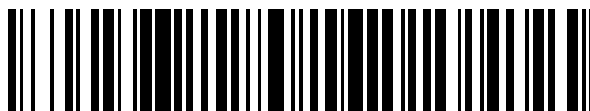


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 785 149**

51 Int. Cl.:

C12Q 1/6886 (2008.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **23.01.2018** **E 18153056 (9)**

97 Fecha y número de publicación de la concesión europea: **11.03.2020** **EP 3514247**

54 Título: **Panel de biomarcadores y métodos para detectar la inestabilidad de los microsatélites en cánceres**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
06.10.2020

73 Titular/es:

BIOCARTIS NV (100.0%)
Generaal De Wittelaan 11 B3
2800 Mechelen , BE

72 Inventor/es:

DE CRAENE, BRAM;
DECANNIERE, KLAAS y
VAN DE VELDE, JAN

74 Agente/Representante:

ELZABURU, S.L.P

ES 2 785 149 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

Panel de biomarcadores y métodos para detectar la inestabilidad de los microsatélites en cánceres

Campo técnico

5 La presente invención se relaciona, en general, con el campo del cáncer, en particular con cánceres que tienen inestabilidad de microsatélites (IMS) y/o deficiencia en la reparación de desapareamientos (RDA-). Como ejemplos de dichos cánceres, se incluyen muchos tumores colorrectales, gástricos y endometriales. Por consiguiente, la presente invención proporciona un nuevo panel de marcadores diagnósticos para analizar loci de IMS, junto con métodos y kits de uso de dicho panel en la detección de cánceres que tienen inestabilidad de microsatélites (IMS) y/o deficiencia en la reparación de desapareamientos (RDA-).

10 Antecedentes de la invención

15 Cada año se diagnostica en Europa y EE.UU. a aproximadamente 440.000 pacientes con cáncer colorrectal (CRC). Directrices recientes, tales como las Directrices NCCN para Pacientes: Cáncer de Colon Versión 1.2017, y las Directrices de Práctica Clínica ESMO sobre cáncer colorrectal de riesgo familiar, recomiendan el estudio de tumores en cuanto a la deficiencia en la reparación de desapareamientos (RDA) de ADN y/o el estado de IMS en todos los pacientes con CCR. A pesar de ello, actualmente estas pruebas están aún muy infrutilizadas debido a su complejidad técnica subyacente. En particular, es difícil estudiar todas las posibles mutaciones que subyacen a la deficiencia en la RDA y, aunque existen enfoques alternativos para su rastreo, los presentes ensayos aún requieren un amplio tiempo de manipulación en el laboratorio y, por lo tanto, no están listos para llegar a ser una rutina diagnóstica.

20 Los anteriores obstáculos afectan de manera natural al tratamiento y, por lo tanto también potencialmente, a los índices de supervivencia de muchos pacientes con cáncer. Ciertamente, en un subgrupo sustancial de casos de carcinoma colorrectal (CCR), se ve que las deficiencias en los genes de RDA son cruciales para la tumorigénesis y la progresión de la enfermedad. Por ejemplo, la silenciación epigenética del gen de RDA MLH1 representa alrededor del 12% del CCR. Otro 2-5% de casos están causados por una mutación de pérdida de función predominantemente heredada autosómica en uno de los genes de RDA MLH1, MSH2, PMS2 o MSH6. Este trastorno de susceptibilidad al cáncer familiar se conoce como el síndrome de Lynch o CCR hereditario no polipósico (CCRHNPP) y además lleva a un mayor riesgo de carcinoma gástrico y endometrial también (entre otros).

25 La ruta RDA implica a una gran cantidad de genes y se identificaron muchas lesiones genéticas y epigenéticas diversas que la afectan. Probablemente, aún quedan por identificar otras varias. Por lo tanto, es más práctico diagnosticar deficiencias en la maquinaria de RDA mediante rastreo de su resultado directo. Siendo esto último una acumulación a nivel de todo el genoma de errores de replicación del ADN que pueden observarse como cambios en el número de nucleótidos debido a delección o inserción en secuencias de repetición de uno sólo y de dos nucleótidos, por ejemplo (A)_n o (CA)_n. Este fenómeno se conoce como inestabilidad de los microsatélites o IMS. Cuando la deficiencia en la RDA conduce a IMS en regiones codificantes, da lo más frecuentemente como resultado mutaciones de promotores o de desplazamiento de marco que dan lugar a falta de expresión, expresión de proteínas truncadas y/o proteínas que contienen nuevas secuencias extensas que contienen neoantígenos. Además, se vio que la IMS en regiones límite de intrones-exones afecta a los mecanismos de ajuste del ARN y, por lo tanto, también interfiere con la traducción de las proteínas. En general, el fenotipo de IMS guarda correlación con la inestabilidad genómica, el mayor índice de mutación y, por consiguiente, el diferente comportamiento y pronóstico de los tumores.

30 Los tumores con IMS alta (IMS-A) tienen, en general, un mejor pronóstico y una probabilidad reducida de metástasis en comparación con tumores estables en cuanto a microsatélites (EMS). Además, estos dos tipos de tumores también reaccionan de manera diferente a diferentes tratamientos. Por ejemplo, los CCR con IMS precoz generalmente no responden a la quimioterapia basada en 5-fluorouracilo, que es actualmente el criterio estándar en el tratamiento del CCR (por ej., Webber *et al.*, 2015). Por otro lado, los tumores con IMS exhiben mayores niveles de al menos cinco moléculas de punto de control inmunitario que son dianas para inhibidores terapéuticos que están siendo actualmente estudiados clínicamente (Llosa *et al.* 2014). Por ejemplo, se espera que los pacientes con cáncer colorrectal con deficiencia en la reparación de desapareamientos responderán particularmente bien a la inmunoterapia anti-PD-1, que actúa bloqueando la interacción entre los receptores de PD-1 en las células T y los receptores de PD-L1 y PD-L2 en las células tumorales, incapacitando los mecanismos de evitación de su sistema inmunitario (Le *et al.*, 2015). Actualmente, se están estudiando también muchos otros compuestos y sustancias bioactivas, tales como camptotecina o irinotecano, para terapias tumorales dirigidas en base a la impronta molecular única de la IMS.

35 Por lo tanto, ya se reconoce que la identificación del estado de IMS en tumores puede tener un enorme impacto en el resultado del tratamiento y, por consiguiente, también en la calidad de vida y la esperanza de vida de muchos pacientes con cáncer. Se demuestra esto mejor por el hecho de que muchas guías oficiales ya recomiendan abiertamente el estudio de la IMS en el cáncer de colon y el síndrome de Lynch. Éstas incluyen, por ej., las Guías NCCN para el Cáncer de Colon, las Guías de Práctica Clínica ESMO sobre el cáncer colorrectal de riesgo familiar, las Guías de Bethesda Revisadas, los criterios clínicos de Ámsterdam II, el Equipo de Trabajo Multisociedad de EE.UU. sobre el Cáncer Colorrectal, etc.

Actualmente, se usan lo más frecuentemente dos técnicas para el estudio de la IMS: inmunohistoquímica (IHQ) y electroforesis capilar. La IHQ es una técnica cara, de mucho trabajo y que lleva tiempo, con un alto índice de resultados falso negativo. En la electroforesis capilar, se usa PCR fluorescente para amplificar regiones genómicas específicas que contienen repeticiones de nucleótidos en células tumorales y células normales, después de lo cual se determina la presencia de la inestabilidad comparando la longitud de los productos de amplificación. Existen cientos de miles de loci de microsatélites por todo el genoma que pueden usarse potencialmente en el análisis de la IMS (*Ellegren, Nat Rev Genet. 2004*).

Por ejemplo, el panel de consenso de la IMS establecido en 1997 y conocido como el panel Bethesda incluye 5 marcadores microsatélite, que incluyen 2 repeticiones mono- u homonucleotídicas de 25 y 26 nucleótidos de longitud (BAT25 y BAT26, respectivamente) y 3 repeticiones dinucleotídicas (D2S123, D5S346, D17S250) (*Boland et al., 1998*). Se designa una muestra estudiada con el panel Bethesda como poseedora de una alta frecuencia de IMS o un fenotipo "IMS-A" si un 30% o más de los marcadores (por lo tanto, al menos 2 en el panel de 5 marcadores) se estudiaron como inestables. Si un marcador de los cinco (o <30% de los marcadores tumorales) puntúa como positivo a IMS, se designa una muestra como baja en IMS o "IMS-B". Finalmente, si no se encuentra alterado ningún marcador, se considera que una muestra es estable a la IMS o "EMS" (*Boland et al., 1998*).

Sin embargo, a pesar de ser el patrón actual de estudio de la IMS, el panel Bethesda adolece de varios inconvenientes, tales como una desigual prevalencia de los loci en diferentes poblaciones étnicas y diferentes tipos de tumores. En particular, tiende a mostrar baja sensibilidad, especialmente en cánceres distintos del cáncer colorrectal, en vista de lo cual se desarrolló inicialmente (*Boland et al., 1998*). Éstos y otros factores han llevado a su expansión y/o diversificación con marcadores adicionales por parte de clínicos individuales y laboratorios de investigación, lo que por consiguiente ha dado como resultado una pérdida de estandarización y una escasa reproducibilidad. Como ejemplos de lo anterior, se incluyen, por ej., *Murphy et al., 2006* y WO2006047412 (Promega). De manera alternativa, también se describieron marcadores microsatélite completamente nuevos que no se solapan con ninguno del panel Bethesda, por ej., en WO2013153130 (VIB), Zhao *et al.*, 2014 (eLife), Maruvka *et al.*, 2017 (Nature Biotechnology), y Endris *et al.*, 2017 (AACR Annual Meeting abstract).

Otro inconveniente de los enfoques actualmente conocidos es su nivel de complicación y la necesidad de instrumentos especializados que se extienden más allá de los termocicladores de laboratorio estándar, así como su limitada idoneidad para automatización. El propio estudio con el panel Bethesda clásico es una prueba de tubo abierto, que aumenta la probabilidad de contaminación cruzada. Además, requiere personal de laboratorio especializado y lleva tiempo, es caro y de mucho trabajo. En general, las técnicas de detección actualmente existentes de IMS aplican uno de los principios siguientes: (i) uso de cebadores fluorescentemente marcados para la detección de los marcadores del panel Bethesda, seguido de electroforesis capilar; (ii) análisis de la curva de fusión de alta resolución de los 5 marcadores del panel Bethesda usando un tinte que se intercala en el ADNdh; (iii) detección espectrométrica de masas de alelos de una longitud diferente; y (iv) secuenciación de la siguiente generación (SSG) de grandes regiones de ADN (por ej., exoma), seguido de recuento del número de mutaciones, o de un número de regiones homopoliméricas en una disposición no apareada (Campbell *et al.*, 2017, Cell).

En (i), por ejemplo, la estrategia de rastreo Bethesda basada en PCR inicial requiere una interpretación de un observador experto que impide una automatización efectiva y sencilla. Luego, en cuanto a (ii), el análisis de la curva de fusión de alta resolución con tintes que se intercalan en el ADNdh sufre de capacidades de multiplexación muy limitadas para rastrear varios marcadores de IMS diferentes en una tanda, ya que la temperatura de fusión para cada amplicón de marcador necesita ser suficientemente diferente para no producir señales de solapamiento. Además, como esta estrategia se basa en la formación de heterodúplex entre alelos de longitud normal y mutante, también es menos sensible en comparación con las otras alternativas. A continuación, en cuanto a (iii), el método basado en espectrometría de masas (Zhao *et al.*, 2014) es, en principio, también susceptible de automatización, pero requiere instrumentación especializada y personal altamente cualificado para la interpretación de los datos. Finalmente, con respecto a (iv), aunque SSG tiene sin duda la ventaja de mirar a un muy gran número de posiciones indicativas de IMS en el genoma o exoma, más que sólo a los marcadores selectivos, y, aunque este método es también, en principio, al menos parcialmente automatizable, es actualmente muy caro y requiere hardware de SSG especializado. Con respecto a la puntuación de homopolímeros, la SSG aún no es lo suficientemente robusta como para puntuar repetidamente repeticiones homopoliméricas individuales, ya que aún es propensa a perder información sobre indeles de nucleótidos simples en una hebra de nucleótidos repetitivos. Además, debido a la generación de grandes cantidades de datos, sigue llevando tiempo, es complicada y requiere a un analista altamente entrenado.

