <213> Secuencia artificial
 45
 <220>
 <223> /mol_type="ADN no asignado"
 /note="cebador directo de secuencia de ADN ribosómico 16S de B10"

ES 2 734 143 T3

<400> 5
caacaaggta agtgacggc 19

5 <210> 6
<211> 20
<212> ADN
<213> Secuencia artificial

10 <220>
<223> /mol_type="ADN no asignado"
/note="cebador inverso de secuencia de ADNr 16S de B10"

15 <400> 6
cgctactg tgactactc 20

20 <210> 7
<211> 474
<212> ADN
<213> Secuencia artificial

25 <220>
<223> /mol_type="ADN no asignado"
/note="secuencia de ADN ribosómico 16S de B46, secuencia parcial (número de registro GQ411150.1)"

<400> 7

gtaacatttc tgtaacaaga catacagcag tttctcagag ttcccaaact caaaggaatt 60

gacggaagcc gcggttccac gtaagtcaca agcgttgctc ggaattactg ggtgtaaagg 120

gagcgcaggc ggaagacaa gttggaagtg aatccatgg gctcaacca taaactgctt 180

tcctaactgt ttttcttgag tagtgcacag gtaggcgaa ttcccgggtg agcgggtggaa 240

tgcttagata tcgggaggaa caccattggc gaaggcggcc tactggactg caactgacgc 300

tgaggctcga aagtgtgggt agcaaacagg attagatacc ctggtagtcc actccgtaaa 360

ccatgattac tacgtgttg aggattgacc ctttctgtgc cgcagataac acaataagta 420

atccacctgg ggagtacgac cgcccgggtg agactcacag gaattgacgg actc 474

30 <210> 8
<211> 19
<212> ADN
<213> Secuencia artificial

35 <220>
<223> /mol_type="ADN no asignado"
/note="cebador directo de secuencia de ADN ribosómico 16S de B46"

40 <400> 8
tccacgtaag tcacaagcg 19

45 <210> 9
<211> 20
<212> ADN
<213> Secuencia artificial

50 <220>
<223> /mol_type="ADN no asignado"
/note="cebador inverso de secuencia de ADN ribosómico 16S de B46"

ES 2 734 143 T3

<400> 9

5 cgcttacctg tgactactc 20

<210> 10
<211> 566
<212> ADN
<213> Secuencia artificial

10 <220>
<223> /mol_type="ADN no asignado"
/note="secuencia de ADN ribosómico 16S de B48, secuencia parcial (número de registro GQ411152.1)"

15 <400> 10

tctcttacgg ggagcagcag tggggaatat tgcacaatgg gggaaaccct gatgcagcga 60

cgccgcgtga gcgatgaagt atttcggtat gtaaagctct atcagcaggg aagaaaatga 120

cggtagctga ctaagaagcc ccggctaaat acgtgccagc agccgcggta atacgtatgg 180

tgcaagcgtt atccggattt actgggtgta aaggagcgt agacggagtg gcaagtctga 240

tgtgaaaacc cggggctcaa ccccgggact gcattggaaa ctgtgcatct agagtgtcgg 300

agaggtaagc ggaattccta gtgtagcggg gaaatgcgta gatattagga ggaacaccag 360

tggcgaaggc ggcttactgg acgataactg acgctgaggg tcgaaagcgt ggggagcaaa 420

caggattaga taccctggta gtccacgccg taaacgatga ctactaggtg tcggggagca 480

cagctcttcg gtgccgcagc aaacgcaata agtattccac ctggggagta cgttcgcaag 540

aatgaaactc acagaatttg acggag 566

<210> 11
20 <211> 18
<212> ADN
<213> Secuencia artificial

<220>
25 <223> /mol_type="ADN no asignado"
/note="cebador directo de secuencia de ADN ribosómico 16S de B48"

<400> 11

30 gtacggggag cagcagtg 18

<210> 12
<211> 23
<212> ADN
35 <213> Secuencia artificial

<220>
<223> /mol_type="ADN no asignado"
/note="cebador inverso de secuencia de ADN ribosómico 16S de B48"

40 <400> 12

gacacttag atgcacagtt tcc 23

45 <210> 13
<211> 18
<212> ADN

ES 2 734 143 T3

<213> Secuencia artificial

<220>

5 <223> /mol_type="ADN no asignado"
/note="cebador inverso de secuencia de ADN ribosómico 16S de B3"

<400> 13

10 aggtccggg ggcttcgg

18

REIVINDICACIONES

1. Método para determinar el riesgo de desarrollar cáncer colorrectal (CCR) en un sujeto humano que comprende:
 - i. cuantificar al menos una secuencia bacteriana de ADNr 16S seleccionada de la lista que consiste en SEQ ID NO: 1, SEQ ID NO: 7 y SEQ ID NO: 10, a partir de una muestra de heces de dicho sujeto;
 - ii. comparar los niveles de cuantificación de la muestra del sujeto en (i) con los niveles de cuantificación en una muestra de control; y
 - iii. determinar el riesgo de desarrollar CCR en dicho sujeto humano, cuando existe una desviación de los valores de muestra del sujeto a partir de los valores en dicha muestra de control.
2. Método de cribado de cáncer colorrectal (CCR) en un sujeto humano que comprende:
 - i. cuantificar al menos una secuencia bacteriana de ADNr 16S seleccionada de la lista que consiste en SEQ ID NO: 1, SEQ ID NO: 4, SEQ ID NO: 7 y SEQ ID NO: 10 a partir de una muestra de heces de dicho sujeto;
 - ii. comparar los niveles de cuantificación de la muestra del sujeto en (i) con los niveles de cuantificación en una muestra de control; y
 - iii. diagnosticar, detectar de manera temprana o determinar recaída de CCR en dicho sujeto humano, o determinar que debe realizarse una colonoscopia en dicho sujeto humano, cuando existe una desviación de los valores de muestra del sujeto a partir de los valores en dicha muestra de control.
3. Método según cualquiera de las reivindicaciones 1 ó 2, en el que dicho método comprende en la etapa (i) cuantificar 2, 3 o todas de dichas secuencias bacterianas, preferiblemente todas de dichas secuencias bacterianas.
4. Método según cualquiera de las reivindicaciones 1 a 3, en el que al menos dos de dichas secuencias bacterianas se cuantifican en la etapa i) y al menos se determina una de las razones de los niveles de cuantificación de dichas secuencias, que comprende además comparar al menos una de dichas razones en dicho sujeto con la razón en una muestra de control, en el que una desviación de la razón de dicha muestra de control es indicadora de CCR.
5. Método según cualquiera de las reivindicaciones 1 a 4, en el que antes de la cuantificación de dichas secuencias bacterianas, se extrae ADN de la muestra de heces lo que implica alteración física de la muestra por cualquier medio y el ADN se purifica sobre columnas de membrana de sílice.
6. Método según cualquiera de las reivindicaciones 1 a 5, en el que la cuantificación de dichas secuencias bacterianas se realiza mediante PCR cuantitativa.
7. Método según la reivindicación 6, en el que los niveles de cuantificación de las secuencias se expresan como valor umbral de ciclo (Ct).
8. Método según cualquiera de las reivindicaciones 1 a 7, en el que dicho al menos un nivel de cuantificación de ADNr 16S se determina mediante PCR cuantitativa a partir de dicha muestra de heces humanas y en el que los niveles de cuantificación de SEQ ID NO: 1 se determinan usando cebadores con al menos una identidad del 90% con SEQ ID NO: 2 y 13, respectivamente; los niveles de cuantificación de SEQ ID NO: 4 se determinan usando cebadores con al menos una identidad del 90% con SEQ ID NO: 5 y 6, respectivamente; los niveles de cuantificación de SEQ ID NO: 7 se determinan usando cebadores con al menos una identidad del 90% con SEQ ID NO: 8 y 9, respectivamente; y los niveles de cuantificación de SEQ ID NO: 10 se determinan usando cebadores con al menos una identidad del 90% con SEQ ID NO: 11 y 12, respectivamente.
9. Método según cualquiera de las reivindicaciones 1 a 8, en el que los niveles de SEQ ID NO: 1 se determinan usando cebadores con secuencia SEQ ID NO: 2 y 13, los niveles de SEQ ID NO: 4 se determinan usando cebadores con secuencia SEQ ID NO: 5 y 6; los niveles de SEQ ID NO: 7 se determinan usando cebadores con secuencia SEQ ID NO: 8 y 9; y los niveles de SEQ ID NO: 10 se determinan usando cebadores con secuencia SEQ ID NO: 11 y 12.
10. Método según cualquiera de las reivindicaciones 1 a 9, en el que dicho método comprende además detectar y/o cuantificar uno o más biomarcadores moleculares que se sabe que son indicadores de CCR.
11. Método según cualquiera de las reivindicaciones 1 a 10, en el que dicho método comprende además

almacenar los resultados del método en un medio legible por ordenador.

5 12. Uso de un kit para diagnosticar, detectar de manera temprana o determinar recaída de CCR, determinar el riesgo de desarrollar CCR o determinar si debe realizarse una colonoscopia, a partir de una muestra de heces de un sujeto humano, según el método de cualquiera de las reivindicaciones 1 a 11, comprendiendo dicho kit:

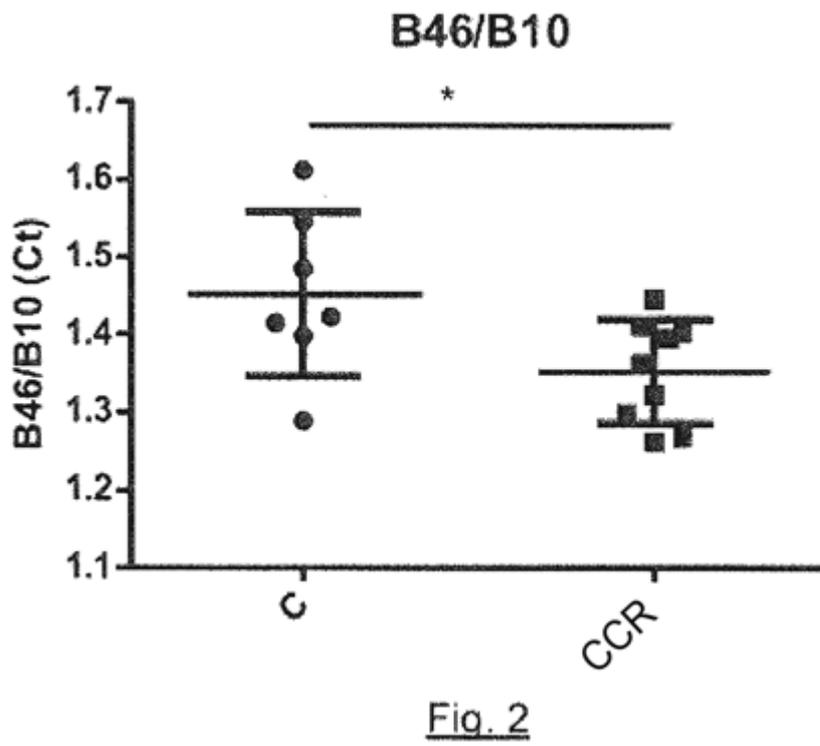
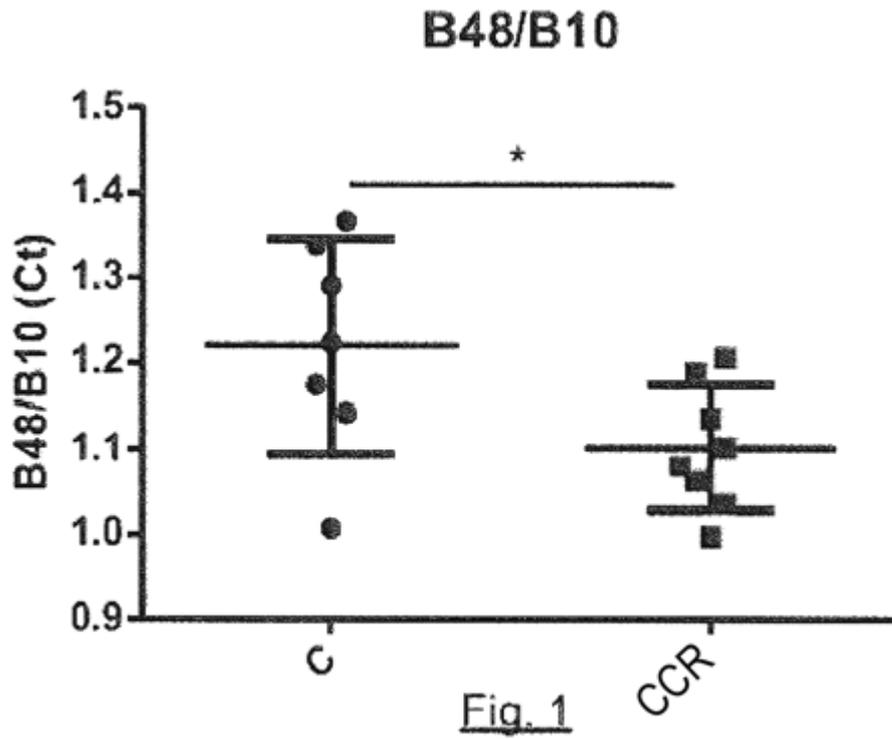
a. un reactivo seleccionado del grupo que consiste en:

10 i. sondas de ácido nucleico capaces de hibridarse específicamente con al menos una secuencia de ADNr 16S seleccionada de la lista que consiste en SEQ ID NO: 1, SEQ ID NO: 4, SEQ ID NO: 7 y SEQ ID NO: 10; o

15 ii. un par de cebadores de ácido nucleico capaces de amplificar específicamente al menos una secuencia de ADNr 16S seleccionada de la lista que consiste en SEQ ID NO: 1, SEQ ID NO: 4, SEQ ID NO: 7 y SEQ ID NO: 10; y

20 b. instrucciones para cuantificar los niveles de una o más de dichas secuencias a partir de una muestra de heces humanas.