En conclusión, el estudio de la IMS representa una necesidad médica muy grande que actualmente sólo se satisface parcialmente mediante los métodos de diagnóstico existentes debido a sus restricciones técnicas. Estas incluyen de manera importante capacidades de detección limitadas, costes y/o tiempo de respuesta elevados, requerimiento de equipo especializado y/o interpretación de un experto altamente entrenado. La presente invención resuelve los inconvenientes antes enumerados proporcionando un conjunto altamente sensible de sólo unos pocos marcadores de IMS homopoliméricos cortos del tipo descrito en WO2013153130, junto con un método extremadamente robusto para detectar inserciones o deleciones de homonucleótidos (indeles) en sus secuencias. Este método se adapta muy bien a la automatización, no requiere infraestructuras moleculares específicas y se puede llevar a cabo usando un equipo de laboratorio estándar, tal como un simple termociclador conectado a un ordenador. Además, permite una fácil duplexación o incluso una multiplexación de mayor nivel de los marcadores seleccionados, lo que confiere la ventaja

de limitar incluso más el material de laboratorio requerido y, por lo tanto, de facilitar la implementación en las plataformas basadas en PCR existentes. Es importante el hecho de que el método proporciona resultados muy consistentes y permite una interpretación fácil y totalmente automática con un informe directo como resultado. En los marcos actuales, desde la recepción de la muestra de tejido de un paciente, mostramos que podemos obtener dichas lecturas totales de los estados de los indeles en menos de 3 horas. Por lo tanto, el nuevo panel de marcadores aquí presentado y el método para su detección proporcionan una nueva alternativa altamente ventajosa para detectar IMS en muestras de CCR y de otros cánceres, tales como cáncer ovárico, endometrial y gástrico, así como para estudios predictivos y de seguimiento en el contexto de la inmunoterapia. A continuación, se presentan estas y otras ventajas y usos de la presente invención.

10 Compendio de la invención

La presente invención se define en las reivindicaciones independientes adjuntas. Se definen las realizaciones preferidas en las reivindicaciones dependientes. En particular, la presente invención se relaciona con

un método para analizar loci de IMS en una muestra biológica de cáncer colorrectal, endometrial o gástrico, comprendiendo el método la etapa de:

- 15 - determinar el número de nucleótidos en las siguientes repeticiones homopoliméricas mapeadas en el genoma de referencia humano GRCh38/hg38:

repetición homopolimérica que comprende 11 adeninas consecutivas localizadas en el gen DDO1 humano y que comienza en la posición chr20:62,905,340;

- 20 repetición homopolimérica que comprende 11 adeninas consecutivas localizadas en el gen MRE11 humano y que comienza en la posición chr11:94,479,765;

repetición homopolimérica que comprende 10 adeninas consecutivas localizadas en el gen SULF2 humano y que comienza en la posición chr20:47,657,577; y

repetición homopolimérica que comprende 8 adeninas consecutivas localizadas en el gen ACVR2A humano y que comienza en la posición chr2:147,926,117.

- 25 En relación con lo anterior, la presente invención también se relaciona con un kit para analizar loci de IMS en una muestra biológica, comprendiendo el kit herramientas para amplificar regiones de ácido nucleico que comprenden al menos las repeticiones homopoliméricas antes mencionadas.

Finalmente, pero no menos importante, la presente invención también se relaciona con una célula o cualquier otro material, en particular material genético, derivados de una estirpe celular HTC116 cl.110268743 que comprende una delección homonucleotídica en cada una de las repeticiones homopoliméricas antes mencionadas y en otras varias repeticiones homopoliméricas ventajosas. Se depositó la estirpe celular HTC116 cl.110268743 con éxito de acuerdo con el Tratado de Budapest en la autoridad depositaria BCCM/GeneCorner de Bélgica bajo el número de acceso LMBP 12278CB.

Breve descripción de las figuras

- 35 Para una mayor comprensión de la naturaleza de la presente invención, se hace referencia a la siguiente descripción detallada tomada conjuntamente con los dibujos adjuntos, en los cuales:

Figura 1: muestra el estado de IMS de 7 marcadores de microsatélites (BTBD7, RYR3, SEC31A, ACVR2A, DDO1, MRE11 y SULF2) en 128 muestras de cáncer colorrectal IMS-A. Diferentes paneles muestran el estado de IMS (blanco, EMS; gris oscuro, IMS-A) cuando se evalúan las muestras para el conjunto mínimo de A) 4 marcadores y B), C) y D) añadiendo secuencialmente un marcador al conjunto mínimo de 4 marcadores. Se indica el estado de los marcadores para muestras individuales como tipo salvaje (gris pálido) o mutante (gris);

y

- 45 Figura 2: muestra el estado de IMS de 7 marcadores de microsatélites (BTBD7, RYR3, SEC31A, ACVR2A, DDO1, MRE11 y SULF2) en 15 muestras de cáncer gástrico IMS-A y 19 muestras de cáncer de endometrio. Diferentes paneles muestran el estado de IMS (blanco, EMS; gris oscuro, IMS-A) cuando se evalúan las muestras para el conjunto mínimo de A) 4 marcadores y B), C) y D) añadiendo secuencialmente un marcador al conjunto mínimo de 4 marcadores. Se indica el estado de los marcadores para muestras individuales como tipo salvaje (gris pálido), mutante (gris) o sin resultados (franja diagonal).

Descripción detallada de la invención

La presente invención se relaciona, en general, con un nuevo panel de biomarcadores de IMS, con métodos que explotan este panel, con sistemas automatizados y con kits para realizar dichos métodos, mediante los cuales los kits pueden preferiblemente comprender o presentarse en forma de un cartucho compatible con dichos sistemas automatizados y comprenden herramientas y preferiblemente también material de control positivo para detectar indeles en dicho panel.

En una realización preferida, la invención proporciona un método para analizar el estado de IMS de un cáncer colorrectal, endometrial o gástrico, analizando los loci de IMS en una muestra biológica mediante el uso de un panel de biomarcadores que comprende al menos las siguientes regiones de repetición homopoliméricas, o formas mutadas de las mismas (mediante lo cual la mutación es la presencia de al menos un indel en la secuencia de repetición homopolimérica), mapeadas en el genoma de referencia humano GRCh38/hg38:

repetición homopolimérica que comprende 11 adeninas consecutivas localizadas en el gen DDO1 humano y que comienza en la posición chr20:62,905,340;

repetición homopolimérica que comprende 11 adeninas consecutivas localizadas en el gen MRE11 humano y que comienza en la posición chr11:94,479,765;

repetición homopolimérica que comprende 10 adeninas consecutivas localizadas en el gen SULF2 humano y que comienza en la posición chr20:47,657,577; y

repetición homopolimérica que comprende 8 adeninas consecutivas localizadas en el gen ACVR2A humano y que comienza en la posición chr2:147,926,117.

Estudiamos muchos marcadores aleatoriamente seleccionados del tipo descrito en WO2013153130 (VIB), que son homopolímeros marcadamente más cortos que los del panel Bethesda. Selecciones aleatorias de un menor número de marcadores del conjunto preferido allí descrito de 56 marcadores no dio un ensayo robusto capaz de detectar repetidamente el fenotipo IMS-A en una gran gama de muestras de cáncer humano con los recursos de laboratorio mínimos y más básicos. Los marcadores, o bien no se podían detectar en una reacción múltiplex o incluso dúplex, o bien variaban en cuanto a número de nucleótidos entre diferentes etnias humanas. La elección fortuita de una repetición homopolimérica exónica que comprende 8 adeninas consecutivas localizadas en el gen ACVR2A humano y que comienza en la posición chr2:147,926,117, que no se describe en el conjunto preferido de 56 en WO2013153130, y la repetición homopolimérica exónica que comprende 11 adeninas consecutivas localizadas en el gen MRE11 humano y que comienza en la posición chr11:94,479,765, que no se describe en WO2013153130 en absoluto, condujo sorprendentemente a la obtención de un panel de gran rendimiento para la detección de IMS en tumores humanos deficientes en RDA.

Observamos que el rendimiento del panel mínimo aquí presentado de sólo 4 marcadores sobrepasa el rendimiento aceptable mínimo establecido en el 95% para la correcta identificación de muestras de cáncer colorrectal humanas con IMS-A verificada. Como se muestra en la sección de ejemplos más adelante, mostramos que el anterior panel mínimo de la invención podía recuperar con éxito 123 muestras positivas a IMS-A del conjunto de 128 muestras de CCR IMS-A, lo que representa que un 96% de las muestras se identifican correctamente como IMS-A. Por lo tanto, en una realización preferida, se proporciona un panel que tiene el rendimiento de identificar correctamente al menos un 90%, preferiblemente al menos un 95%, de muestras tumorales IMS-A.

Lógicamente, la adición de más marcadores puede aumentar gradualmente este rendimiento. Por lo tanto, en una realización preferida, la presente invención proporciona el método según la realización previa, que además comprende una cualquiera, dos o todas de las siguientes tres regiones de repetición homopolimérica, o formas mutantes de las mismas, mapeadas en el genoma de referencia humano GRCh38/hg38:

repetición homopolimérica que comprende 10 adeninas consecutivas localizadas en el gen BTBD7 humano y que comienza en la posición chr14:93,241,685;

repetición homopolimérica que comprende 9 timinas consecutivas localizadas en el gen SEC31A humano y que comienza en la posición chr4:82,864,412;

repetición homopolimérica que comprende 10 adeninas consecutivas localizadas en el gen RYR3 humano y que comienza en la posición chr15:33,865,341.

Los marcadores de repeticiones homopoliméricas aquí presentados que constituyen el panel de biomarcadores de la invención son simplemente hebras cortas de a lo sumo 11 homonucleótidos repetitivos, por ej., 11 adeninas consecutivas en DDO1 3'UTR. Como apreciará cualquier persona experta en la técnica, también se construirían secuencias complementarias a éstas, por ej., 11 timinas consecutivas complementarias a las 11 adeninas consecutivas en la secuencia DDO1 3'UTR como dentro del alcance de los términos anteriores aquí utilizados.

En una realización particularmente preferida, se proporciona el panel para uso en el método de la invención que comprende las cinco siguientes regiones de repeticiones homopoliméricas, o formas mutadas de las mismas mapeadas en el genoma de referencia humano GRCh38/hg38:

- 5 repetición homopolimérica que comprende 11 adeninas consecutivas localizadas en el gen DIDO1 humano y que comienza en la posición chr20:62,905,340;
- repetición homopolimérica que comprende 11 adeninas consecutivas localizadas en el gen MRE11 humano y que comienza en la posición chr11:94,479,765;
- repetición homopolimérica que comprende 10 adeninas consecutivas localizadas en el gen SULF2 humano y que comienza en la posición chr20:47,657,577;
- 10 repetición homopolimérica que comprende 8 adeninas consecutivas localizadas en el gen ACVR2A humano y que comienza en la posición chr2:147,926,117; y
- repetición homopolimérica que comprende 10 adeninas consecutivas localizadas en el gen BTBD7 humano y que comienza en la posición chr14:93,241,685.
- 15 Usando este conjunto nuclear de cinco marcadores, se pudieron identificar 124 de 128, es decir, un 97% de las muestras tumorales como IMS-A.

En otra realización preferida, se proporciona el panel que comprende las seis siguientes regiones de repeticiones homopoliméricas, o formas mutadas de las mismas, mapeadas en el genoma de referencia humano GRCh38/hg38:

- repetición homopolimérica que comprende 11 adeninas consecutivas localizadas en el gen DIDO1 humano y que comienza en la posición chr20:62,905,340;
- 20 repetición homopolimérica que comprende 11 adeninas consecutivas localizadas en el gen MRE11 humano y que comienza en la posición chr11:94,479,765;
- repetición homopolimérica que comprende 10 adeninas consecutivas localizadas en el gen SULF2 humano y que comienza en la posición chr20:47,657,577;
- 25 repetición homopolimérica que comprende 8 adeninas consecutivas localizadas en el gen ACVR2A humano y que comienza en la posición chr2:147,926,117;
- repetición homopolimérica que comprende 10 adeninas consecutivas localizadas en el gen BTBD7 humano y que comienza en la posición chr14:93,241,685; y
- repetición homopolimérica que comprende 9 timinas consecutivas localizadas en el gen SEC31A humano y que comienza en la posición chr4:82,864,412.
- 30 Añadiendo además el marcador adicional localizado en el gen SEC31A al conjunto nuclear de marcadores, se pudieron puntuar la totalidad de las 128 muestras como IMS-A, lo que hace que el panel sea incluso más eficiente en la definición del estado de IMS.

En aún otra realización preferida, se proporciona el panel que comprende las siete siguientes regiones de repeticiones homopoliméricas, o formas mutadas de las mismas, mapeadas en el genoma de referencia humano GRCh38/hg38:

- 35 repetición homopolimérica que comprende 11 adeninas consecutivas localizadas en el gen DIDO1 humano y que comienza en la posición chr20:62,905,340;
- repetición homopolimérica que comprende 11 adeninas consecutivas localizadas en el gen MRE11 humano y que comienza en la posición chr11:94,479,765;
- 40 repetición homopolimérica que comprende 10 adeninas consecutivas localizadas en el gen SULF2 humano y que comienza en la posición chr20:47,657,577;
- repetición homopolimérica que comprende 8 adeninas consecutivas localizadas en el gen ACVR2A humano y que comienza en la posición chr2:147,926,117;
- repetición homopolimérica que comprende 10 adeninas consecutivas localizadas en el gen BTBD7 humano y que comienza en la posición chr14:93,241,685;
- 45 repetición homopolimérica que comprende 9 timinas consecutivas localizadas en el gen SEC31A humano y que comienza en la posición chr4:82,864,412; y
- repetición homopolimérica que comprende 10 adeninas consecutivas localizadas en el gen RYR3 humano y que comienza en la posición chr15:33,865,341.

Se calcula que un conjunto de los anteriores siete marcadores reduce el índice de falsos negativos potenciales a ~ 1/1900. Por lo tanto, la adición de aún otro marcador compatible con los paneles de las realizaciones previas proporciona una mayor salvaguarda de rendimiento, en particular para implementar el panel en la detección de la IMS en cánceres distintos del colorrectal, tales como ovárico, endometrial o gástrico.

- 5 En una realización preferida, la muestra biológica se origina de un individuo sospechoso de tener un tumor. En otra realización, la muestra biológica es una muestra tumoral, posiblemente una muestra de tejido fresco o de tumor fijado, por ej., una muestra congelada o FFPE. En una realización particularmente preferida, se selecciona el tumor de un tumor colorrectal, ovárico, endometrial o gástrico. En otra realización posible, la muestra es una muestra de biopsia líquida. En otra realización posible, la muestra es cualquier muestra de tejido, tal como células mononucleares de sangre periférica (PBMNC) u otras células blancas de la sangre, o tejido cutáneo de un paciente sospechoso de sufrir el síndrome de Lynch.

Es otro objeto de la invención proporcionar un método de análisis del estado de IMS de un cáncer colorrectal, endometrial o gástrico analizando los loci de IMS en una muestra biológica, comprendiendo el método la etapa de determinar el número de nucleótidos en el panel de biomarcadores de las realizaciones antes descritas.

- 15 Por lo tanto, en una realización de la invención, se proporciona un método para analizar los loci de IMS en una muestra biológica que comprende la etapa de:

- determinar el número de nucleótidos en las siguientes repeticiones homopoliméricas mapeadas en el genoma de referencia humano GRCh38/hg38:

- 20 repetición homopolimérica que comprende 11 adeninas consecutivas localizadas en el gen DDO1 humano y que comienza en la posición chr20:62,905,340;

repetición homopolimérica que comprende 11 adeninas consecutivas localizadas en el gen MRE11 humano y que comienza en la posición chr11:94,479,765;

repetición homopolimérica que comprende 10 adeninas consecutivas localizadas en el gen SULF2 humano y que comienza en la posición chr20:47,657,577; y

- 25 repetición homopolimérica que comprende 8 adeninas consecutivas localizadas en el gen ACVR2A humano y que comienza en la posición chr2:147,926,117.

Por las mismas razones que las descritas anteriormente, en una realización preferida, el método de la invención comprende además determinar el número de nucleótidos también en una cualquiera, dos o la totalidad de las siguientes regiones de repeticiones homopoliméricas mapeadas en el genoma de referencia humano GRCh38/hg38:

- 30 repetición homopolimérica que comprende 10 adeninas consecutivas localizadas en el gen BTBD7 humano y que comienza en la posición chr14:93,241,685;

repetición homopolimérica que comprende 9 timinas consecutivas localizadas en el gen SEC31A humano y que comienza en la posición chr4:82,864,412;

- 35 repetición homopolimérica que comprende 10 adeninas consecutivas localizadas en el gen RYR3 humano y que comienza en la posición chr15:33,865,341.

En una realización particular, se proporciona el método que comprende la etapa de:

- determinar el número de nucleótidos en las cinco siguientes repeticiones homopoliméricas mapeadas en el genoma de referencia humano GRCh38/hg38:

- 40 repetición homopolimérica que comprende 11 adeninas consecutivas localizadas en el gen DDO1 humano y que comienza en la posición chr20:62,905,340;

repetición homopolimérica que comprende 11 adeninas consecutivas localizadas en el gen MRE11 humano y que comienza en la posición chr11:94,479,765;

repetición homopolimérica que comprende 10 adeninas consecutivas localizadas en el gen SULF2 humano y que comienza en la posición chr20:47,657,577;

- 45 repetición homopolimérica que comprende 8 adeninas consecutivas localizadas en el gen ACVR2A humano y que comienza en la posición chr2:147,926,117; y

repetición homopolimérica que comprende 10 adeninas consecutivas localizadas en el gen BTBD7 humano y que comienza en la posición chr14:93,241,685.

En una realización más específica, se proporciona el método que comprende la etapa de:

- determinar el número de nucleótidos en las seis siguientes repeticiones homopoliméricas mapeadas en el genoma de referencia humano GRCh38/hg38:

repetición homopolimérica que comprende 11 adeninas consecutivas localizadas en el gen DIDO1 humano y que comienza en la posición chr20:62,905,340;

5 repetición homopolimérica que comprende 11 adeninas consecutivas localizadas en el gen MRE11 humano y que comienza en la posición chr11:94,479,765;

repetición homopolimérica que comprende 10 adeninas consecutivas localizadas en el gen SULF2 humano y que comienza en la posición chr20:47,657,577;

10 repetición homopolimérica que comprende 8 adeninas consecutivas localizadas en el gen ACVR2A humano y que comienza en la posición chr2:147,926,117;

repetición homopolimérica que comprende 10 adeninas consecutivas localizadas en el gen BTBD7 humano y que comienza en la posición chr14:93,241,685; y

repetición homopolimérica que comprende 9 timinas consecutivas localizadas en el gen SEC31A humano y que comienza en la posición chr4:82,864,412.

15 En otra realización específica, se proporciona el método que comprende la etapa de:

- determinar el número de nucleótidos en las siete siguientes repeticiones homopoliméricas mapeadas en el genoma de referencia humano GRCh38/hg38:

repetición homopolimérica que comprende 11 adeninas consecutivas localizadas en el gen DIDO1 humano y que comienza en la posición chr20:62,905,340;

20 repetición homopolimérica que comprende 11 adeninas consecutivas localizadas en el gen MRE11 humano y que comienza en la posición chr11:94,479,765;

repetición homopolimérica que comprende 10 adeninas consecutivas localizadas en el gen SULF2 humano y que comienza en la posición chr20:47,657,577;

25 repetición homopolimérica que comprende 8 adeninas consecutivas localizadas en el gen ACVR2A humano y que comienza en la posición chr2:147,926,117;

repetición homopolimérica que comprende 10 adeninas consecutivas localizadas en el gen BTBD7 humano y que comienza en la posición chr14:93,241,685;

repetición homopolimérica que comprende 9 timinas consecutivas localizadas en el gen SEC31A humano y que comienza en la posición chr4:82,864,412; y

30 repetición homopolimérica que comprende 10 adeninas consecutivas localizadas en el gen RYR3 humano y que comienza en la posición chr15:33,865,341.

En posibles realizaciones, el método de la invención puede además comprender una etapa de diagnóstico del estado de IMS de la muestra biológica si se detecta un indel en al menos dos de las repeticiones homopoliméricas.

35 Preferiblemente, se proporcionan los métodos de la invención en donde la muestra biológica obtenida de un sujeto es un tumor colorrectal, gástrico, ovárico o endometrial o un tumor colorrectal, gástrico, ovárico o endometrial potencial.

Como apreciarán los expertos en la técnica, la naturaleza del panel de marcadores de repeticiones homopoliméricas de la invención determina que los métodos de la invención se llevarán a cabo preferiblemente usando el ADN genómico presente en las muestras biológicas. Dependiendo del tipo de muestra, en una realización preferida, el método de la invención va precedido por cualquiera de las etapas siguientes:

40 - liberar y/o aislar el ácido nucleico que comprende potencialmente la secuencia diana de la fuente de un ácido nucleico,

- proporcionar dicho ácido nucleico liberado y/o purificado que comprende potencialmente la diana a la etapa de amplificación de dicho ácido nucleico.

45 Como el ADN genómico es un material de ácido nucleico rico y complejo, es ventajoso que se amplifiquen las secuencias que flanquean las regiones de repeticiones homopoliméricas como se ha definido anteriormente antes de la etapa de determinación del número de nucleótidos en ellas. Por lo tanto, en una realización preferida, se proporciona el método que además comprende la etapa de:

- amplificar regiones de ácido nucleico que comprenden las repeticiones homopoliméricas antes enumeradas. Como resultará evidente para cualquier experto, dicha amplificación dará como resultado productos de amplificación que comprenden la secuencia de repeticiones homopoliméricas independientemente de su estado de IMS. Es decir, que dichos productos de amplificación pueden comprender la versión de tipo salvaje (WT) de una repetición homopolimérica dada o su variante IMS, es decir, un mutante que comprende un indel de al menos un homonucleótido en la secuencia de repeticiones homopoliméricas.

Naturalmente, en una realización obvia, se realiza preferiblemente la amplificación por reacción en cadena de polimerasa (PCR), por ej., usando un medio para realizar PCR, tal como reactivos y/o engranaje apropiados, incluyendo un termociclador. Sin embargo, también se pueden usar otras técnicas de amplificación conocidas en este campo. Éstas incluyen, aunque sin limitación, Amplificación Isotérmica Mediada por Bucle (LAMP), Amplificación Basada en la Secuencia de Ácidos Nucleicos (NASBA), Amplificación con Desplazamiento de Hebra (SDA), Amplificación con Desplazamiento Múltiple (MDA), Amplificación de Círculo Rodante (RCA), Reacción en Cadena de Ligasa (LCR), Amplificación Dependiente de Helicasa (HDA) o Método de Amplificación con Ramificación (RAM).

En una realización preferida, se proporciona el método en donde la etapa de amplificación comprende el uso de al menos un cebador que tiene la secuencia identificada por cualquiera de las siguientes SEQ ID NO.:1-14:

Para el marcador DIDO1:

SEQ ID NO.:1 - TAGCGTGTGAATCGGACAT

SEQ ID NO.:2 - TTGACTGGGCAGATAGGGGA

Para el marcador MRE11:

SEQ ID NO.:3 - ATAGTTCACCCATGGAAACC

SEQ ID NO.:4 - GGAGGAGAATCTTAGGGAAA

Para el marcador BTBD7:

SEQ ID NO.:5 - ACTGGACTCCCGCTGG

SEQ ID NO.:6 - CGCTCAGCCTCCATAAATC

Para el marcador SULF2:

SEQ ID NO.:7 - CAACTTCATTTCTTTTCAGTACCTT

SEQ ID NO.:8 - CTGTCCAGATACCATTTCTC

Para el marcador ACVR2A:

SEQ ID NO.:9 - AGCATCCATCTCTTGAAGACAT

SEQ ID NO.:10 - GCATGTTTCTGCCAATAATCTCT

Para el marcador SEC31A:

SEQ ID NO.:11 - CAACTTCAGCAGGCTGT

SEQ ID NO.:12 - AGTCTGAGAAGCATCAATTTT

Para el marcador RYR3:

SEQ ID NO.:13 - CATTTTCTAAATGCCTCCCTTAAA

SEQ ID NO.:14 - GTCCATTAGGCACAAAAAG

En una realización más específica, la etapa de amplificación comprende el uso de al menos un par de cebadores seleccionado de los siguientes: SEQ ID NO.:1 y SEQ ID NO.:2; SEQ ID NO.:3 y SEQ ID NO.:4; SEQ ID NO.:5 y SEQ ID NO.:6; SEQ ID NO.:7 y SEQ ID NO.:8; SEQ ID NO.:9 y SEQ ID NO.:10; SEQ ID NO.:11 y SEQ ID NO.:12; o SEQ ID NO.:13 y SEQ ID NO.:14.

Como apreciará cualquier experto, dependiendo de las condiciones de amplificación, las secuencias de cebadores antes enumeradas probablemente también funcionarán en caso de que se alteren en ellas 1, 2 o, en algunos casos, incluso posiblemente 3, nucleótidos, es decir, que se añadan, eliminen o sustituyan por un nucleótido diferente o un nucleótido modificado. Por lo tanto, en una realización posible, la presente invención también proporciona al menos una secuencia de cebadores identificada por cualquiera de las SEQ ID NO.:1-14 anteriores, en donde se alteran 1, 2

ó 3 nucleótidos. En una realización alternativa, la presente invención también proporciona al menos una secuencia de cebadores que es al menos un 80%, preferiblemente al menos un 85%, más preferiblemente al menos un 90% o lo más preferiblemente al menos un 95%, idéntica a cualquiera de las SEQ ID NO.:1-14 anteriores. Como apreciarán los expertos, con objeto de generar amplicones que cubran las regiones de repeticiones homopoliméricas de interés, se pueden diseñar cebadores alternativos 5, 10, 20, 50 ó 100 nucleótidos secuencia arriba o secuencia abajo con respecto a las posiciones de los pares de cebadores antes descritos. Por lo tanto, también habría que considerar a dichos pares de cebadores alternativos como una realización obvia alternativa de la presente invención.