25 13. Uso de un kit según la reivindicación 12, en el que dichas sondas de ácido nucleico en (i) o par de cebadores de ácido nucleico en (ii) se seleccionan del grupo que consiste en SEQ ID NO: 2, SEQ ID NO: 13, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 11, y SEQ ID NO: 12; o con una identidad de al menos el 90% de los mismos.



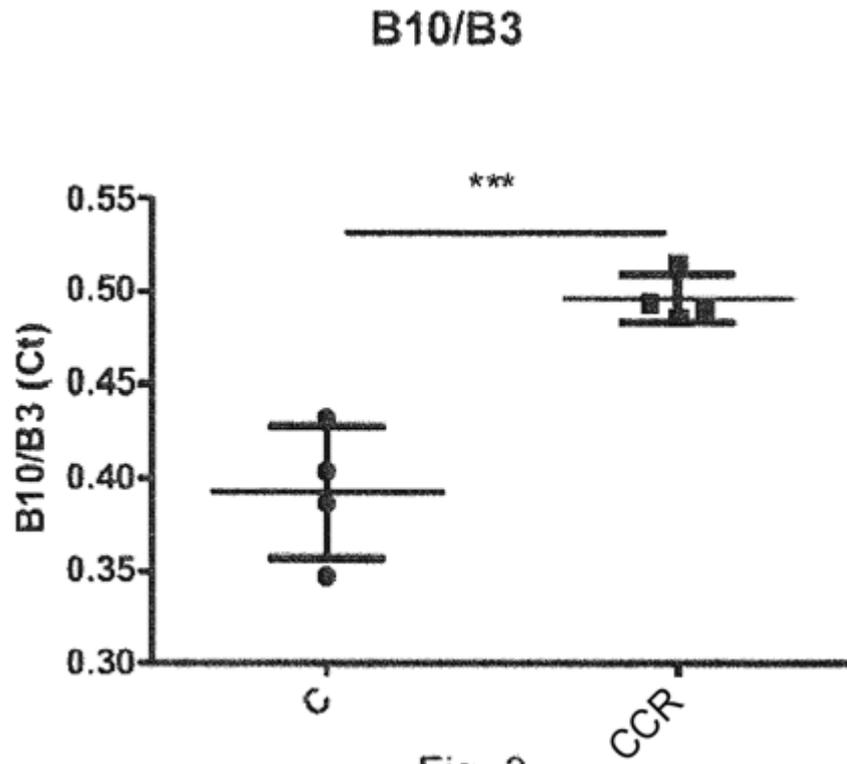