El método aquí proporcionado tiene la ventaja de ser totalmente automatizable y adaptable a cualquier instrumento termociclador de PCR cuantitativa estándar, lo que permite que lo realice un personal regular de laboratorio sin necesidad de entrenamiento especializado. Además de lo anterior, el método es altamente sensible, adecuado para multiplexación, puede proporcionar una estimación de las cantidades relativas de las secuencias de repeticiones nucleotídicas homopoliméricas detectadas y sus variantes. Por lo tanto, en una realización posible, la PCR puede ser una PCR cuantitativa o semicuantitativa.

Como los métodos de la invención se relacionan con la detección de cambios en el número de homonucleótidos en una hebra de secuencias de repetición de homonucleótidos muy corta (es decir, < 12 nt), es ventajoso que sean altamente específicos. Por ejemplo, durante la amplificación de regiones de repeticiones homopoliméricas, se sabe que se produce deslizamiento de polimerasa. Esto conduce a errores al copiar el número original de nucleótidos repetidos, lo que causa la acumulación de deleciones o inserciones artificiales en el producto de PCR amplificado. Por lo tanto, en una realización preferida, se lleva a cabo la etapa de amplificación usando una polimerasa correctora, es decir, una polimerasa que tiene actividad 3'-5' exonucleasa. Se conocen muchas de tales polimerasas de grado de PCR y se dispone de ellas comercialmente. Como ejemplos se incluyen, aunque sin limitación, polimerasas como Q5, Pfx, Pfu, Ex Taq, etc.

En las realizaciones más preferidas, se usa el análisis de curva de fusión de los productos de ácidos nucleicos amplificados en la etapa de determinación del número de nucleótidos. Por lo tanto, en una realización particularmente ventajosa de los métodos de la invención, la etapa de amplificación da lugar a la generación de datos de curva de fusión.

El análisis de la curva de fusión es una valoración de las características de disociación o asociación de una molécula de ácido nucleico de doble hebra durante la variación de temperatura. Por consiguiente, se han de entender los datos de la curva de fusión como cualesquiera datos capturados que representan las características bien de disociación, bien de asociación de la molécula de ácido nucleico bajo investigación, tal como el producto diana de una amplificación de ácido nucleico. Se pueden obtener los datos de la curva de fusión incluyendo restos fluorescentes apropiados en las muestras bajo investigación que se procesan mediante cualquier instrumento o un método para llevar a cabo la amplificación, tal como ciclación térmica, PCR, PCR cuantitativa, etc. Se puede obtener a partir de cualquier aparato equipado con un medio de ajuste de la temperatura de la muestra por encima de la temperatura de fusión de la muestra de ADN, que esté equipado con medios fluorométricos o espectrofotométricos conocidos. Como ejemplos de dichos instrumentos, se incluyen, aunque sin limitación, termocicladores ópticos regulares comúnmente usados para PCR o fluorómetros con control de temperatura, etc.

El análisis de la curva de fusión y el análisis de fusión de alta resolución (HRM) son métodos comúnmente utilizados para detectar y analizar la presencia de secuencias de ácidos nucleicos en una muestra. Una forma de monitorizar las características de disociación y asociación de un ácido nucleico es con ayuda de tintes. Las químicas de detección usadas para PCR y el análisis de la curva de fusión se basan en (a) químicas que normalmente detectan fluorescencia de un tinte que se une a la diana, por ej., un fluoróforo de unión a ADN, tal como LC Green, LC Green+, Eva Green, SYTO9 CYBR Green, o (b) químicas específicas de diana que normalmente utilizan sondas de ADN marcadas con fluoróforos, tales como, por ej., sondas baliza, y/o cebadores, tales como, por ej., cebadores de escorpión. Es bien sabido en la técnica que se pueden aplicar otras químicas de detección en el análisis de la curva de fusión.

En una realización de la invención, se calientan los productos de amplificación en presencia de uno o más tintes intercalantes durante un procedimiento de ensayo de la curva de fusión. La disociación del ADN durante el calentamiento es medible por la gran reducción de fluorescencia que se produce. En otra realización particular, se calientan los productos de amplificación en presencia de uno o más ácidos nucleicos marcados con tinte, por ej., una o más sondas, durante un procedimiento de ensayo de la curva de fusión. En el caso del análisis de curva de fusión de fluorescencia basado en sondas, la detección de variación en ácidos nucleicos se basa en la temperatura de fusión generada por desnaturalización térmica del híbrido sonda-diana. A medida que progresa el calentamiento de los amplicones generados, se detectan los cambios en la fuerza de la señal en función de la temperatura, típicamente en un intervalo de temperatura, para obtener datos brutos de la curva de fusión.

En realizaciones preferidas de los métodos de la invención, la amplificación comprende el uso de una sonda. En principio, en posibles realizaciones, se puede usar cualquier sonda oligonucleotídica específica de diana adecuada para realizar el análisis de la curva de fusión. Las sondas conocidas preferidas pueden comprender un par consistente en un fluoróforo y un desactivador, y también pueden ventajosamente formar estructuras secundarias, tales como bucles u horquillas.

Son particularmente preferidas las sondas de baliza molecular, o balizas moleculares, que son moléculas en forma de horquilla con un fluoróforo internamente desactivado cuya fluorescencia se restaura cuando se unen a una secuencia de ácido nucleico diana. Por esta razón, las balizas moleculares no se degradan por la acción de la polimerasa y se pueden emplear en el estudio de su cinética de hibridación a su diana mediante interpretación de la curva de fusión.

Por lo tanto, en una realización particularmente preferida, se proporciona el método en donde la etapa de amplificación comprende una utilización de al menos una sonda de baliza molecular.

En una realización preferida de la anterior realización, la sonda de baliza molecular comprende una secuencia idéntica a, o complementaria de, la secuencia de repetición nucleotídica homopolimérica mutante que comprende una delección de al menos un homonucleótido en la secuencia de repetición nucleotídica homopolimérica diana. Dicho diseño de balizas moleculares permite detectar con gran sensibilidad y especificidad el marcador de IMS mutado seleccionado, mientras que al mismo tiempo sigue siendo suficientemente sensible al marcador de tipo salvaje (es decir, esperado). Habría que señalar que, con el término "secuencia de repetición nucleotídica homopolimérica diana" se quiere decir la secuencia de repetición homopolimérica de tipo salvaje o de referencia, ya que se espera en las condiciones en que no hay presencia de IMS. Por el contrario, por "secuencia de repetición nucleotídica homopolimérica mutante", se quiere decir una secuencia de repetición nucleotídica homopolimérica que comprende una inserción o una delección de al menos un homonucleótido en la secuencia de repetición homopolimérica. Se medirá entonces la varianza entre los datos de fusión brutos del tipo salvaje y del mutante y será característica de los datos brutos de la curva de fusión.

En una realización específica, se proporciona el método en donde la al menos una sonda de baliza molecular tiene una secuencia identificada por una cualquiera de las siguientes SEQ ID NO.:

Para el marcador DIDO1:	SEQ ID NO.:15	<u>CGCACGACATGGAAAAAAAAAATC</u> <u>CGTGCGTAAA</u>
Para el marcador MRE11:	SEQ ID NO.:16	<u>CGTCGAACCTTAAAAAAAAAAGTTAC</u> <u>CGACGAA</u>
Para el marcador BTBD7:	SEQ ID NO.:17	<u>CGCACGACTTATTA</u> <u>AAAAAAAAAATGACAGTGCGTAAA</u>
Para el marcador SULF2:	SEQ ID NO.:18	<u>CGTCGGTACCTT</u> <u>AAAAAAAAAACATCACGACGAA</u>
Para el marcador ACVR2A:	SEQ ID NO.:19	<u>GTGCATA</u> <u>AAAAAAGAGCAC</u> <u>TAAA</u>
Para el marcador SEC31A:	SEQ ID NO.:20	<u>CGCACTTGCC</u> <u>AAAAAAAAAATTGATGGTGCGTAAA</u>
Para el marcador RYR3:	SEQ ID NO.:21	<u>CGTCGCCCTT</u> <u>AAAAAAAAAACTGC</u> <u>CGACGAA</u>

En una realización posible, se proporciona la al menos una sonda de baliza molecular que tiene un cierto grado de variación de secuencia con respecto a las anteriores SEQ ID NO. 15-21. Dicha variación podría justificar el uso de una secuencia de diferente tallo de baliza (subrayada y en itálica anteriormente), o deberse a la eliminación o adición de nucleótidos a la parte hibridante de la baliza que es específica para la secuencia que se ha de detectar (indicada en negrita anteriormente). Esto último podría incluir la adición o eliminación en 1 ó 2 nucleótidos de las secuencias de repeticiones homopoliméricas, o la inclusión de más o menos nucleótidos en dicha secuencia flanqueante de la repetición.

Debido a la especificidad así conferida de una sonda de baliza molecular dada hacia un marcador de repetición homopolimérica y a sus variantes inestables (mutantes), es también posible diseñar un ensayo de multiplexación, en donde se usan al menos dos, posiblemente más, sondas de baliza molecular en un tubo o compartimento de reacción.

Por lo tanto, en otra realización preferida, se proporciona el método en donde la etapa de amplificación comprende al menos una amplificación dúplex de un par de repeticiones homopoliméricas, siendo dicho par seleccionado de las siguientes combinaciones:

- amplificación dúplex de

la repetición homopolimérica que comprende 11 adeninas consecutivas localizadas en el gen DIDO1 humano y que comienza en la posición chr20:62,905,340

junto con

la repetición homopolimérica que comprende 11 adeninas consecutivas localizadas en el gen MRE11 humano y que comienza en la posición chr11:94,479,765;

- amplificación dúplex de

la repetición homopolimérica que comprende 8 adeninas consecutivas localizadas en el gen ACVR2A humano y que comienza en la posición chr2:147,926,117;

junto con

- 5 la repetición homopolimérica que comprende 9 timinas consecutivas localizadas en el gen SEC31A humano y que comienza en la posición chr4:82,864,412; y

- amplificación dúplex de

la repetición homopolimérica que comprende 10 adeninas consecutivas localizadas en el gen BTBD7 humano y que comienza en la posición chr14:93,241,685;

- 10 junto con

la repetición homopolimérica que comprende 10 adeninas consecutivas localizadas en el gen SULF2 humano y que comienza en la posición chr20:47,657,577.

En realizaciones particularmente ventajosas que mejoran mucho la robustez de los métodos de la invención, en particular cuando se usa multiplexación, se usa un nuevo enfoque en donde se aplica una función de transformación de ondícula a los datos brutos de la curva de fusión.

- 15 Las ondículas son funciones matemáticas que cortan los datos en componentes de diferente frecuencia, y entonces estudian cada componente con una resolución equiparada a su escala. Estas funciones de base son ondas cortas con duración limitada. Se llevan a escala las funciones de base de la transformación de ondículas con respecto a la frecuencia. Hay muchas ondículas diferentes que se pueden usar como funciones de base. La función de base $\psi(t)$, también llamada la ondícula madre, es la función transformante. El término madre implica que las funciones con diferente región de soporte que se usan en el proceso de transformación derivan de una función principal, o la ondícula madre. En otras palabras, la ondícula madre es un prototipo para generar las otras funciones de ventana. En general, la ondícula $\psi(t)$ es una función valiosa compleja. Una función de ondícula general se define como:

$$\psi_{s,\tau}(t) = |s|^{-1/2} \psi[(t-\tau)/s]$$

- 25 Este parámetro de desplazamiento ' τ ' determina la posición de la ventana en el tiempo y, por lo tanto, define qué parte de la señal $x(t)$ se está analizando. En el análisis de la transformación de ondículas, se sustituye la variable de frecuencia ' ω ' por la variable de escala ' s ' y la variable de desplazamiento temporal se representa por ' τ '.

La transformación de ondículas utiliza estas funciones de ondícula madre y realiza la descomposición de la señal $x(t)$ en un conjunto ponderado de funciones de ondícula llevadas a escala $\psi(t)$. La principal ventaja de usar ondículas es que permiten capturar una firma característica y única de un conjunto de datos dado mayor y más complejo sin que se produzca la pérdida de datos.

- 30 Por ejemplo, dos grandes conjuntos de datos brutos de curva de fusión obtenidos de dos productos de amplificación que difieren sólo en un nucleótido de longitud son altamente similares, pero, después de aplicar una transformación de ondícula, producirán dos firmas distintas. Dichas firmas serán entonces más fáciles de comparar entre sí para concluir consistentemente que hubo una inserción o delección presente en uno de los productos de amplificación. En conclusión, la aplicación de la función de ondícula da como resultado reducción del ruido y aumento en la eficacia y velocidad computacional cuando se manejan conjuntos de datos grandes y similares. Como consecuencia, los datos procesados con ondículas son particularmente adecuados para clasificar muestras que conllevan el análisis combinado de varias dianas multiplexadas en un experimento, especialmente cuando se generan grandes conjuntos de datos brutos que requieren discriminación de minúsculas variaciones de datos.

Los métodos actualmente existentes para la detección de IMS tienen los siguientes inconvenientes: (a) para determinar la longitud de repetición, requieren equipo especializado adicional para realizar el análisis post-PCR y/o este análisis típicamente necesita ser interpretado por un experto altamente entrenado; o (b) en caso de curva de fusión de alta resolución con tintes de intercalación en el ADNdh, el inconveniente es la muy limitada capacidad de multiplexación con objeto de evitar señales de fusión solapantes de diferentes amplicones, y, además, no proporciona capacidad para cuantificar las cantidades relativas de secuencias inestables (mutantes) con respecto a las estables (tipo salvaje). Observamos que la aplicación de una transformación de ondículas discreta a los datos de la curva de fusión lleva a una interpretación muy robusta y consistente de los resultados de un modo totalmente automatizado y, por lo tanto, resuelve estos inconvenientes.

- 50 Por lo tanto, en una realización preferida, el método de la invención comprende además las etapas de:

(a) aplicar transformación de ondículas a los datos de la curva de fusión, y

(b) usar los resultados obtenidos de (a) en la determinación del número de nucleótidos en cualquiera de las repeticiones homopoliméricas anteriormente enumeradas. En otras palabras, en una realización, se proporcionan métodos que aplican funciones de transformación de ondícula para analizar los datos de la curva de fusión de ácidos nucleicos de una muestra de ensayo para determinar una presencia o una ausencia de un indel en cada una de las repeticiones homopoliméricas del panel de biomarcadores de elección, pudiendo usarse entonces esta información para clasificar dicha muestra de ensayo como poseedora o no de IMS.

En una realización preferida, los datos de la curva de fusión son datos brutos de la curva de fusión, es decir, datos que representan la métrica bruta de una señal obtenida de un experimento de disociación o asociación de ácidos nucleicos. En otras palabras, dichas métricas brutas no se procesan matemáticamente, por ej., aplicando análisis de curva de fusión de primera o segunda derivada, como frecuentemente se hace en la técnica, sino que, tras su recogida por un detector, se envían a un ordenador en donde se les aplica la función de transformación de ondícula.

En una más preferida realización, la transformación de ondícula es una transformación de ondícula discreta o "TOD". La TOD es cualquier transformación de ondícula para la cual se muestrean discretamente las ondículas. Como con otras transformaciones de ondícula, una ventaja clave que tiene sobre las transformaciones de Fourier es la resolución temporal: captura información tanto de frecuencia como de localización (localización en el tiempo). La aplicación de la transformación de ondícula discreta a la métrica bruta produce un conjunto de coeficientes de ondículas de salida de reconstrucción a diferentes escalas: (a) uno es la salida de aproximación, que es el contenido de baja frecuencia del componente de la señal de entrada, y (b) el otro es la salida multidimensional, que da los componentes de alta frecuencia, que son los detalles de la señal de entrada a diversos niveles. Se hace además referencia a estos coeficientes como coeficientes de transformación de ondícula discreta o coeficientes tod. La separación de características en diferentes escalas (o frecuencias) permite que un operador o un algoritmo de ordenador seleccione los coeficientes tod más relevantes para ciertas decisiones o análisis, un proceso al que con frecuencia se hace referencia como filtración de ondículas. Este proceso puede aplicarse repetidamente, dividiendo la señal en múltiples bandas de frecuencia. Cuando se aplican a los datos de la curva de fusión, los coeficientes de ondículas de mayor frecuencia son mayormente ruido, mientras que los coeficientes de menor resolución capturan información relacionada con la ganancia del instrumento o la eficacia de la amplificación en la reacción de amplificación anterior. Ambos tienen poca o ninguna relevancia para la identificación de un oligonucleótido específico en una muestra sometida al propio análisis de curva de fusión, pero potencialmente tienen relevancia con respecto a la fiabilidad de dicha identificación. Se han descrito paquetes que contienen todas las funciones necesarias para computar y representar TOD (Aldrich, 2015) y serán conocidos para los programadores y matemáticos expertos.

En una realización preferida del método, la etapa de realización de una transformación de ondícula discreta sobre los datos de la curva de fusión para producir coeficientes tod calculará en un marco particular una transformación de ondícula de una dimensión (1D) de los datos brutos o los datos de reducción usando una ondícula madre de la familia Daubechies. La ondícula madre es la ondícula no modificada escogida como base para la transformación de ondícula discreta (Daubechies, 1992). Se obtuvieron buenos resultados cuando se usó la ondícula madre DB8. Se dilata, desplaza y lleva a escala a continuación la ondícula madre, usando el algoritmo tod en pirámide, para generar un conjunto de ondículas hijas que mejor representan la señal de curva de fusión de fluorescencia que se ha de analizar, siendo el conjunto de coeficientes de ondículas y escalas obtenidos del algoritmo el resultado de la transformación de ondícula discreta. En el ejemplo especificado, las condiciones de los límites para la TOD son periódicas. La entrada de datos brutos a la transformación puede ser la totalidad de los datos medidos o un subconjunto de los mismos que cubre todos los sucesos significativos de un experimento dado.

En línea con lo anterior, para producir coeficientes tod, los métodos de la presente invención pueden aplicar la transformación de ondícula discreta a los datos brutos de la curva de fusión o a unos datos de curva de fusión matemáticamente transformados o reducidos, es decir, sólo a una selección de datos brutos.

Además, no todos los coeficientes tod necesitan ser siempre utilizados para la determinación final del número de nucleótidos. Para aumentar la velocidad computacional, puede bastar sólo una selección de coeficientes tod. Preferiblemente, se realiza la transformación de ondícula discreta sobre datos brutos de la curva de fusión. Opcionalmente, sin embargo, se puede realizar la reducción de datos sobre los datos brutos de acuerdo con cualquier método matemático conocido en la técnica para generar una selección de datos brutos. En este último caso, se aplicará la transformación de ondícula discreta a dicha selección de datos brutos para producir también coeficientes tod. En resumen, en una realización particular, los resultados obtenidos de (a) pueden ser coeficientes tod obtenidos de datos brutos de curva de fusión. En una realización alternativa, los resultados obtenidos de (a) pueden ser coeficientes tod obtenidos de una selección de datos brutos de la curva de fusión. En aún otra realización específica, los resultados obtenidos de (a) pueden ser una selección de coeficientes tod obtenidos de cualquiera de las anteriores realizaciones alternativas.

En una realización particular, la transformación de ondícula discreta es una transformación de ondícula discreta 1D. En una realización incluso más específica, la transformación de ondícula discreta 1D es una transformación de ondícula de Daubechies 1D.

Con objeto de aplicar una transformación de ondícula discreta, se necesita escoger una ondícula madre. En otra realización preferida, se aplica la transformación de ondícula discreta de Daubechies, la cual usa una ondícula madre de la familia de Daubechies, siendo lo más preferiblemente la ondícula madre DB8.

En principio, en posibles realizaciones alternativas, se puede usar cualquier función de transformación de ondícula adecuada para generar coeficientes significativos que capturen información que permita la discriminación a nivel de un solo nucleótido en el método de la invención. Como posibles ejemplos, se incluyen la ondícula Haar (que también se puede considerar parte de la familia de Daubechies), menos asimétrica, coiflet o mejor localizada. Realizaciones alternativas pueden usar algoritmos alternativos para calcular la tod, incluyendo el algoritmo lifting o la transformación de ondícula compleja de doble árbol. Otras formas de transformación de ondícula discreta incluyen la transformación de ondícula no diezmada o sin diezmar, en donde se omite la reducción de muestreo, o la transformación de Newland, en donde se forma una base ortonormal de ondículas a partir de filtros chistera apropiadamente construidos en espacio de frecuencia. Posiblemente existan otros ejemplos, y serán fácilmente aplicables a los métodos aquí descritos por una persona apropiadamente entrenada.

Una de las principales ventajas de los métodos de la invención es su sencilla automatización y adaptación, especialmente a los sistemas de PC estándar conocidos. Por lo tanto, en una realización particular, se proporciona un método en donde se realiza la determinación del número de nucleótidos en las repeticiones homopoliméricas antes enumeradas de un modo automatizado, por ej., mediante un programa. Esto se podría hacer en un sistema automatizado, por ejemplo, equipado con disposiciones apropiadas de hardware y software que pueden leer las señales obtenidas de los métodos de la invención, analizarlas y proporcionar una conclusión con respecto a la presencia o ausencia de un indel en el marcador de elección de una muestra dada. Un sistema particularmente adecuado para dicha automatización es la plataforma Idylla™ de Biocartis, que, además de realizar PCR y proporcionar interpretación de sus resultados, también automatiza completamente todo el procesamiento de la muestra y el proceso de trabajo de aislamiento de ácidos nucleicos. Por consiguiente, en una posible realización, la presente invención proporciona un método de muestra-a-resultado totalmente automatizado para analizar loci de IMS.

En aún otra realización atractiva de la invención, se realiza el método en donde la determinación del número de nucleótidos en cualquiera de las repeticiones homopoliméricas antes enumeradas se realiza además también en una muestra biológica de control. Dicha muestra estándar de control o referencia podría, por ej., ser un material derivado de un tumor IMS-A que se ha confirmado que tiene un indel en cualquiera de las repeticiones homopoliméricas de elección antes enumeradas, o una construcción de ácido nucleico sintético o aislado, por ej., un plásmido. Un estándar de referencia particularmente ventajoso podría ser, por ej., uno de los estándares Acrometrix, que contiene una mezcla de ADN tanto sintético como genómico. La tecnología usa una línea celular muy altamente caracterizada y secuenciada GM2438 como ADN de fondo genómico en la cual se introducen dianas sintéticas secuenciadas. En el enfoque Acrometrix, estas dianas son moléculas de ADN sintético lineal que comprenden una secuencia que imita una alteración asociada a, por ej., un biomarcador de elección, que en el contexto de la presente invención podría ser una secuencia de cualquiera de las repeticiones homopoliméricas antes descritas que comprenden un indel, preferiblemente con fines de PCR, junto con sus secuencias flanqueantes. Las dianas además comprenden una secuencia de "cola" copulada a la secuencia de alternancia-imitación antes descrita, lo que sirve además con fines de identificación y cuantificación. Se considera que la secuencia resultante es una secuencia híbrida, que comprende la secuencia de alteración-imitación y la secuencia de cola. La cola, por ej., puede estar imitando una alteración conocida, tal como una SNP en un gen para el que se dispone de ensayos de detección, y, por lo tanto, puede proporcionar un medio adicional para cuantificar indirecta y absolutamente la secuencia de alteración-imitación, como en este caso un indel en el marcador o marcadores de elección. Dicho estándar podría ser, por ej., útil con fines de verificación y validación, por ejemplo en caso de que se contemplara otra investigación SSG, especialmente que los enfoques SSG actuales aún tienden a perder información sobre indeles en secuencias de repeticiones homopoliméricas.

De manera alternativa, en una realización preferida de la invención, la muestra biológica de control comprende material derivado de la línea celular HTC116 cl.110268743 que se generó para los fines de la presente invención y se depositó bajo el Tratado de Budapest el 28 de noviembre de 2017 en la autoridad depositaria BCCM/GeneCorner en Bélgica con el número de acceso LMBP 12278CB. La línea celular comprende una delección de homonucleótido en cada una de las repeticiones homopoliméricas antes mencionadas. Esto significa que el genoma de la línea celular contiene las siguientes repeticiones homopoliméricas mutantes (es decir, variantes IMS):

- 10 adeninas localizadas en el gen DIDO1 humano y que comienzan en la posición chr20:62,905,340;
- 10 adeninas localizadas en el gen MRE11 humano y que comienzan en la posición chr11:94,479,765;
- 9 adeninas localizadas en el gen SULF2 humano y que comienzan en la posición chr20:47,657,577;
- 7 adeninas localizadas en el gen ACVR2A humano y que comienzan en la posición chr2:147,926,117;
- 9 adeninas localizadas en el gen BTBD7 humano y que comienzan en la posición chr14:93,241,685;
- 8 timinas localizadas en el gen SEC31A humano y que comienzan en la posición chr4:82,864,412; y
- 9 adeninas localizadas en el gen RYR3 humano y que comienzan en la posición chr15:33,865,341.