Fig. 3

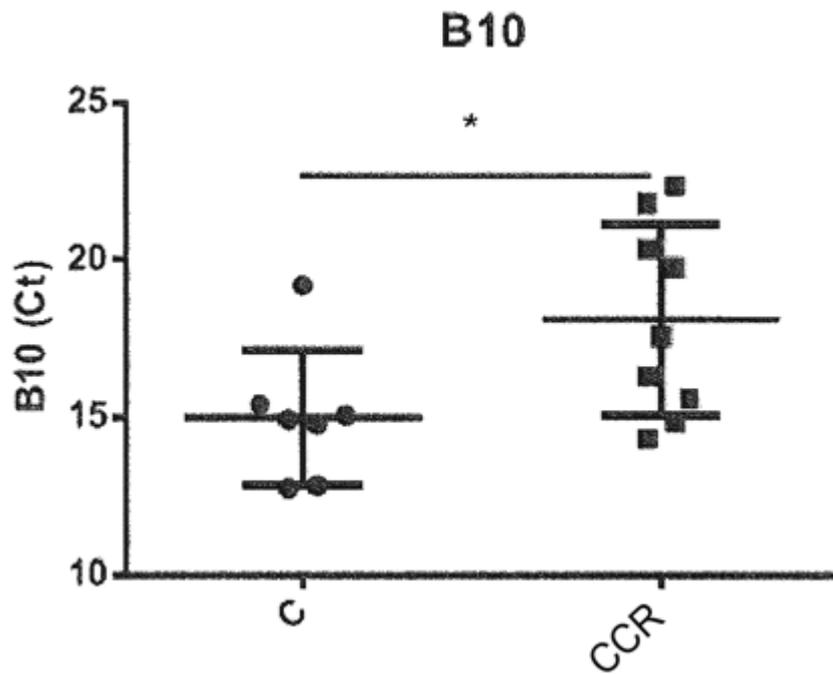


Fig. 4

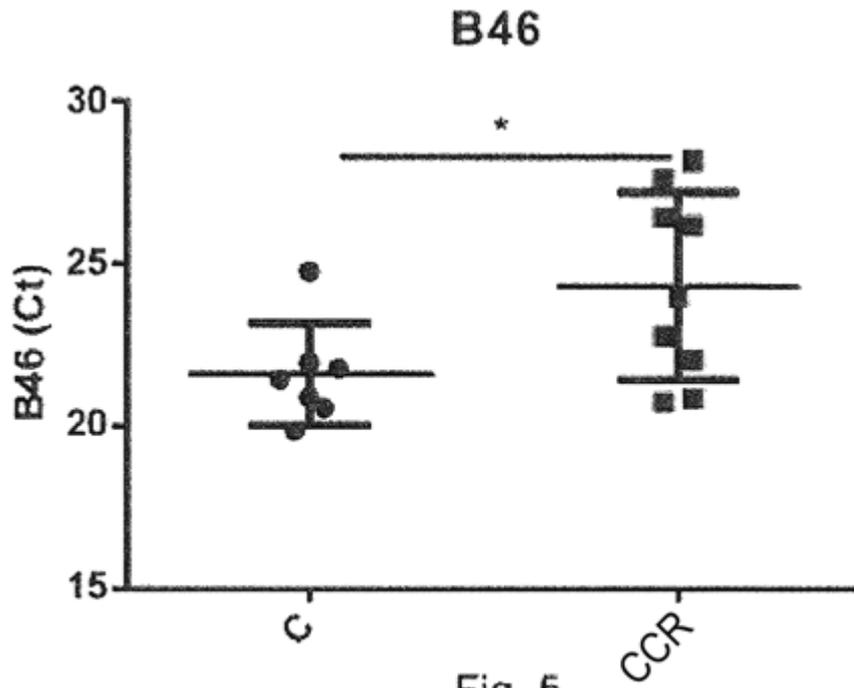


Fig. 5

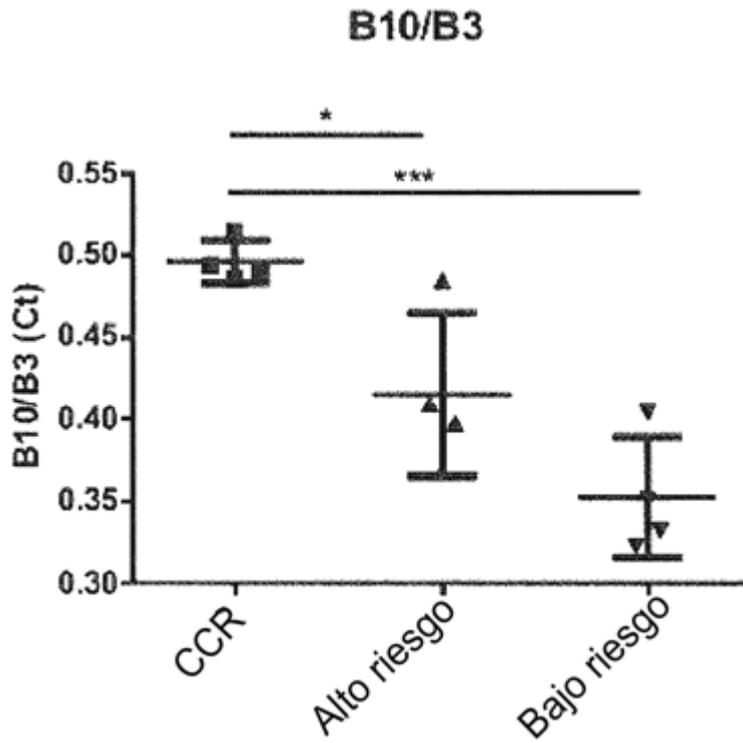


Fig. 6

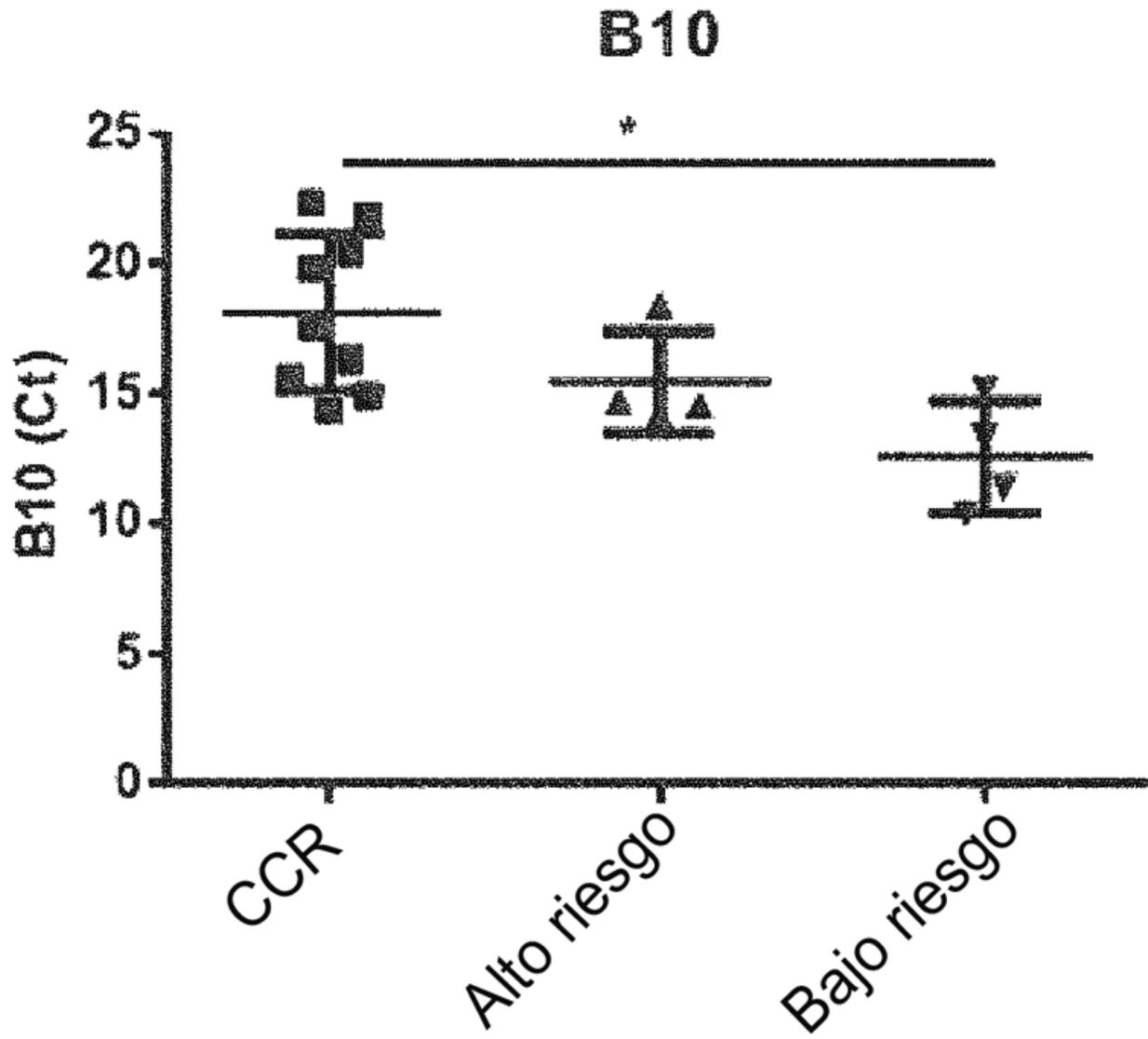


Fig. 7

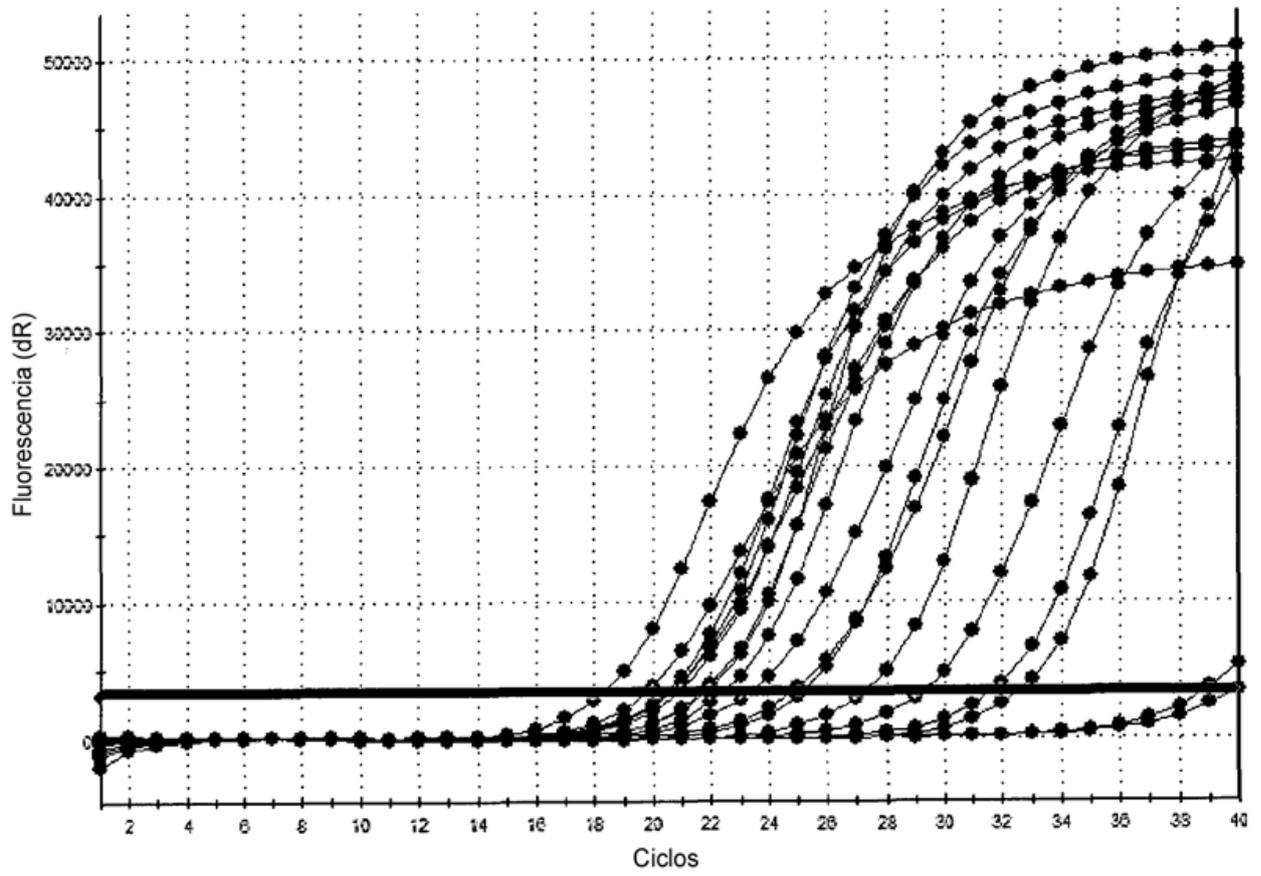


Fig. 8a

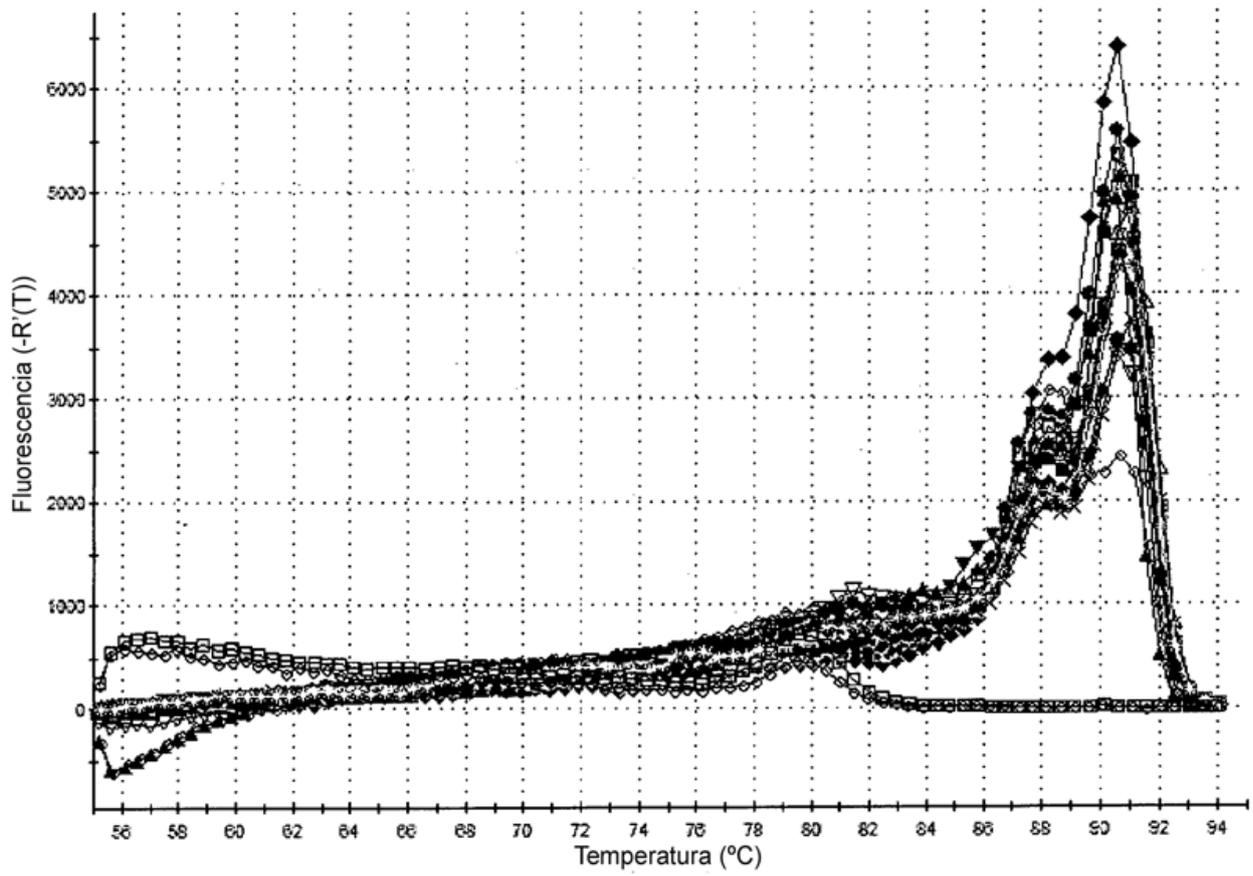


Fig. 8b

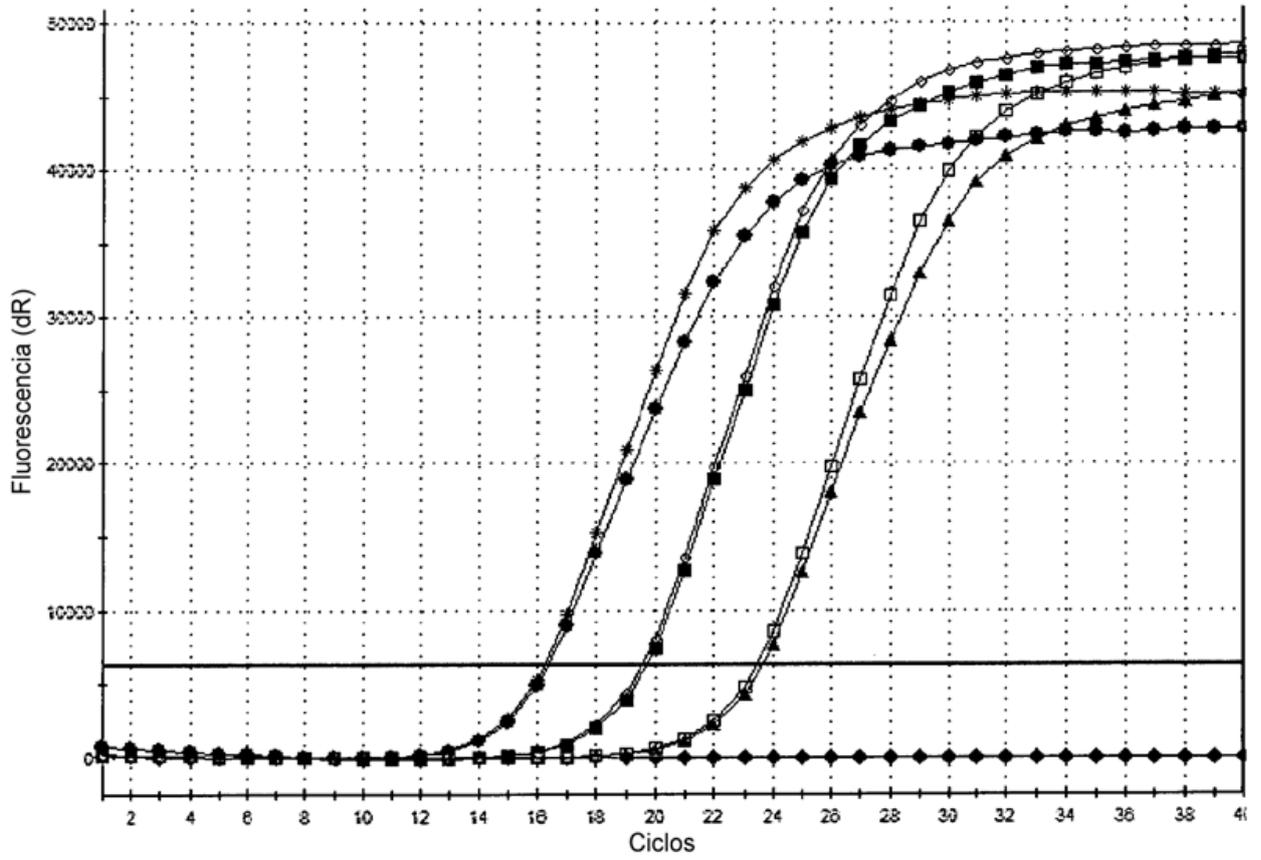


Fig. 9a