Además, la línea celular también contiene un indel para otras varias repeticiones asociadas a IMS, por ej., BAT25 y BAT26 del panel Bethesda, que puede usarse en estudios comparativos.

En un aspecto relacionado, la presente invención también proporciona una célula o cualquier otro material, en particular material genético, que deriva de la línea celular HTC116 cl.110268743. Dicho material podría ser ADN genómico aislado o un lisado celular. Otras formas apropiadas de dicho material serán obvias para el experto dependiendo del diseño final de los métodos aquí proporcionados y kits basados en dichos métodos.

En otro aspecto, la presente invención también proporciona la realización de los métodos usando kits para detectar indeles en el panel de biomarcadores IMS de la invención. En una realización particular, los presentes métodos de la invención usan un kit para analizar loci de IMS en una muestra biológica, comprendiendo el kit herramientas para detectar regiones de ácido nucleico que comprenden las repeticiones homopoliméricas antes descritas proporcionadas en el panel de biomarcadores de la invención. Preferiblemente, dichas herramientas son específicas de secuencia, es decir, que están diseñadas para reconocer de un modo específico de secuencia dichas repeticiones homopoliméricas con sus regiones flanqueantes de longitud seleccionada. En una realización específica, las herramientas comprenden al menos un cebador o par de cebadores seleccionado de las SEQ ID NO.: 1-14. En una realización específica alternativa, las herramientas comprenden al menos una sonda de baliza molecular seleccionada de las SEQ ID NO.: 15-21. En una realización posible, las herramientas comprenden al menos un cebador o par de cebadores seleccionado de las SEQ ID NO.: 1-14 y al menos una sonda de baliza molecular seleccionada de las SEQ ID NO.: 15-21. Las herramientas pueden además comprender, por ej., una polimerasa correctora, sistemas tampón apropiados, dNTP, una selección de tintes posiblemente con desactivadores compatibles, etc. En otra realización, se proporciona el kit que comprende material de muestra biológica de control, siendo preferiblemente el material derivado de la línea celular HTC116 cl.110268743

En una realización preferida, se proporciona el kit que además comprende un cartucho. Posiblemente, se puede proporcionar el kit en forma de un cartucho. Por lo tanto, ventajosamente, la presente invención proporciona un kit en donde se proporcionan dichas herramientas para detectar regiones de ácidos nucleicos que comprenden las repeticiones homopoliméricas del panel de biomarcadores antes descritas en un cartucho conectable con un sistema automatizado. Como se ha descrito anteriormente, un ejemplo adecuado de un cartucho y un sistema automatizado conectable con el mismo es la plataforma Idylla™ de Biocartis. Se pueden encontrar más detalles de éste y sistemas de manera similar aplicables a la presente invención en WO2007004103, EP1896180, EP1904234, y EP2419705. Como puede apreciarse por los documentos aquí citados, los cartuchos ventajosos no sólo comprenden medios para llevar a cabo la PCR, sino que también pueden estar diseñados para aceptar directamente una fuente de ácido nucleico o una muestra, aislado o ácidos nucleicos liberados de dicha fuente de ácido nucleico, y proporcionan (por ej., bombeando) el ácido nucleico así liberado para el posterior ensayo basado en PCR.

En una realización preferida, se pueden proporcionar las herramientas, tales como cebadores, sondas y/u otros reactivos, incluyendo la polimerasa correctora, en dicho cartucho en un formato de puntos, que contribuye a una mayor vida útil.

En otro aspecto relacionado, la presente invención también proporciona la realización de los presentes métodos en sistemas automatizados.

En una realización posible, dicho sistema automatizado puede comprender una consola y un instrumento compatible con los cartuchos reutilizables de la invención. El instrumento comprende módulos de control para realizar ensayos. La consola es un ordenador para controlar y monitorizar las acciones del instrumento y el estado del cartucho durante los ensayos. El ensayo se realizará preferiblemente por entero dentro del cartucho y puede incluir, por ejemplo, una PCR en tiempo real. Después de insertar una muestra en dicho cartucho de la invención que está precargado con reactivos como se ha descrito anteriormente, se carga el cartucho en el instrumento y el instrumento controla el ensayo que se realiza de manera autónoma en el cartucho. Después de realizado el ensayo, el software de la consola procesa los resultados y genera un informe accesible para el usuario final del sistema automatizado.

El sistema automatizado puede ser un sistema automatizado abierto o cerrado. Después de añadir una muestra o de insertarla en el cartucho, se carga el cartucho en el sistema, que se cierra luego y permanece cerrado durante la operación del sistema. El sistema cerrado contiene todos los reactivos necesarios a bordo, por lo que la configuración cerrada proporciona la ventaja de que el sistema realiza una detección libre de contaminación. De manera alternativa, se puede usar un cartucho accesible abierto en un sistema automatizado. Se añaden los reactivos necesarios al cartucho abierto según se requiera, se puede insertar a continuación una muestra en el cartucho abierto y se puede usar el cartucho en un sistema automatizado cerrado.

Preferiblemente, se usan sistemas basados en cartucho que contienen una o más cámaras de reacción y una o más cámaras de fluido. Algunas de las cámaras de fluido pueden contener fluido que se usa para producir un lisado a partir de la muestra. Otras cámaras pueden contener fluidos tales como tampones de reacción, fluidos de lavado y soluciones de amplificación. Se usan las cámaras de reacción para llevar a cabo las diferentes etapas de la detección, tal como lavado, lisis y amplificación.

En una realización particularmente deseada según las realizaciones antes enumeradas, para mejorar y facilitar la interpretación de los resultados del método según la presente invención, también se realiza el análisis sobre la curva de fusión de un modo automatizado por medio de un método implementado por ordenador.

- 5 Por último, es también el objeto de la presente invención proporcionar una utilización de los paneles de biomarcadores, los métodos, los kits que incluyen cartuchos y los sistemas automatizados según la invención, en el análisis de loci de IMS en una muestra tumoral o una muestra biológica que se espera comprenda material tumoral.

En una realización preferida, el tumor es un cáncer colorrectal (CCR). En una realización alternativa, el tumor es un cáncer ovárico o endometrial. En aún otra realización, el tumor es un cáncer gástrico.

- 10 En una realización posible, la invención también proporciona una utilización de los métodos según la invención en el análisis de loci de IMS en una muestra de tumor y en la predicción de la respuesta a la inmunoterapia del sujeto del que derivó la muestra tumoral, en base al análisis. Se puede contemplar este último uso en vista de los informes recientes de la literatura, notablemente Turajlic *et al.*, 2017, Lancet Oncology, en donde se vio que una mayor acumulación de indeles en el genoma se correlacionaba con la generación de nuevos marcos abiertos de lectura codificantes de una gran cantidad de secuencias neoantigénicas. En base a esto, se podría asumir que la detección de al menos 2 ó 3 indeles en el panel de biomarcadores de la invención guardará correlación con el número total de indeles y neoantígenos, y, por lo tanto, un fenotipo inmunogénico distinto de un tumor según se analiza mediante los métodos o kits de la invención. Como resultado, se esperaría que el tumor del que derivó dicha muestra fuera altamente respondedor a la inmunoterapia con anticuerpos dirigidos a moléculas de punto de control inmune, tales como PD-1, PD-L1 o PD-L2, y que dicho tumor también fuera altamente respondedor a enfoques que se dirigen específicamente a los neoantígenos generados por medio de célula T antigénica quimérica o terapias con vacunas terapéuticas. Este y otros usos de la presente invención en el diagnóstico, pronóstico y seguimiento clínico de sujetos serán además fácilmente derivables para el experto en la técnica.

Ejemplos

- 25 Un conjunto mínimo de 4 marcadores no es sencillo de derivar de cualquier conjunto dado de marcadores. Por ejemplo, el descrito por Zhao *et al.*, 2014, eLife, el análisis con Sequenom de 18 muestras IMS-A usando un panel de 59 marcadores reveló que un marcador se identifica como media como mutante en un 44,26% de las muestras. Aunque este gran panel de marcadores es altamente eficaz en la detección del estado de IMS, conjuntos aleatorios derivados del mismo de 4 marcadores seleccionados muestran un rendimiento teórico mucho peor en comparación con el conjunto nuclear aquí propuesto que comprende ACRV2A, DIDO1, MRE11, SULF2. Dichos paneles aleatoriamente seleccionados son adicionalmente proclives a sufrir el inconveniente de que pueden contener marcadores que exhiben diferencias dependientes de la etnicidad en la región homopolimérica, tal como se vio, por ej., en el marcador TMEM65 para subpoblaciones caribeñas. Dichas diferencias hacen extremadamente difícil diseñar un panel de marcadores de bajo número robusto y eficaz, ya que pueden comprometer la correcta interpretación de los cambios dirigidos por IMS. Esto último resulta especialmente relevante cuando se identifica una baja cantidad de marcadores variables y/o cuando se carece de un control apropiado, lo que se ve comúnmente, por ej., con el conjunto clásico de Bethesda de biomarcadores IMS, que tienen un amplio intervalo de variación individual y múltiples alelos variantes, especialmente en poblaciones africanas (Buhard *et al.*, 2006).

Perfil de IMS en CCR, cáncer gástrico y cáncer de endometrio

- 40 Se hizo el perfil del estado de 7 marcadores de microsatélites (BTBD7, RYR3, SEC31A, ACVR2A, DIDO1, MRE11 y SULF2) en 128 muestras de cáncer colorrectal IMS-A. Se incluyeron varios sitios clínicos y diferentes grupos étnicos para valorar la robustez de la selección de marcadores. Además, se comprobó el estado de los 7 marcadores en 15 muestras de cáncer gástrico IMS-A y 19 muestras de cáncer de endometrio IMS-A. Se determinó la longitud de las repeticiones en ADN FFPE por PCR y se siguió con la caracterización del producto de amplificación con balizas moleculares.

45 Materiales y métodos

Muestras. En total, se obtuvieron 128 muestras FFPE de CCR IMS-A humano de diferentes fuentes, incluyendo Universidad de Cambridge, Instituto Portugués de Oncología de Porto, Cureline, Boca Biologics, Trans-Hit, Geneticist inc, Righshospitalet, Origene y Asterand. Se obtuvieron 15 muestras gástricas IMS-A humanas de Cureline y Trans-Hit y se obtuvieron 19 muestras de endometrio IMS-A humano de IDIBELL.

- 50 Procesamiento de muestras. Se insertó cada una de las muestras FFPE IMS-A en un cartucho de fluido patentado por Biocartis Idylla™. Se cerraron los cartuchos y se cargaron en la plataforma Idylla™ para análisis genéticos basados en PCR automatizados, después de lo cual se inició un procesamiento de muestras totalmente automatizado. Resumiendo, se liberó ADN de las muestras FFPE según el protocolo de licuefacción de FFPE de Biocartis, y se bombeó después a los compartimentos de PCR de los cartuchos según un protocolo Idylla™ estándar.