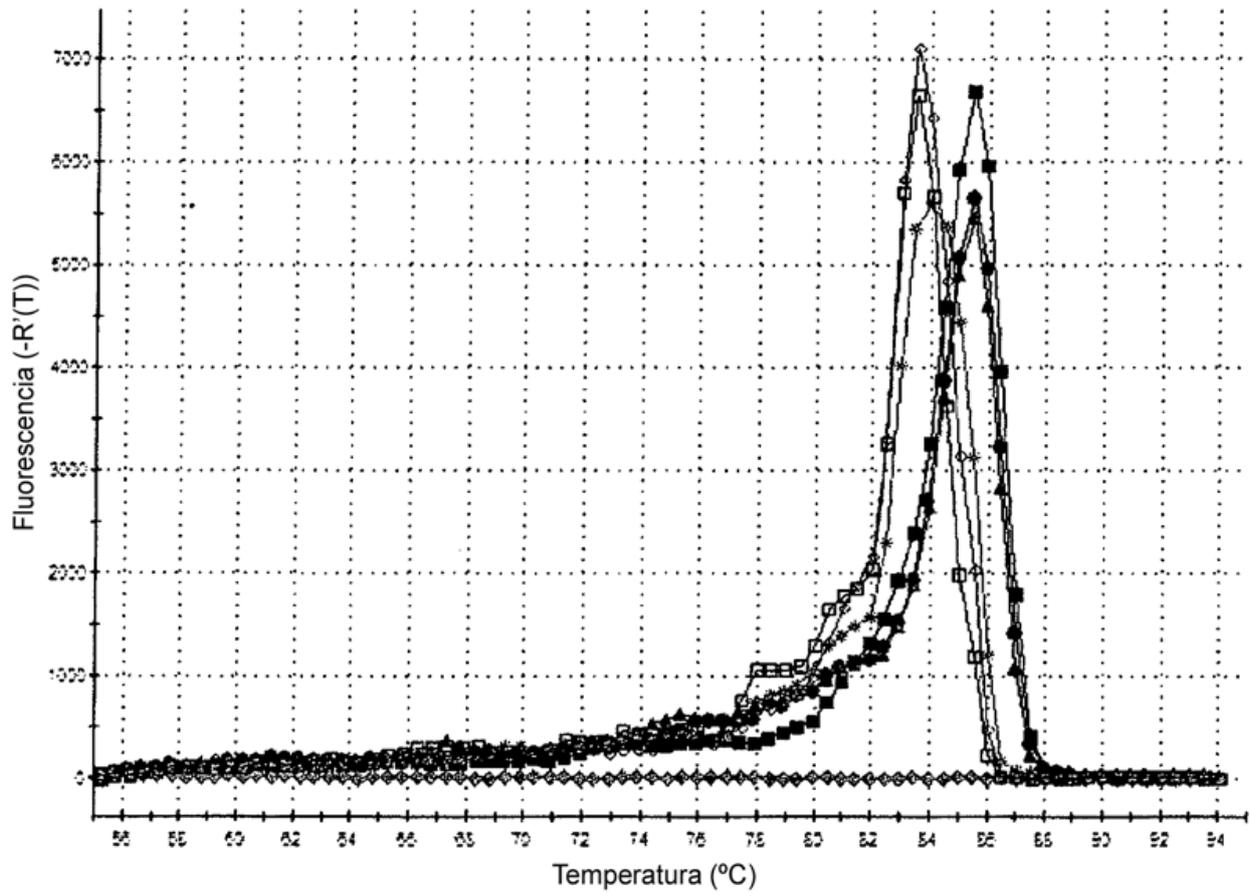


Fig. 9b

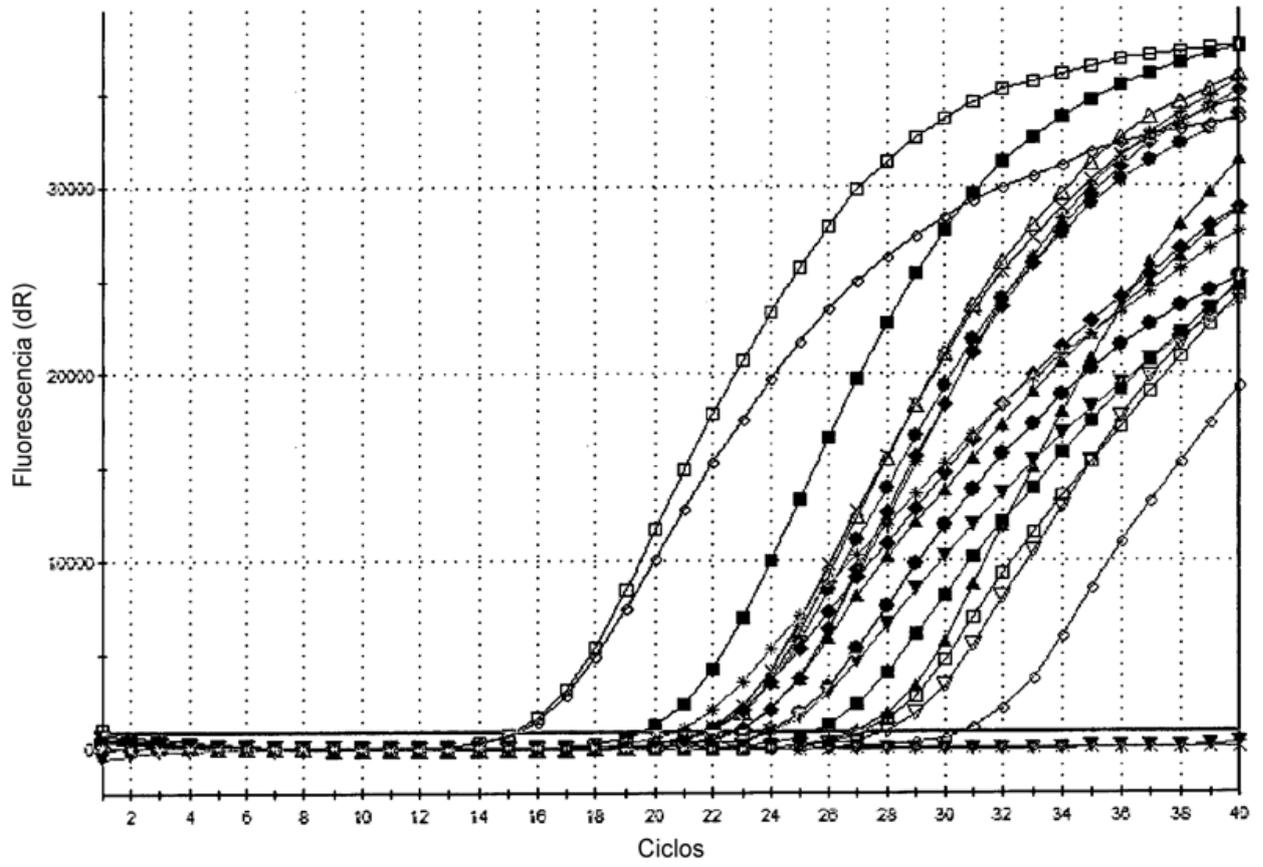


Fig. 10a

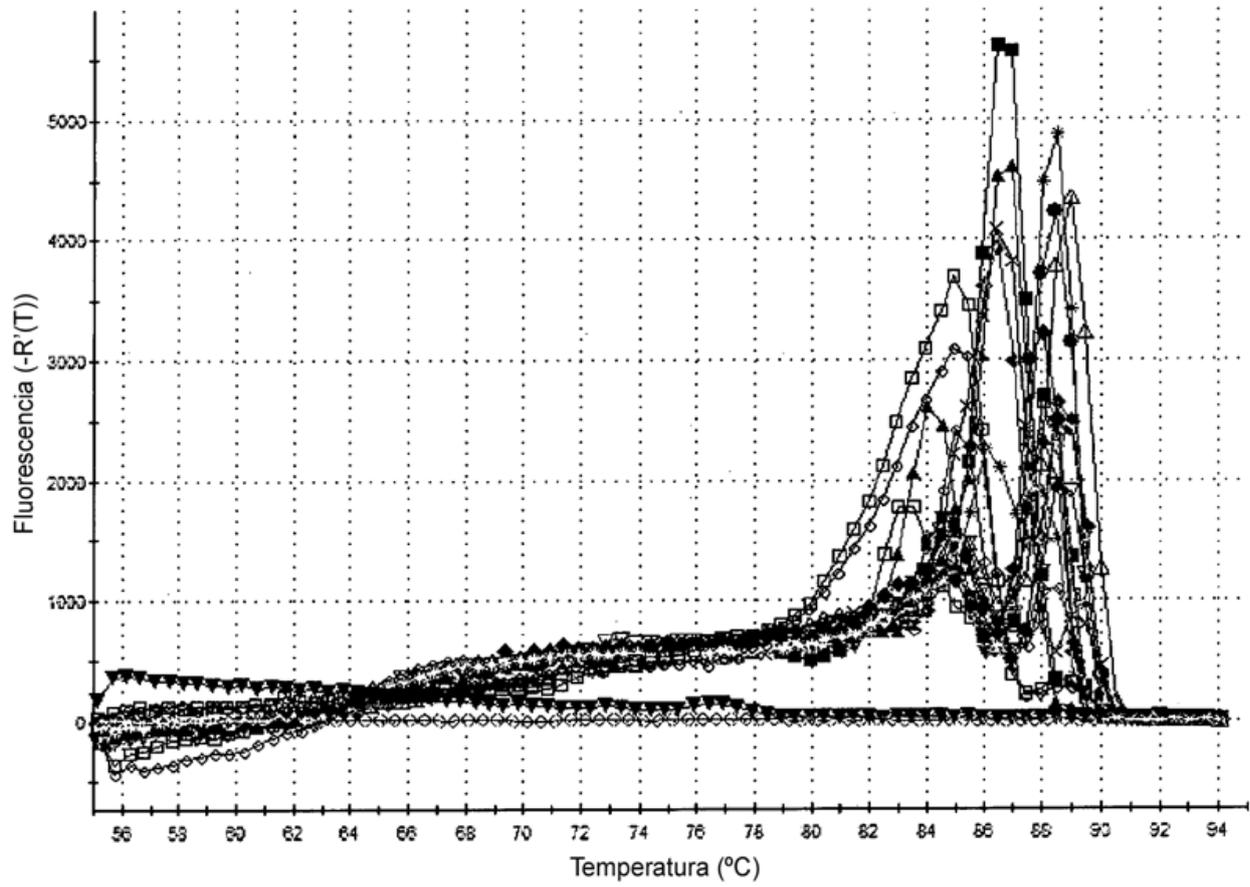


Fig. 10b

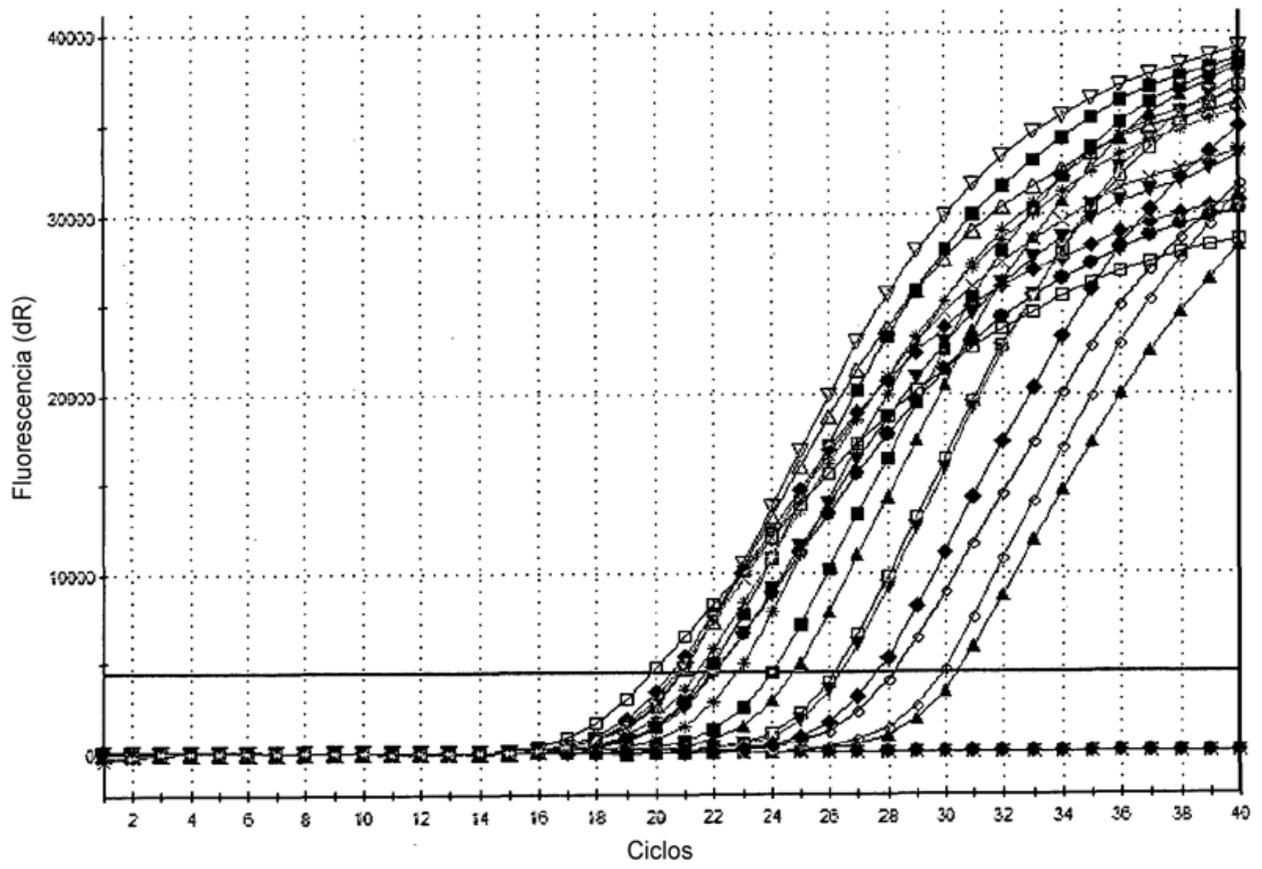


Fig. 11a

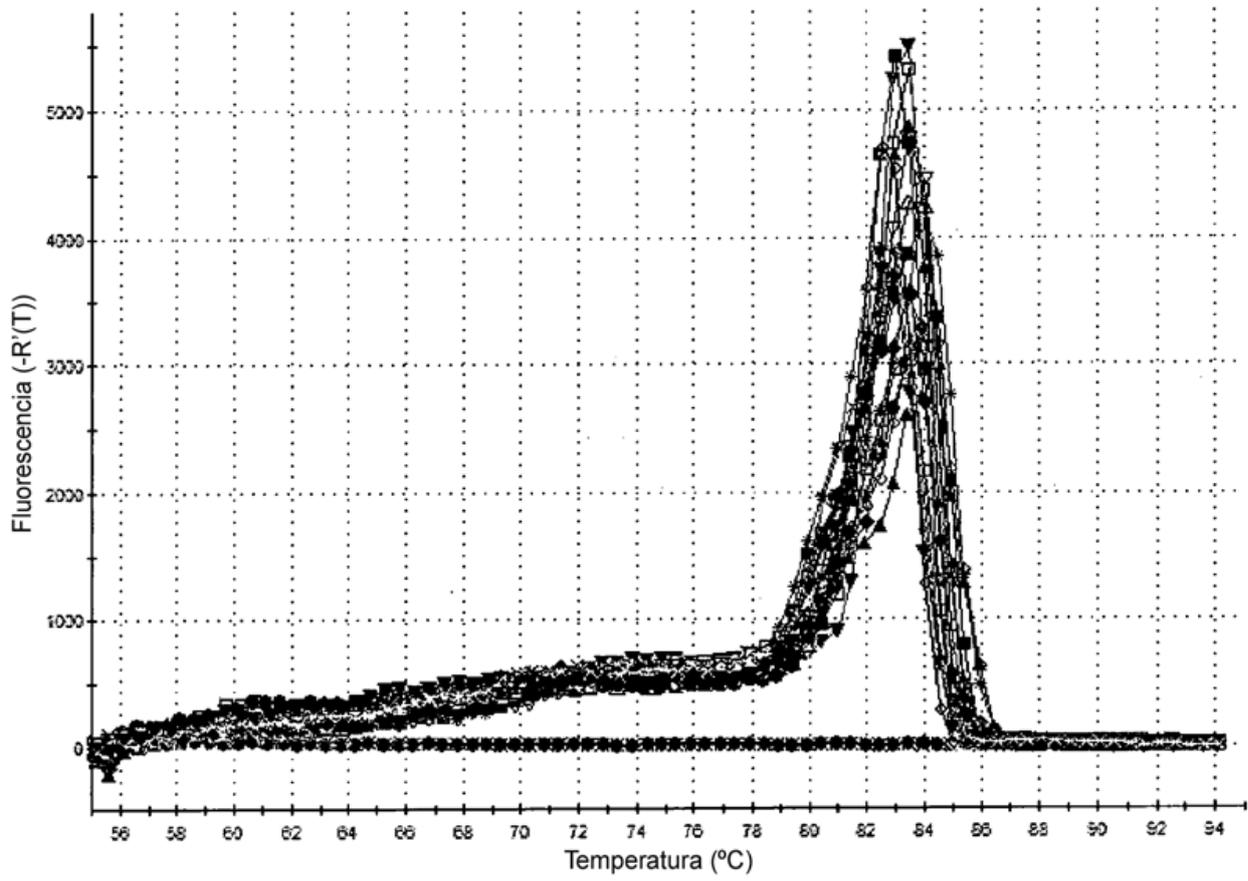


Fig. 11b

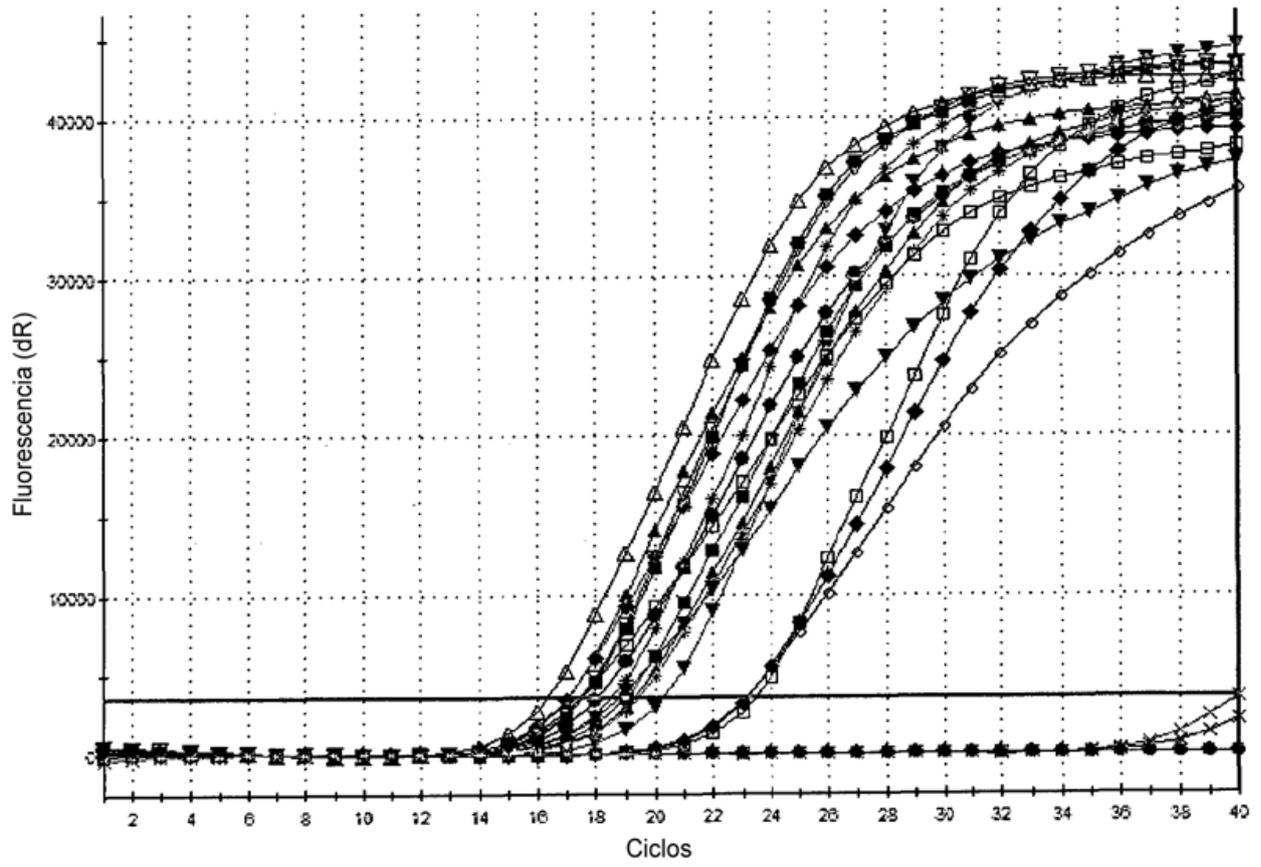


Fig. 12a

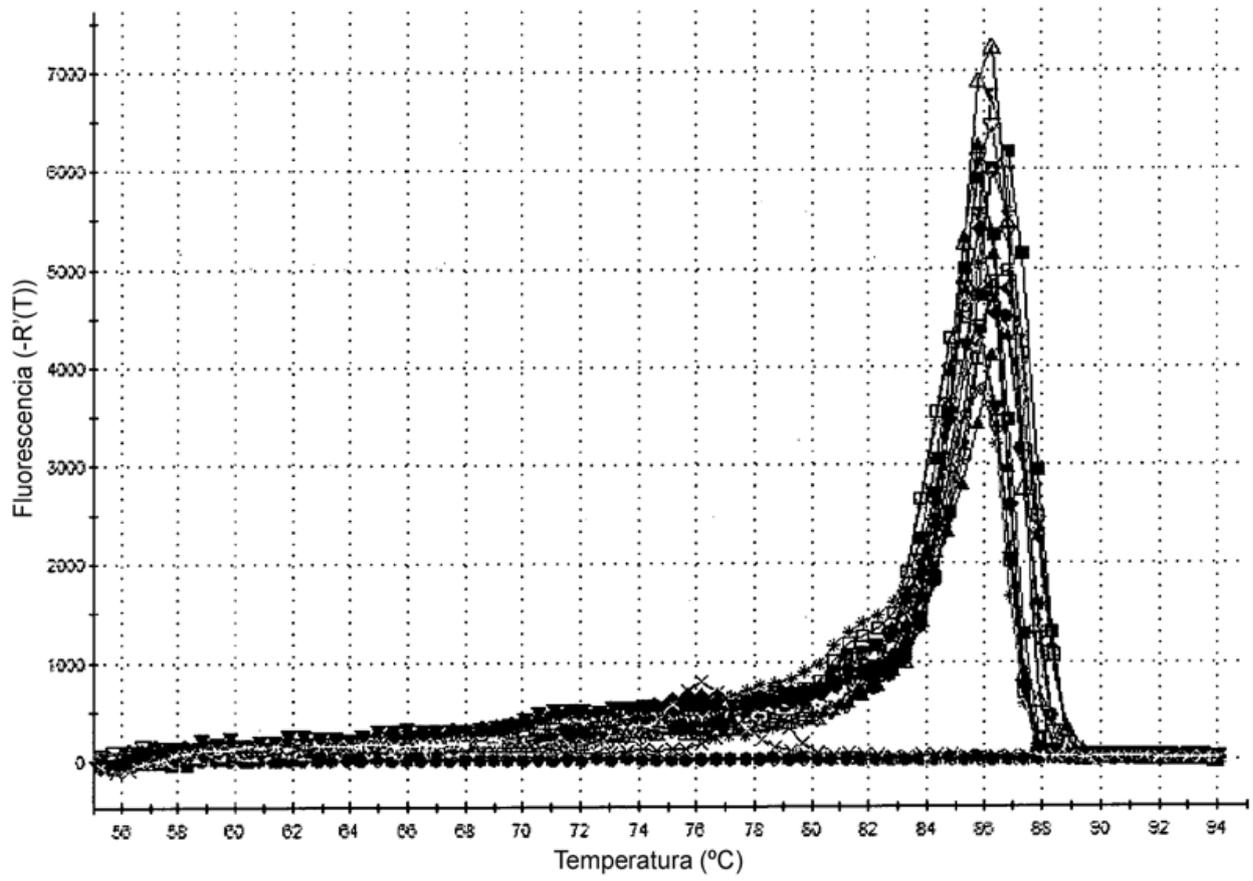


Fig. 12b