- 55 PCR. Se cargaron los compartimentos de PCR del cartucho para que contuvieran las siguientes mezclas de PCR por par de cebadores o dúplex de par de cebadores, como se indica a continuación:

BTBD7	Concentración final en PCR	
	Conc.	Unidades
Cebador directo	0,2	μM
Cebador inverso	1	μM
Baliza molecular	0,2	μM
dNTP	1,1	mM
Trehalosa	24	mM
Enzima: Veraseq Ultra	5	U
Tris pH 8	100	mM
MgCl ₂	3	mM
KCl	37,5	mM
(NH ₄) ₂ SO ₄	20	mM
ProClin 300	0,05	%

RYR3	Concentración final en PCR	
	Conc.	Unidades
Cebador directo	0,2	μM
Cebador inverso	1	μM
Baliza molecular	0,2	μM
dNTP	1	mM
Trehalosa	24	mM
Enzima: Veraseq Ultra	5,0	U
Tris pH 8	100	mM
MgCl ₂	3	mM
KCl	37,5	mM
(NH ₄) ₂ SO ₄	20	mM
ProClin 300	0,05	%

ACVR2A/SEC31A

		Concentración final en PCR	
		Conc.	Unidades
ACVR2A	Cebador directo	0,2	μM
	Cebador inverso	1	μM
	Baliza molecular	0,2	μM
SEC31A	Cebador directo	0,2	μM
	Cebador inverso	1	μM
	Baliza molecular	0,2	μM
dNTP		1,05	mM
Trehalosa		24	mM
Enzima: Veraseq Ultra		5,0	U
Tris pH 8		100	mM
MgCl ₂		3	mM
KCl		37,5	mM
(NH ₄) ₂ SO ₄		20	mM
ProClin 300		0,05	%

DIDO1/MRE11

		Concentración final en PCR	
		Conc.	Unidades
DIDO1	Cebador directo	0,2	μM
	Cebador inverso	3	μM
	Baliza molecular	0,2	μM
MRE11	Cebador directo	0,2	μM
	Cebador inverso	3	μM
	Baliza molecular	0,2	μM
dNTP		0,9	mM
Trehalosa		24	mM
Enzima: Veraseq Ultra		5,0	U
Tris pH 8		100	mM
MgCl ₂		3	mM
KCl		37,5	mM
(NH ₄) ₂ SO ₄		20	mM

ProClin 300	0,05	%
-------------	------	---

SULF2	Concentración final en PCR	
	Conc.	Unidades
Cebador directo	0,2	μM
Cebador inverso	1	μM
Baliza molecular	0,2	μM
dNTP	1	mM
Trehalosa	24	mM
Enzima: Veraseq Ultra	5,0	U
Tris pH 8	100	mM
MgCl ₂	3	mM
KCl	37,5	mM
(NH ₄) ₂ SO ₄	20	mM
ProClin 300	0,05	%

Las secuencias de los pares de cebadores y sondas por marcador eran las siguientes:

marcador	oligo	secuencia	SEQ ID NO.
DIDO1	directo	TAGCGTGAATCGGACAT	SEQ ID NO.: 1
	inverso	TTGACTGGGCAGATAGGGGA	SEQ ID NO.: 2
	sonda	CGCACGCATGGAAAAAAAAAATCCGTGCGTAAA	SEQ ID NO.: 15
MRE11	directo	ATAGTTCACCCATGGAACCC	SEQ ID NO.: 3
	inverso	GGAGGAGAATCTTAGGGAAA	SEQ ID NO.: 4
	sonda	CGTCGAACCTTAAAAAAAAAAGTTACCGACGAA	SEQ ID NO.: 16
BTBD7	directo	ACTGGACTCCCGCTGG	SEQ ID NO.: 5
	inverso	CGCTCAGCCTCCATAAATC	SEQ ID NO.: 6
	sonda	CGCACGACTTATTAATAAAAAAAAAATGACAGTGCGTAAA	SEQ ID NO.: 17
SULF2	directo	CAACTTCATTTCTTTTCAGTACCTT	SEQ ID NO.: 7
	inverso	CTGTCCAGATACCATTCTC	SEQ ID NO.: 8
	sonda	CGTCGGTACCTTAAAAAAAAACATCACGACGAA	SEQ ID NO.: 18
ACVR2A	directo	AGCTCCATCTCTTGAAGACAT	SEQ ID NO.: 9
	inverso	GCATGTTTCTGCCAATAATCTCT	SEQ ID NO.: 10
	sonda	GTGCATAAAAAAAAAAGAGCACTAAA	SEQ ID NO.: 10

			19
SEC31A	directo inverso sonda	CAACTTCAGCAGGCTGT AGTCTGAGAAAGCATCAATTTT CGCACTTGCCAAAAAAATTGATGGTGCGTAA	SEQ ID NO.: 11 SEQ ID NO.: 12 SEQ ID NO.: 20
RYR3	directo inverso sonda	CATTTTCTAAATGCCTCCCTTAAA GTCCATTAGGCACAAAAAG CGTCGCCCTTAAAAAAAAGTCCGACGAA	SEQ ID NO.: 13 SEQ ID NO.: 14 SEQ ID NO.: 21

Las condiciones de PCR eran las siguientes:

	Temperatura (°C)	Tiempo (min:s)	# Ciclos
Desnaturalización inicial	92	00:30	
Amplificación	92 60	00:10 00:10	60
Desnaturalización	92	02:00	
Curva de fusión	de 40 a 76,6°C, incremento 0,3°C	00:05	

5 Fusión de amplicones. Se desnaturalizaron los productos de la PCR en los cartuchos durante 2 min a 92°. A continuación, se recogieron los datos de fluorescencia de la curva de fusión calentando la mezcla de 40°C a 76,6°C en etapas de 0,3°C (12 s por ciclo) y monitorizando al mismo tiempo las señales de fluorescencia después de cada incremento de 0,3°C.

10 Posprocesado. En la primera etapa del análisis de datos, se recuperan los valores brutos de la medición de fluorescencia de la curva de fusión del instrumento Idylla™. En la siguiente etapa, sólo se retienen los primeros 64 ciclos del vector de valores de medición. Este subgrupo se denomina la "región de interés" o RDI, ya que los cambios más importantes en la señal se observan dentro de esta ventana. En posteriores ciclos, se observa mayormente la fusión de balizas. La siguiente etapa en el algoritmo de posprocesado es la aplicación de una transformación de ondícula discreta (TOD) al vector de medición (ROI). Las ondículas son especialmente muy adecuadas para el análisis de una curva de fusión, ya que éste es un fenómeno de baja frecuencia que se produce en una zona de temperatura específica. Las ondículas son capaces de realizar un análisis de tiempo y frecuencia al mismo tiempo. Esto significa que pueden interpretar tanto lo que está sucediendo en términos de cambios de baja frecuencia como cuándo está sucediendo esto. Como tales, las ondículas resumen de manera compacta el procedimiento de la curva de fusión. En este caso específico, se usa la ondícula DB8 y se retienen los coeficientes de tercer nivel. Después de esta transformación, se retienen los coeficientes tanto de escala como de ondícula, para dar como resultado dos grupos de 8 coeficientes. Se calcula un conjunto de 16 coeficientes de ondícula para cada marcador presente en el ensayo. Este conjunto de coeficientes de ondícula por marcador se denomina los resultados de posprocesado para cada marcador.

25 Árbol de decisiones. La segunda etapa del análisis de datos de los datos de la curva de fusión se denomina el árbol de decisiones. En esta etapa, se aplica un algoritmo de reconocimiento de patrón con objeto de clasificar muestras válidas en base a los resultados del posprocesado. Con este fin, se aplica un algoritmo de clasificación, que es una red neural, a los resultados del posprocesado de cada marcador. Se ha entrenado esta red con datos marcados para los que se conoce el genotipo de entrada. En el caso de datos de referencia, este marcaje se basa en el genotipo de entrada que se indicó. Para datos clínicos, se obtiene este marcaje de la puntuación visual de los perfiles de fusión por expertos en curvas de fusión. A través de la optimización iterativa de los pesos en la red neural, el algoritmo puede aprender a distinguir entre una curva de tipo salvaje (WT) y una mutante. El algoritmo da una puntuación de

probabilidad como salida para cada gen marcador que refleja la certeza de la decisión (1 para mutante y 0 para WT). Se puntúa una muestra como IMS-A si al menos dos marcadores tienen una puntuación de probabilidad mayor de 0,5.

Resultados

1. Perfil de IMS en CCR

- 5 Primer análisis. Se valoró un conjunto nuclear de cuatro marcadores que incluía ACVR2A, DIDO1, MRE11 y SULF2 en cuanto a su capacidad para recuperar muestras positivas IMS-A del conjunto de 128 muestras IMS-A. Se puntúa una muestra como positiva cuando el árbol de decisiones de los datos de la curva de fusión posprocesados da como resultado al menos dos marcadores que se detectan como comprendiendo un indel. Usando este conjunto nuclear de cuatro marcadores, se podía identificar el 96% de las muestras como IMS-A. Dado que se ha definido el rendimiento
- 10 mínimo aceptable para recuperar al menos el 95% de las muestras, se ha aceptado la selección hecha anteriormente como el conjunto nuclear de marcadores de IMS.

# Marcadores	Marcadores	# IMS	% IMS
4 Nucleares	ACVR2A, DIDO1, MRE11, SULF2	123	96%

- Para crear otra valoración del rendimiento de estos marcadores, se sometieron todas las permutaciones posibles de 3 marcadores del panel nuclear de 4 al mismo análisis de rendimiento. Los resultados indican que el rendimiento es peor para cada posible subselección, yendo del 83% al 93%. Para suministrar más robustez al ensayo a través de los
- 15 tipos de cáncer, se prefiere un diseño con al menos cuatro marcadores.

# Marcadores	Marcadores	# IMS	% IMS
Sub 3	ACVR2A, DIDO1, SULF2	119	93%
Sub 3	ACVR2A, DIDO1, MRE11	118	92%
Sub 3	ACVR2A, MRE11, SULF2	109	85%
Sub 3	DIDO1, MRE11, SULF2	106	83%

Segundo análisis. Añadiendo un marcador adicional (BTBD7) al conjunto nuclear de cuatro marcadores, se puede puntuar una muestra más como IMS-A, lo que representa el 97% de muestras correctamente puntuadas y, por lo tanto, hace que el panel sea más eficiente al definir el estado de IMS.

# Marcadores	Marcadores	# IMS	% IMS
4 Nucleares + BTBD7	ACVR2A, DIDO1, MRE11, SULF2, BTBD7	124	97%

- 20 Tercer análisis. Añadiendo un marcador adicional (SEC31A) al conjunto de cinco marcadores, todas las muestras en el conjunto de muestras limitado estudiado pueden puntuarse como IMS-A, lo que proporciona incluso otra mejora.

# Marcadores	Marcadores	# IMS	% IMS
4 Nucleares + BTBD7 + SEC31A	ACVR2A, DIDO1, MRE11, SULF2, BTBD7, SEC31A	128	100%

- Cuarto análisis. Añadiendo otro marcador (RYR3) al conjunto de seis marcadores, naturalmente, aún todas las muestras se puntúan como IMS-A. Aunque no es inmediatamente visible por los presentes datos, en teoría, la adición de un 7º marcador es probable que aún aumente el rendimiento del ensayo en caso de analizar mayores conjuntos de muestras. En teoría, dadas las frecuencias observadas de estos marcadores mutados en muestras IMS-A, para 7
- 25 marcadores se predice que el índice de falsos negativos será de tan sólo <1/1900, el cual, para mayores conjuntos de muestras, puede volverse relevante.

# Marcadores	Marcadores	# IMS	% IMS
4 Nucleares + BTBD7 + SEC31A + RYR3	ACVR2A, DIDO1, MRE11, SULF2, BTBD7, SEC31A, RYR3	128	100%

Se muestran los resultados del análisis de las muestras de CCR en la Figura 1.

2. Perfil de IMS en cáncer gástrico y cáncer de endometrio

- 30 Primer análisis. Se valoró también entonces el conjunto nuclear de los cuatro marcadores con mejor rendimiento en CCR (ACVR2A, DIDO1, MRE11 y SULF2) en un conjunto de 34 muestras de cáncer que incluían 15 muestras de cáncer gástrico y 19 muestras de cáncer de endometrio (EN).

# Marcadores	Marcadores	# IMS	% IMS
4 Nucleares	ACVR2A, DIDO1, MRE11, SULF2	32	94%

Los resultados sugieren que la repetición homopolimérica en el ACVR2A es un marcador mucho más potente en CCR que en otros tipos de cáncer. Sin embargo, a pesar de un conjunto de muestras muy pequeño, los resultados del conjunto nuclear en el cáncer gástrico y de EN están muy próximos al umbral aceptable del 95%, lo cual es un buen indicador de que en un conjunto mayor de muestras el conjunto nuclear de 4 marcadores aquí presentado puede aplicarse generalmente a otros tipos de tumores deficientes en RDA o inestables en microsatélites y no sólo CCR. Con objeto de tener una mejor visión del rendimiento del panel nuclear, habría que hacer el perfil de más muestras.

Segundo análisis. Añadiendo un marcador adicional (BTBD7) al conjunto nuclear de cuatro marcadores, se puede puntuar una muestra más como IMS-A, lo que en este pequeño conjunto de muestras ya lleva el rendimiento de dicho panel de cinco marcadores a un valor altamente satisfactorio del 97% de muestras correctamente puntuadas.

# Marcadores	Marcadores	# IMS	% IMS
4 Nucleares + BTBD7	ACVR2A, DIDO1, MRE11, SULF2, BTBD7	33	97%

Tercer análisis. Añadiendo un marcador adicional (SEC31A) al conjunto de cinco marcadores, se pueden puntuar todas las muestras como IMS-A.