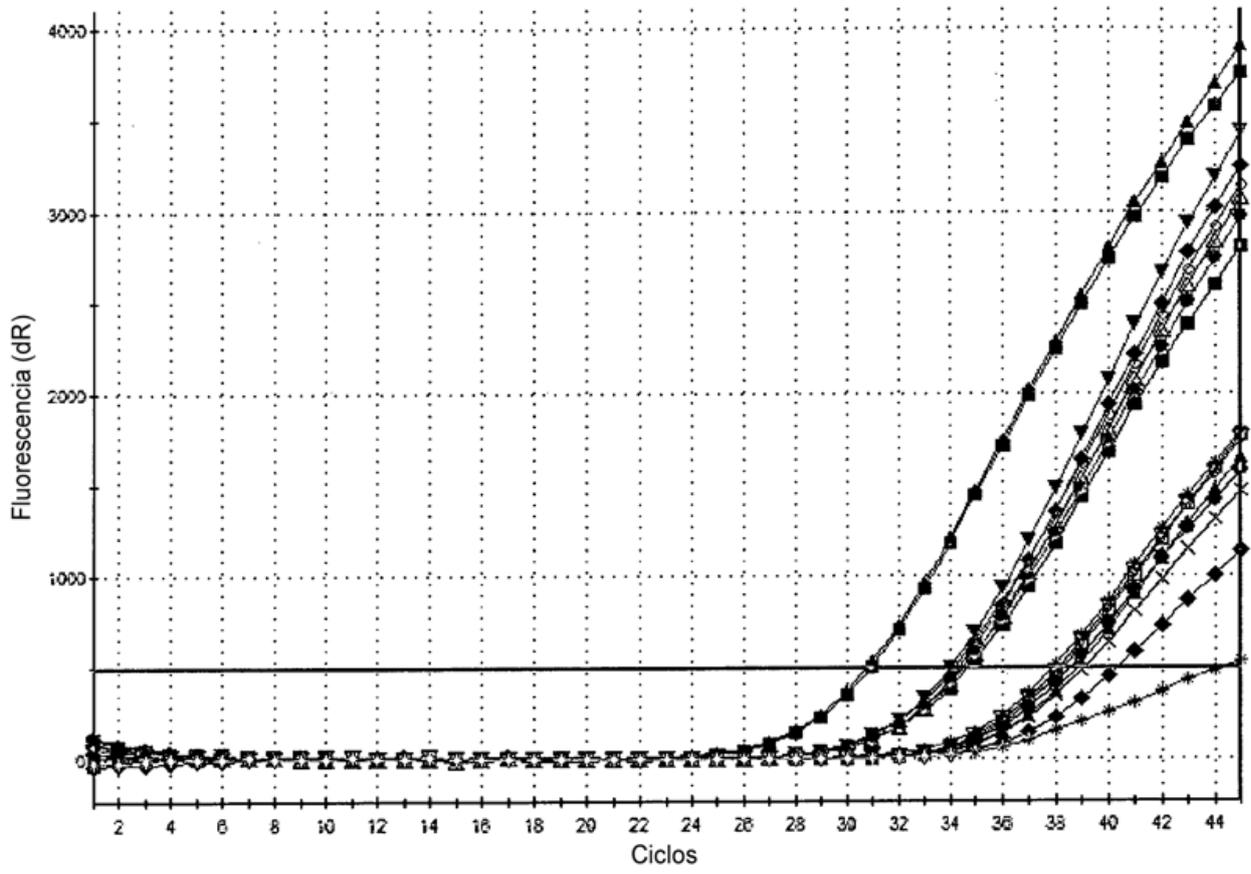


Fig. 13a

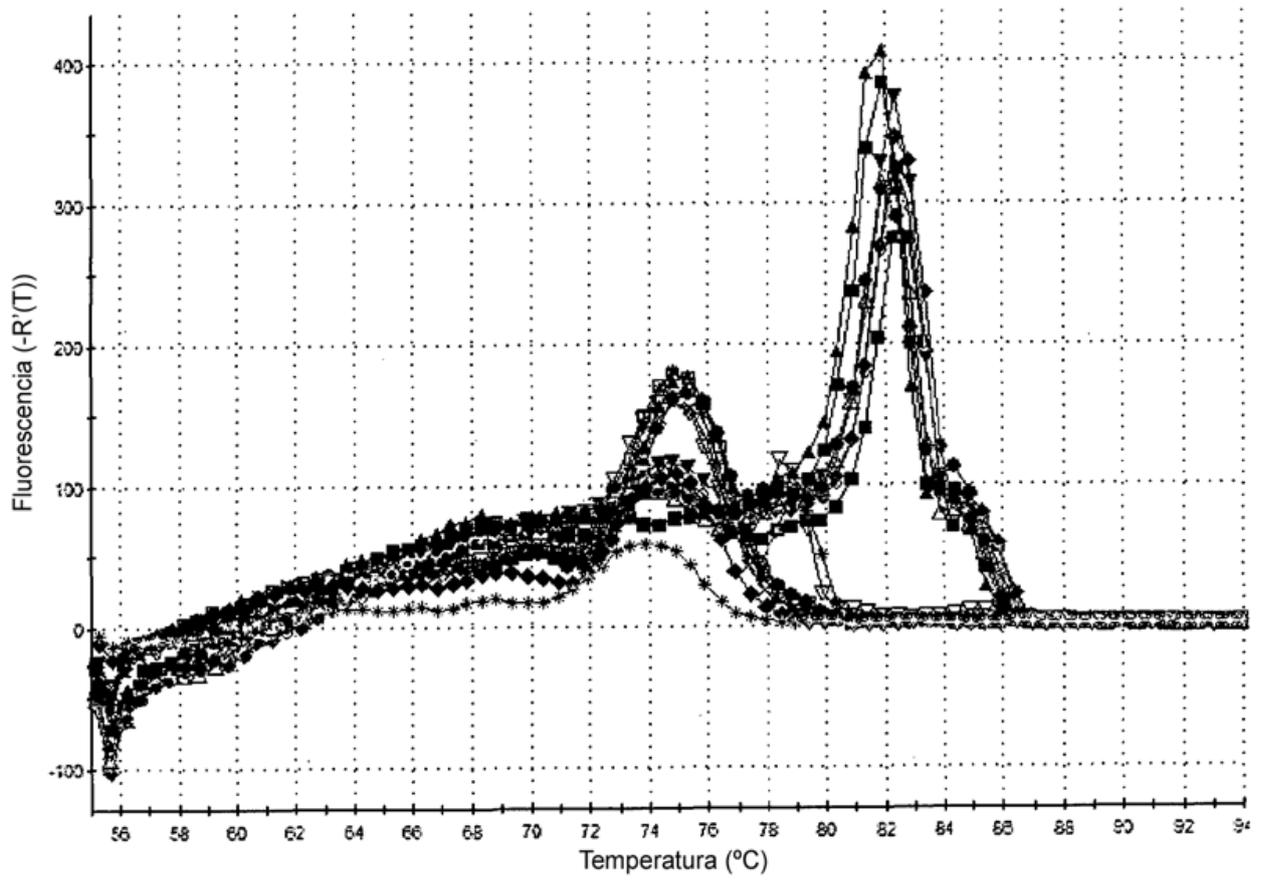


Fig. 13b

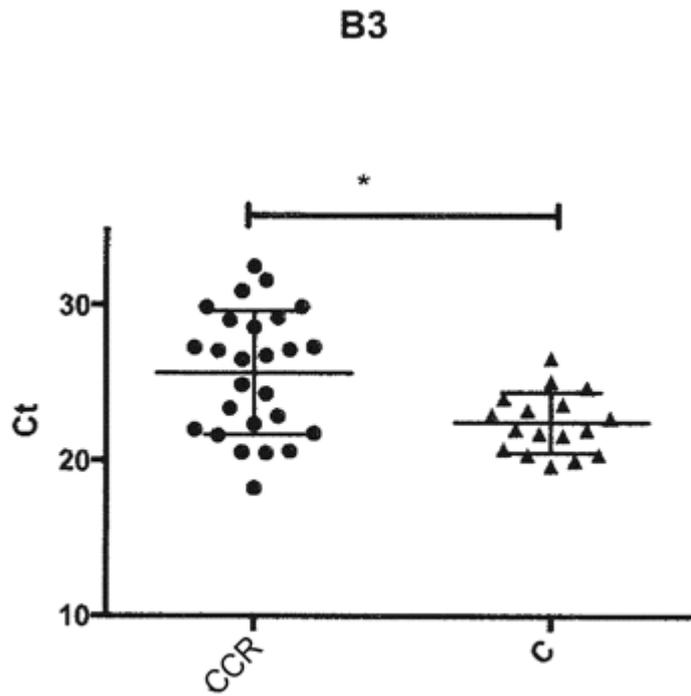


Figura 14. Valor de Ct de B3 en grupos de CCR y C.

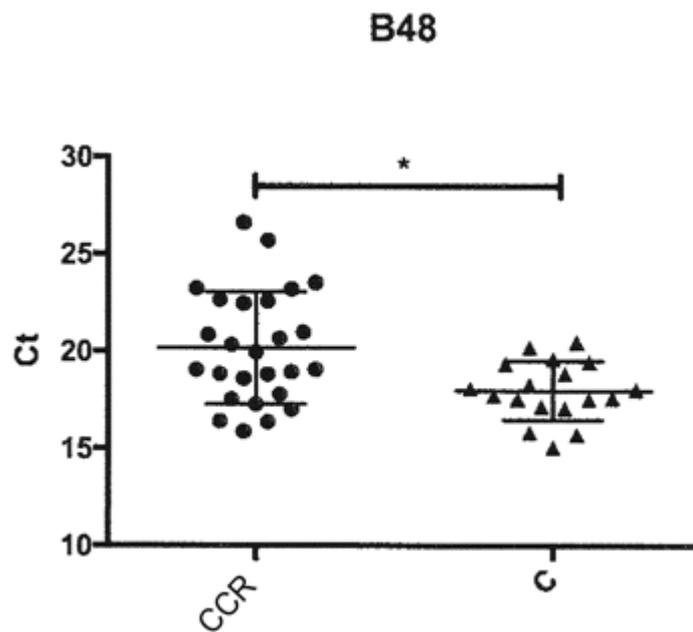


Figura 15. Valor de Ct de B48 en grupos de CCR y C.

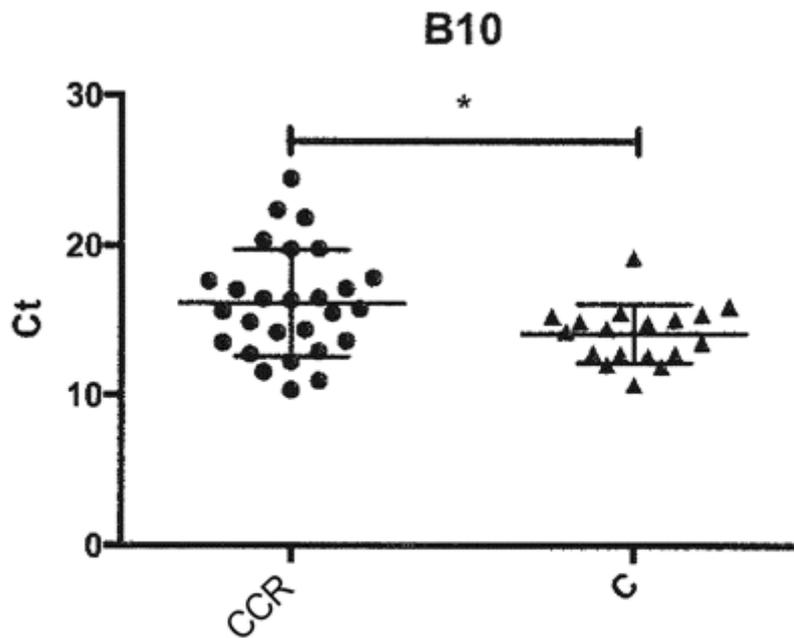


Figura 16. Valor de Ct de B10 en grupos de CCR y C.

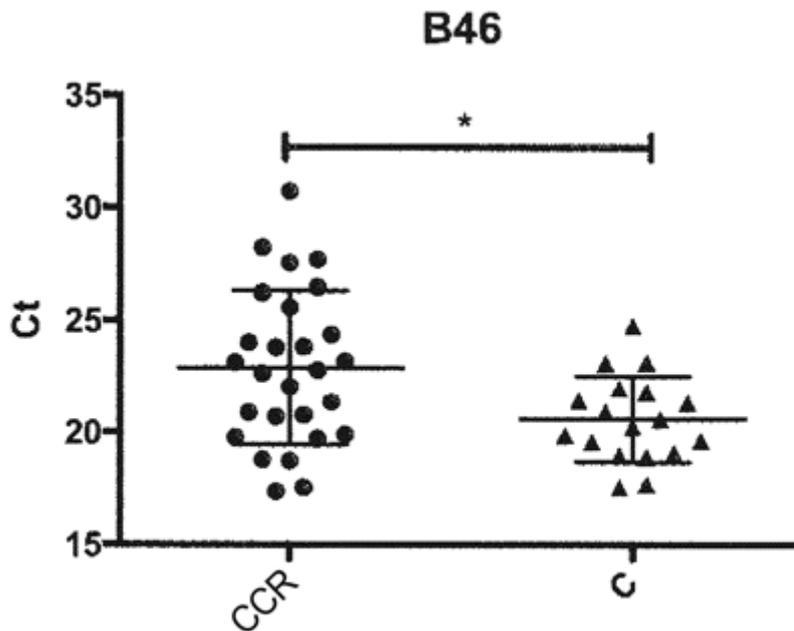


Figura 17. Valor de Ct de B46 en grupos de CCR y C.

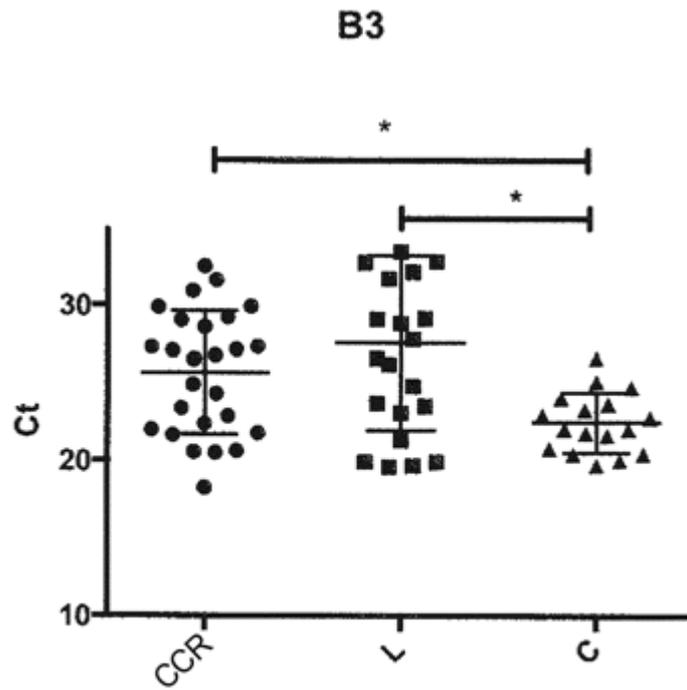


Figura 18. Valor de Ct de B3 en grupos de CCR, L y C.

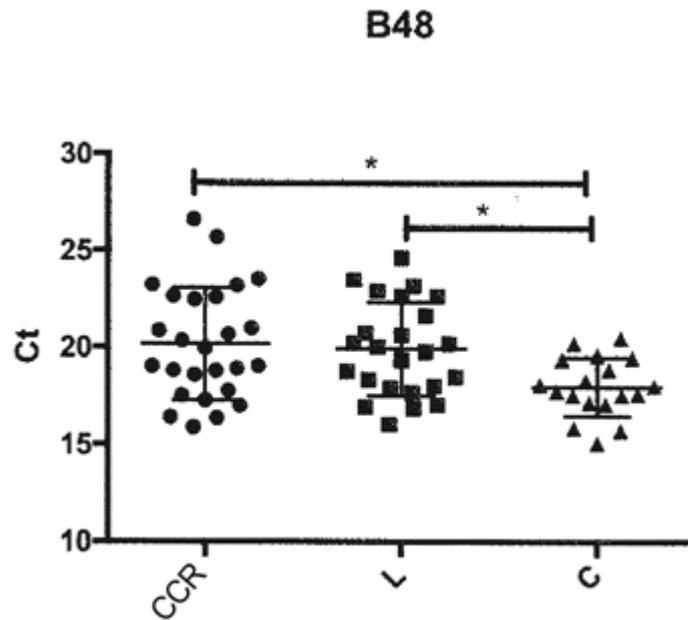


Figura 19. Valor de Ct de B48 en grupos de CCR, L y C.

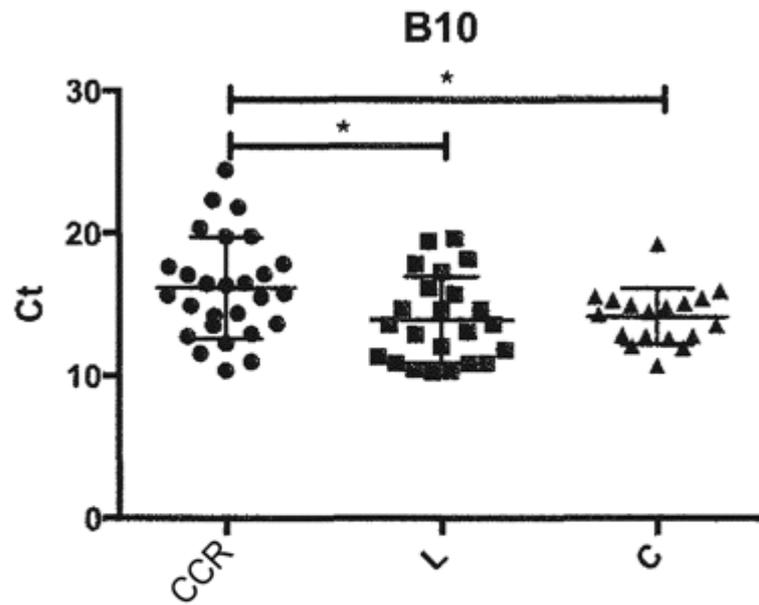


Figura 20. Valor de Ct de B10 en grupos de CCR, L y C.

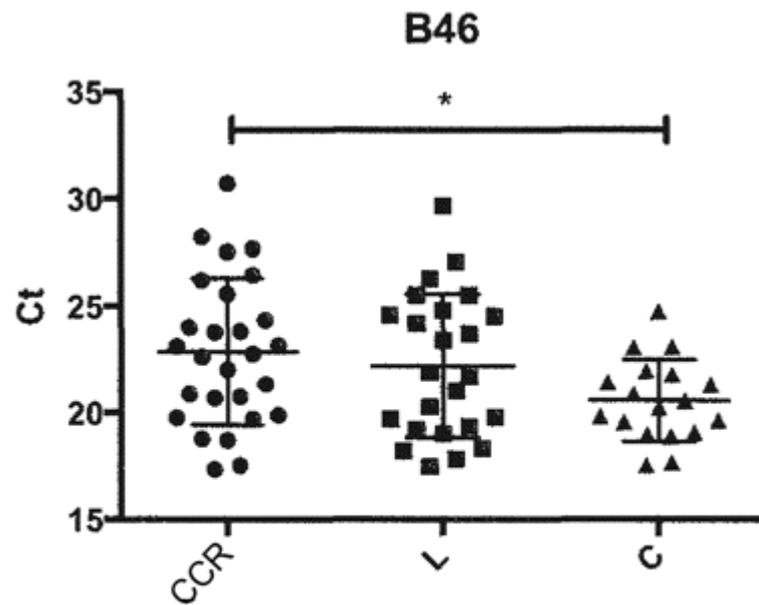


Figura 21. Valor de Ct de B46 en grupos de CCR, L y C.

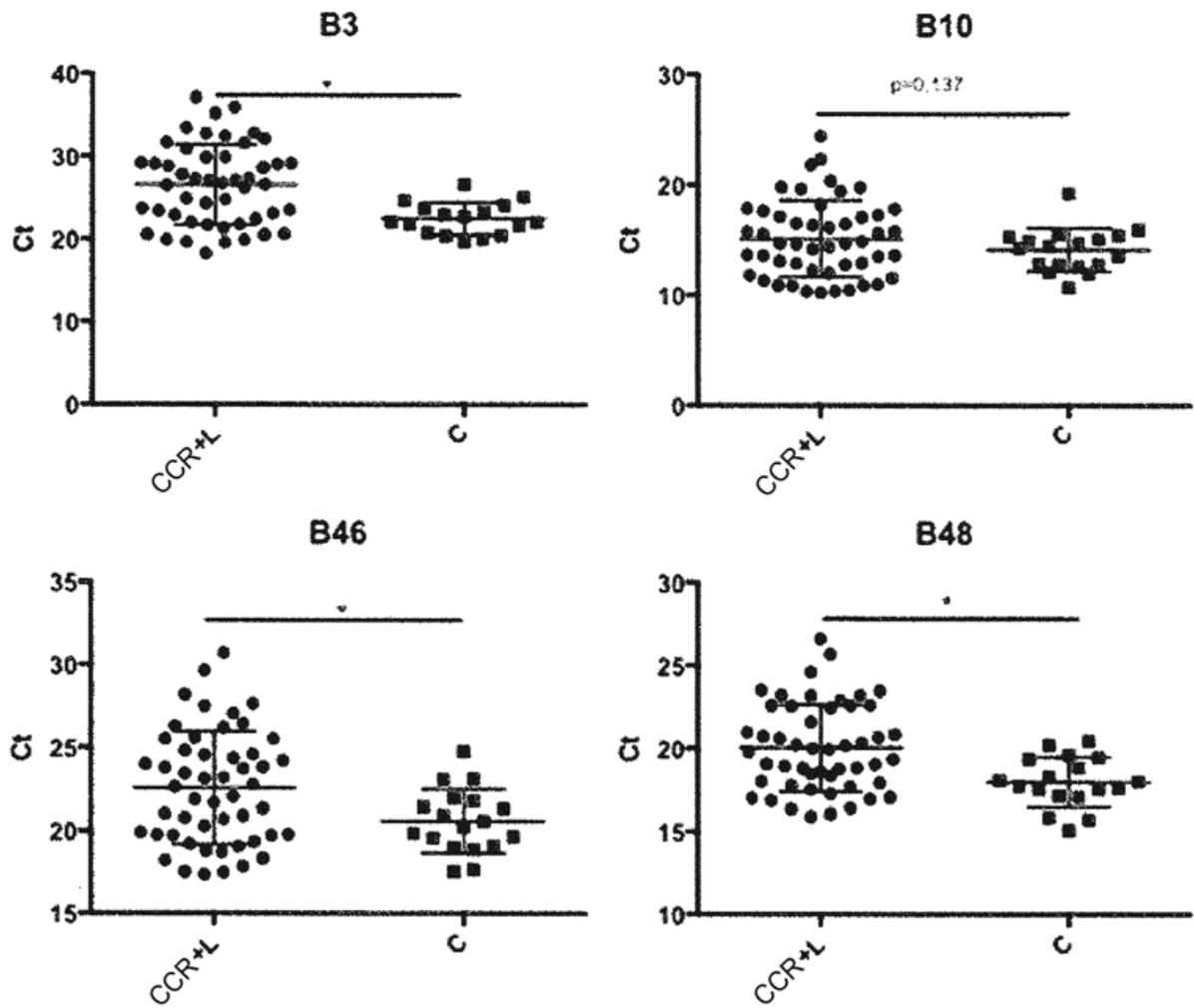


Figura 22: Valor de Ct de B3, B10, B46 y B48 en CCR+L frente a C

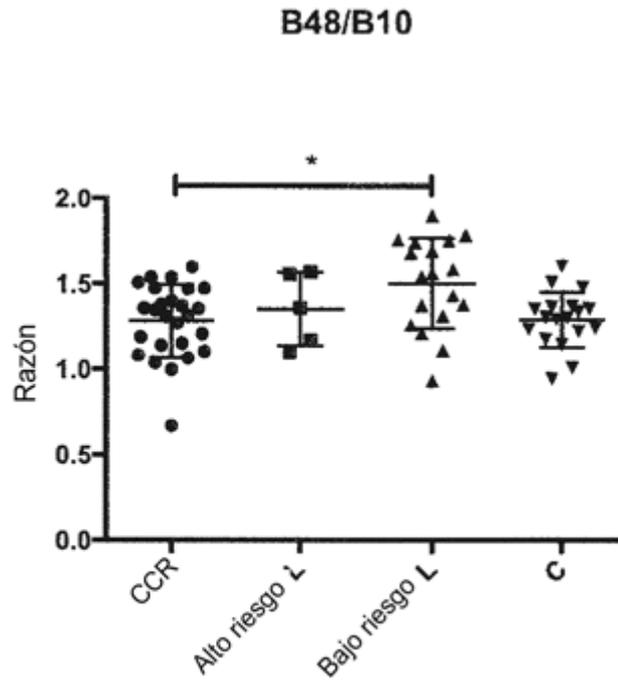


Figura 23. Valor de razón B48/B10 en grupos de CCR, alto riesgo L, bajo riesgo L y C.

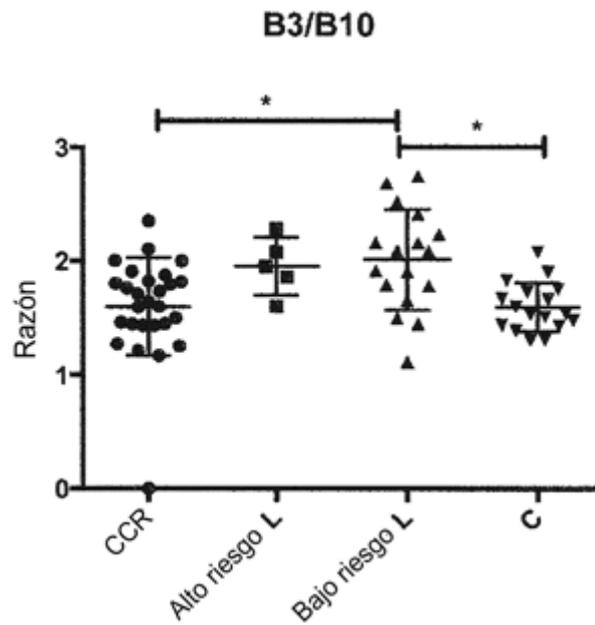


Figura 24. Valor de razón B3/B10 en grupos de CCR, alto riesgo L, bajo riesgo L y C.

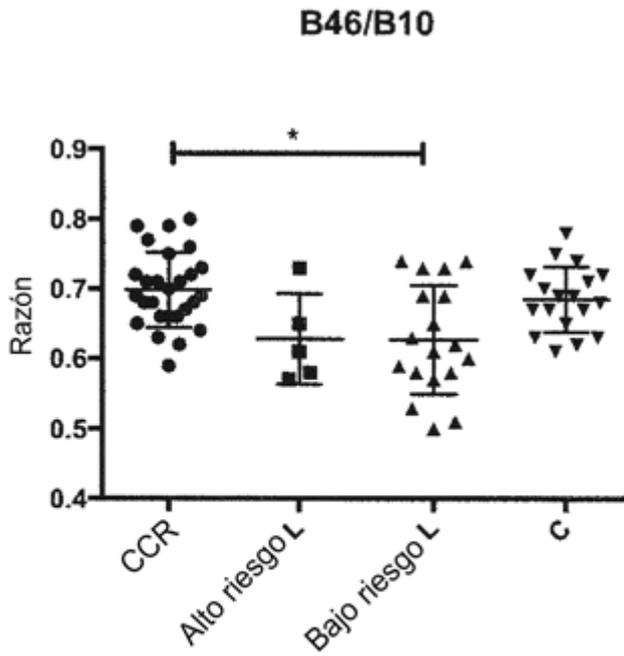


Figura 25. Valor de razón B46/B10 en grupos de CCR, alto riesgo L, bajo riesgo L y C.

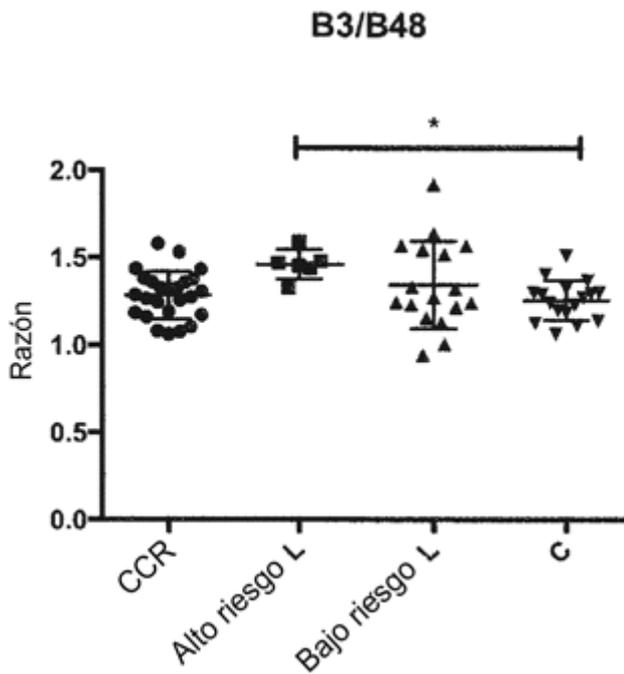


Figura 26. Valor de razón B3/B48 en grupos de CCR, alto riesgo L, bajo riesgo L y C.

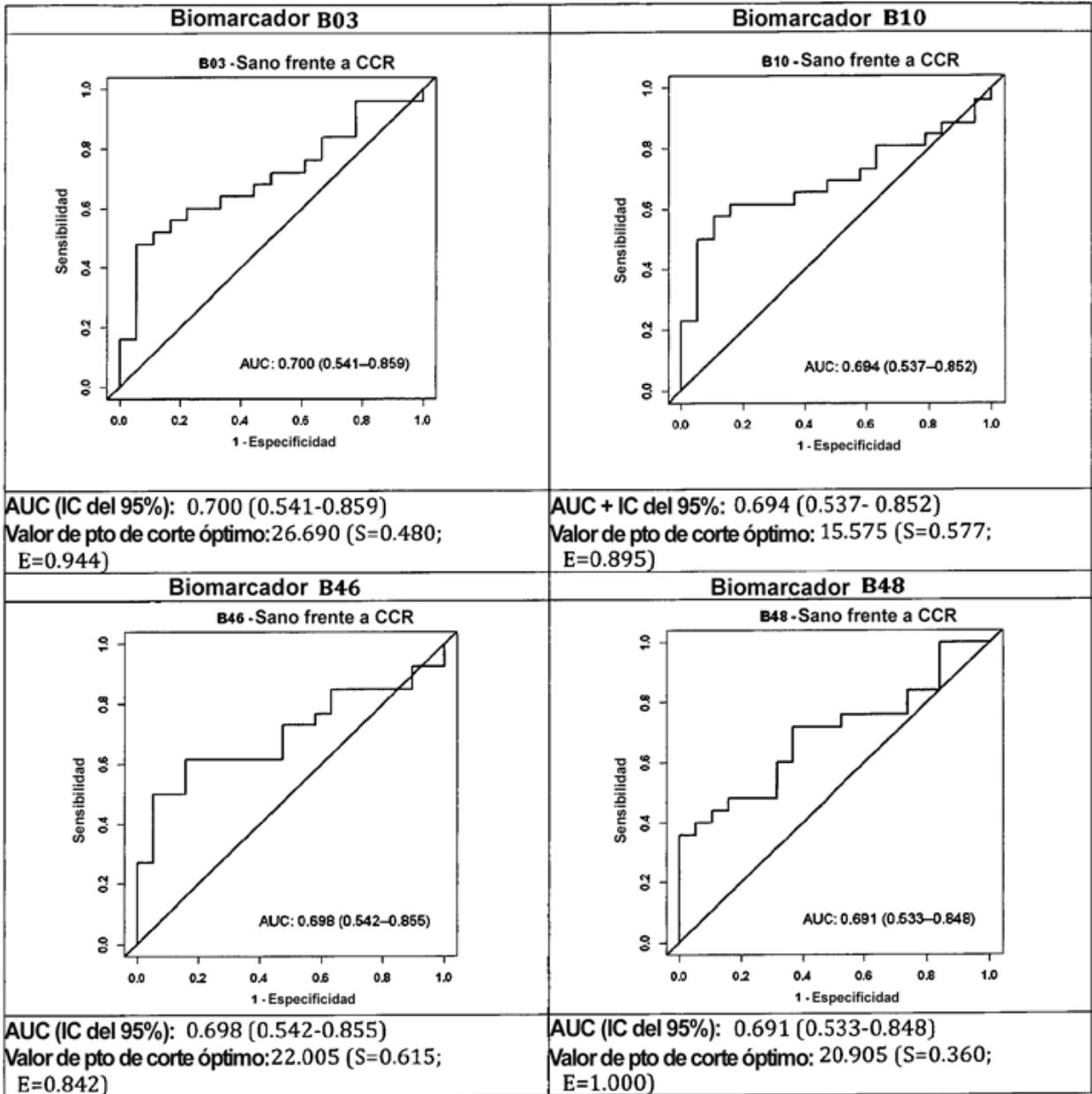


Fig. 27

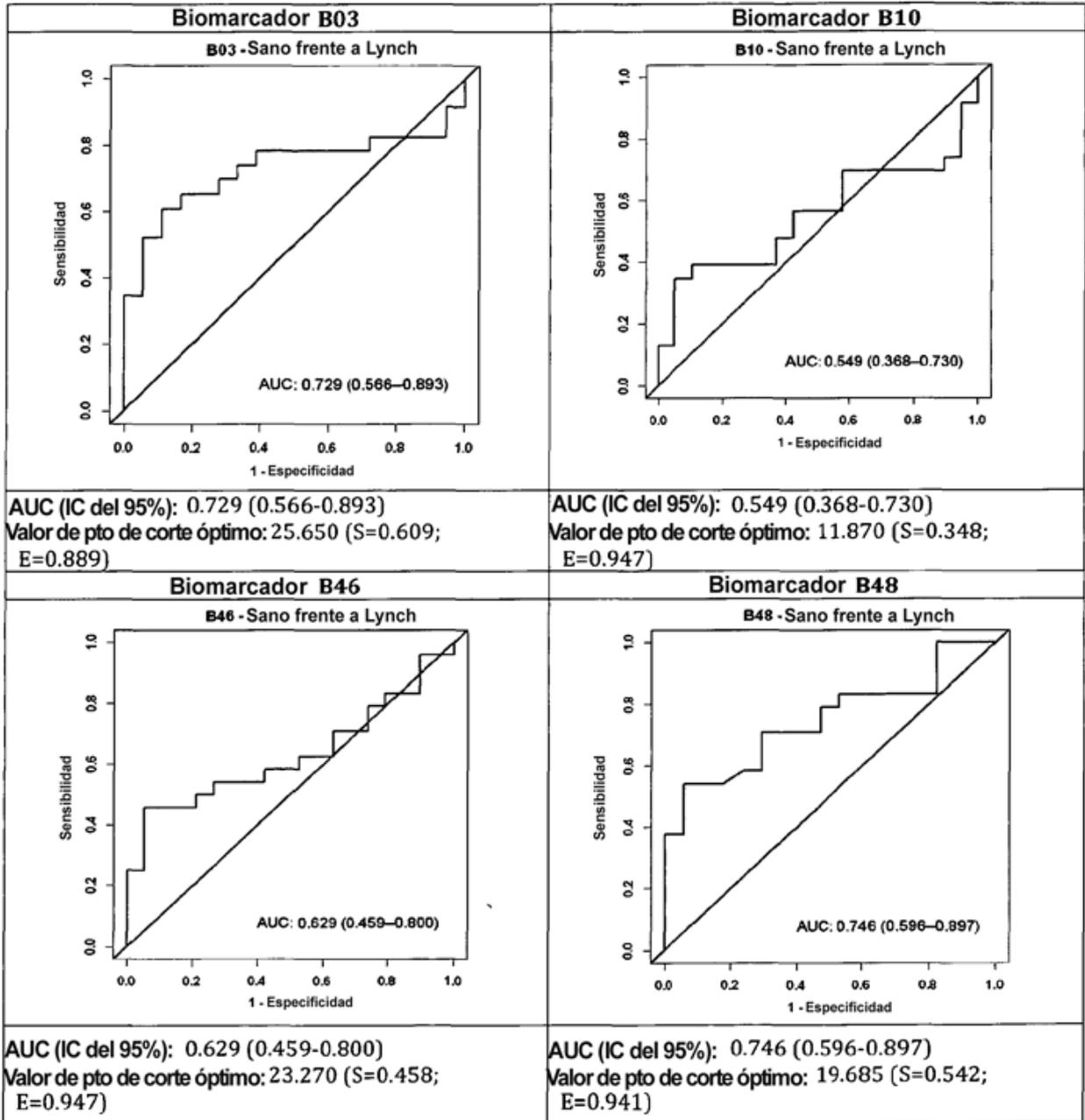


Fig. 28

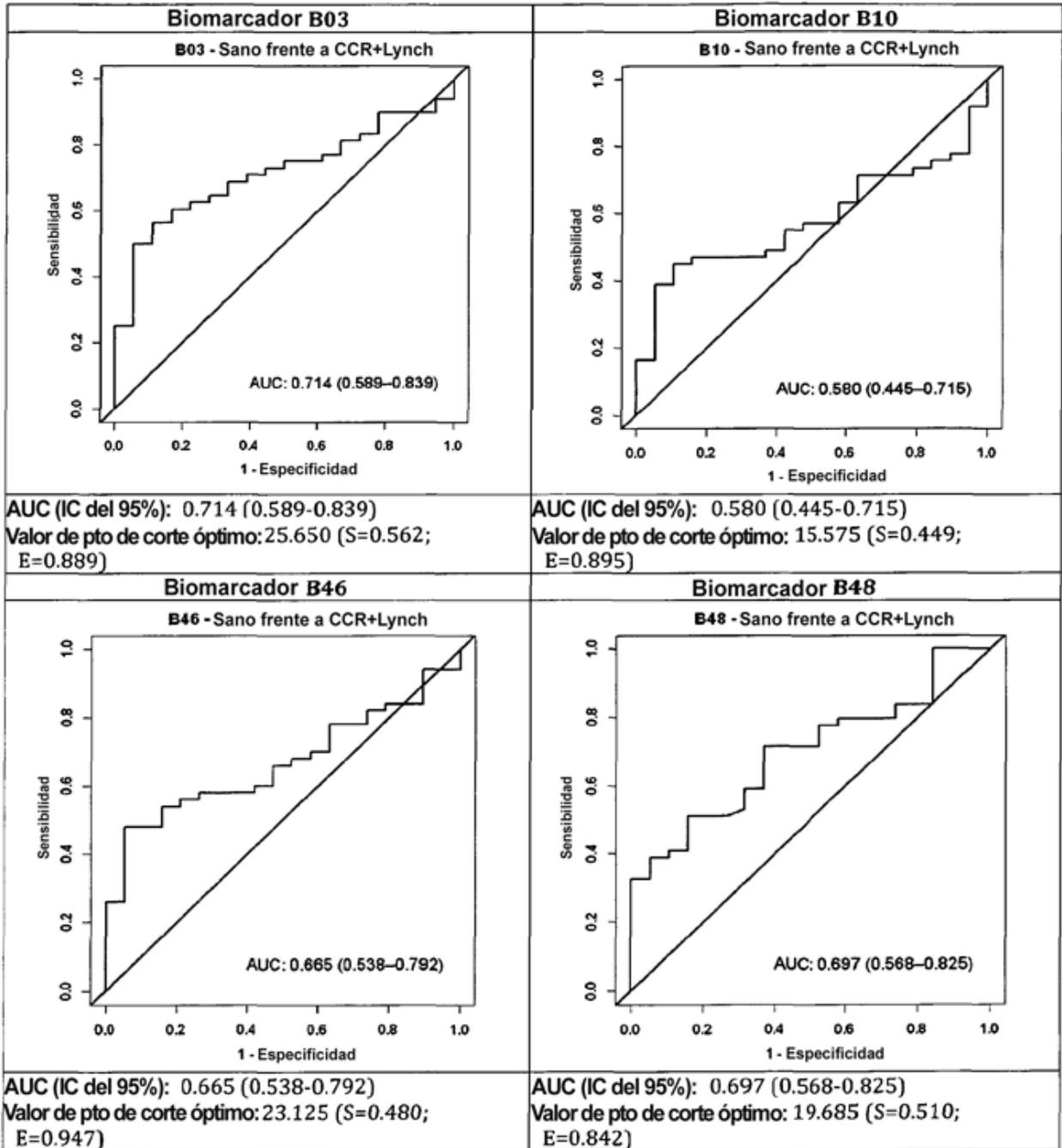


Fig. 29

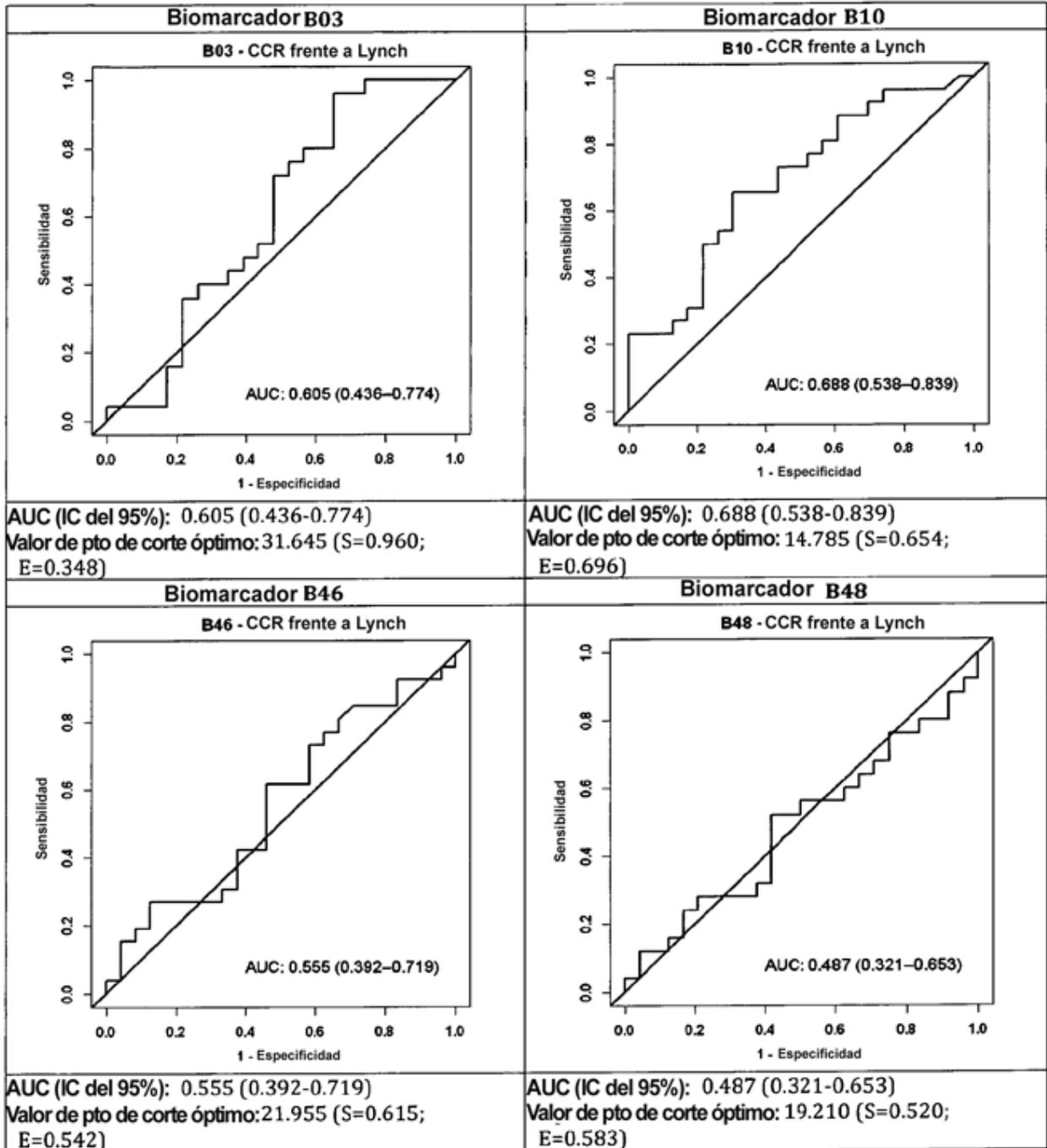


Fig. 30