# Marcadores	Marcadores	# IMS	% IMS
4 Nucleares + BTBD7 + SEC31A	ACVR2A, DIDO1, MRE11, SULF2, BTBD7, SEC31A	34	100%

Conclusiones

El panel aquí presentado de sólo cuatro marcadores nucleares que incluyen las repeticiones homopoliméricas en los genes ACVR2A, DIDO1, MRE11 y SULF2, muestra un rendimiento extremadamente bueno en muestras de CCR. También muestra un rendimiento muy bueno en muestras de cáncer gástrico y endometrial, incluso a pesar del hecho de que tuvimos acceso y pudimos hacer el perfil de sólo muy pocas de ellas. El perfil de más muestras IMS-A de origen distinto del CCR corroborará probablemente la aplicabilidad del panel nuclear mínimo de dos marcadores a un espectro más amplio de cánceres. Actualmente, parece que el homopolímero en el gen ACVR2A es un marcador de IMS particularmente eficaz y bastante específico para CCR. Por consiguiente, en otros tipos de cáncer, como realizaciones alternativas de la invención, posiblemente se podrían proponer y estudiar otros paneles nucleares mínimos. Por los datos preliminares aquí presentados, parece que se podrían proponer los tres paneles de 4 marcadores nucleares siguientes: (1) DIDO1, SULF2, BTBD7 y SEC31A; (2) DIDO1, SULF2, BTBD7 y ACVR2A; y (3) DIDO1, SULF2, SEC31A y RYR3. Independientemente de los tipos de muestra deficiente en RDA, se vio que un panel de 5 marcadores nucleares que comprende ACVR2A, DIDO1, MRE11, SULF2, BTBD7 era generalmente adecuado para diagnosticar muestras de diferentes orígenes y, por lo tanto, constituye una realización particularmente atractiva de la presente invención.

Definiciones

Tal como se usa aquí, el término "*muestra biológica*", o simplemente "*muestra*", pretende incluir una variedad de fuentes biológicas que contienen ácido nucleico y/o material celular, independientemente de si se obtiene fresco de un organismo (es decir, muestra de tejido fresco) o está conservado por cualquier método conocido en la técnica (por ej., una muestra congelada o FFPE). Como ejemplos de muestras biológicas, se incluyen: cultivos de células, tales como células de mamífero, pero también de microorganismos eucarióticos, fluidos corporales, precipitados de fluidos corporales, espécimen de lavado, aspirados con aguja fina, muestras de biopsia, muestras de tejido, células cancerosas, otros tipos de células obtenidas de un paciente, células de un tejido o células cultivadas *in vitro* de un individuo al que se está estudiando y/o tratando por enfermedad o infección, o muestras forenses. Como ejemplos no limitantes de muestras de fluidos corporales, se incluyen sangre entera, médula ósea, líquido cefalorraquídeo (LCR), líquido peritoneal, líquido pleural, líquido linfático, suero, plasma, orina, quilo, heces, eyaculado, esputo, aspirado de pezón, saliva, espécimen de frotis, líquido de enjuagado o lavado y/o especímenes de cepillo.

El término "*ácido nucleico*" y su equivalente "*polinucleótido*", tal como se usan aquí, se refieren a un polímero de ribonucleótidos o desoxirribonucleótidos unidos entre sí por uniones fosfodiéster entre los monómeros de nucleótidos. Los (desoxi)nucleótidos son formas fosforiladas de (desoxi)nucleósidos, que lo más comúnmente incluyen adenosina, guanosina, citidina, timidina o uridina. Estos nucleósidos consisten en un azúcar pentosa, que es ribosa o desoxirribosa, y una base nitrogenada ("nucleobase", o simplemente, "base"), que es adenina, guanina (que son purinas), citosina, timina o uracilo (que son pirimidinas). La secuencia a la que estas bases (o sus nucleósidos, o los nucleótidos de estos últimos) siguen en una hebra de ácido nucleico se denomina "secuencia de ácido nucleico" y se da convencionalmente en una dirección así llamada de extremo 5' a extremo 3' refiriéndose a la orientación química de la hebra del ácido nucleico. El "5'" se origina de la referencia al carbono 5' del primer anillo de (desoxi)ribosa a partir del cual comienza la lectura de la secuencia de ácido nucleico, y el "3'" se origina del carbono 3' del último anillo de (desoxi)ribosa sobre el cual finaliza la lectura de la secuencia de ácido nucleico. Una secuencia de ácido nucleico

puede, por ej., ser ATATGCC, lo que se ha de interpretar aquí que hace referencia a una secuencia de ácido nucleico 5'- ATATGCC - 3'. Bajo la misma convención, esta última secuencia será complementaria a la secuencia 5' - GGCATAT-3', o simplemente GGCATAT. Una secuencia de ácido nucleico puede ser una secuencia de repetición homopolimérica, es decir, una secuencia hecha de un cierto número de nucleótidos consecutivos que comprenden la misma base nitrogenada, a los que también se denomina aquí "homonucleótidos". Por ejemplo, se ha de considerar que el término "repetición homopolimérica que comprende 8 adeninas consecutivas" se refiere a al menos una parte de un ácido nucleico, estando constituida dicha parte por un trayecto que comprende 8 nucleótidos consecutivos, en donde cada uno de dichos nucleótidos comprende una adenina como base nitrogenada. Dicha secuencia se designaría como 5'-AAAAAAA-3' (o simplemente AAAAAAA), mientras que su secuencia complementaria sería 5'-TTTTTTT-3' (o TTTTTTT). Se ha de considerar aquí que los términos "forma mutada de una repetición homopolimérica" o "formas mutadas de las mismas" se refieren a variantes IMS de una repetición homopolimérica dada que comprenden una inserción o una delección (es decir, un "indel") de al menos un homonucleótido. Por ejemplo, una forma mutada de la repetición homopolimérica que comprende 8 adeninas consecutivas sería una repetición homopolimérica que comprende 7 adeninas consecutivas, o una repetición homopolimérica que comprende 9 adeninas consecutivas. Los ácidos nucleicos incluyen, aunque sin limitación, ADN y ARN, incluyendo ADN genómico, ADN mitocondrial o ADNmt, ADNc, ARNm, ARNr, ARNt, ARNnh, microARN, ARNinc, ARNip y diversas versiones modificadas de los mismos. Los ácidos nucleicos pueden obtenerse lo más comúnmente de fuentes naturales, como muestras biológicas obtenidas de diferentes tipos de organismos. Por otro lado, también se pueden sintetizar, recombinar o de algún otro modo producir ácidos nucleicos en cualquiera de los métodos contemplados en humanos conocidos (por ej., PCR).

Se da aquí al término "*PCR cuantitativa*" o simplemente "*PCRc*" la definición de una técnica de laboratorio basada en la reacción en cadena de polimerasa (PCR), que se usa para amplificar y simultáneamente detectar o cuantificar una molécula de ADN diana. Contrariamente a la PCR estándar, en donde se detecta el producto de la reacción en su extremo, es decir, después de finalizar la termociclación, la característica clave de la PCRc es que el producto de ADN se está detectando durante la termociclación a medida que la reacción progresa en "tiempo real"; de ahí el nombre alternativo de la PCRc "*PCR en tiempo real*". Existen actualmente muchos tipos diferentes de PCRc. Por ejemplo, cuando se comienza con una etapa de transcripción inversa (RT), se puede usar la PCRc para cuantificar los números de ARN mensajeros y se llama entonces una PCRc con transcriptasa inversa o una RT-PCRc. Tal como se usan aquí, los términos "*PCR cuantitativa*" o simplemente "*PCRc*" se emplearán con preferencia sobre el término "*PCR en tiempo real*" o "*RT-PCR*" con objeto de evitar confusión con la PCR con transcripción inversa, también frecuentemente abreviada como RT-PCR. La mayoría de las PCRc usan uno de los dos métodos más comunes para detectar la amplificación del producto en tiempo real: (a) intercalación de tintes fluorescentes no específicos con cualquier ADN de doble hélice, o (2) sondas de ADN específicas de secuencia consistentes en oligonucleótidos que están marcados con un indicador fluorescente que permite la detección sólo tras hibridación de la sonda con su secuencia diana complementaria. Se detectan las señales fluorescentes generadas durante la termociclación mediante un sistema de detección óptico apropiado y se rastrean desde el momento en que pasan el umbral de fondo hasta que la reacción alcanza una meseta. Se puede estimar el número de copias de las secuencias diana usando una estrategia de cuantificación relativa o absoluta, típicamente analizando la forma de la curva de amplificación obtenida (estrategia de curva estándar) o determinando cuándo se eleva la señal por encima de algún valor umbral (frecuentemente llamado el valor Ct, pero a veces también valor Cp o valor Cq). En la cuantificación relativa, los niveles de ácido nucleico diana estimados en una muestra dada usando el análisis Ct o de curva estándar se expresan como relativos a los valores obtenidos para la misma diana en otra muestra de referencia, por ejemplo, una muestra de control no tratada. Por el contrario, en la cuantificación absoluta, la señal de PCRc se relaciona con el número de copias de entrada usando una curva estándar, o puede también calcularse según un método de PCR digital más reciente. Por el momento actual, la primera estrategia es aún más prevalente y basa la estimación de la cantidad de ADN diana comparando los valores obtenidos con una curva estándar previamente hecha. Estas y otras estrategias de cuantificación de PCRc son ampliamente conocidas en la técnica y su cálculo puede diferir en menor o mayor grado dependiendo de la aplicación dada y del sistema de PCRc.

Tal como se usa aquí, el término "*medios para realizar PCR cuantitativa*" se entenderá como la disposición mínima necesaria de reactivos y elementos para realizar una PCRc. Normalmente incluirán cualesquiera reactivos que permitan que sea detectable en la termociclación por PCR en tiempo real de una plantilla de ácido nucleico recibida de una fuente de ácido nucleico. Dichos reactivos incluyen, aunque dependiendo del tipo de PCRc no se limitan a ellos, una polimerasa de grado PCR, al menos un conjunto de cebadores, un tinte detectable o una sonda, dNTP, tampón de PCR, etc. Además, los "*medios para realizar PCR cuantitativa*" incluirán también normalmente cualquier estándar conocido en la técnica de montaje mínimo de partes, que normalmente incluye, aunque sin limitación, lo siguiente: (1) un compartimento adecuado (al que además se hace referencia como "*un compartimento de PCRc de termociclación*") en donde puede tener lugar la termociclación detectable en tiempo real. Dichos compartimentos pueden, por ej., estar formados por una cámara adecuada para amplificar ácidos nucleicos, es decir, hecha de material apropiado y que proporcione una regulación suficiente de la temperatura interna, y también que comprenda al menos una pared que permita la detección en tiempo real de las señales generadas durante dicha amplificación, por ej., una pared transparente a la luz. Además, (2) medios para variar la temperatura en esta cámara u otro compartimento, como se sabe ampliamente por diversas máquinas de termociclación existentes. Luego, (3) medios para detectar las señales generadas durante la termociclación por PCRc, como un detector óptico acoplado a un ordenador, etc. Resumiendo, dicho montaje mínimo incluirá normalmente cualquier sistema o sistemas conocidos en la técnica

capaces de iniciar y mantener la reacción de termociclación en el compartimento de PCRC de termociclación, ajustando y regulando la temperatura para asegurar condiciones de termociclación estables en el mismo, etc. Además, también incluirá cualquier dispositivo o dispositivos de detección apropiados, medios para procesamiento de datos (por ej., un ordenador conectado alternativamente a una base de datos), y sistemas de salida que permitan leer y monitorizar la termociclación de la reacción de PCRC en tiempo real (normalmente, una pantalla de ordenador que muestra el progreso de la reacción en una interfaz de usuario gráfica apropiada). Adicionalmente, también contendrá los paquetes de software adecuados para operar la maquinaria y/o mostrar y posiblemente ayudar a la interpretación de los resultados obtenidos.

Tal como se usa aquí, el término "*cartucho*" debe entenderse como un montaje autocontenido de cámaras y/o canales, que se forma como un único objeto que puede transferirse o moverse como un accesorio dentro o fuera de un instrumento mayor adecuado para aceptar o conectar a dicho cartucho. Algunas partes contenidas en el cartucho pueden estar firmemente conectadas, mientras que otras pueden estar conectadas flexiblemente y ser móviles con respecto a otros componentes del cartucho. De manera análoga, tal como se usa aquí el término "*cartucho de fluido*" se entenderá como un cartucho que incluye al menos una cámara o canal adecuado para tratar, procesar, descargar o analizar un fluido, preferiblemente un líquido. Se da un ejemplo de dicho cartucho en WO2007004103. Ventajosamente, un cartucho de fluido puede ser un cartucho de microfluido. En el contexto de cartuchos de fluido, los términos "*aguas abajo*" y "*aguas arriba*" pueden definirse como relacionados con la dirección en la que fluyen los fluidos en dicho cartucho. A saber, una sección de un trayecto de fluido en un cartucho del que fluye un fluido hacia una segunda sección en el mismo cartucho, se ha de interpretar como situada *aguas arriba* de la última. De manera análoga, la sección a la que llega un fluido más tarde se sitúa *aguas abajo* con respecto a una sección por la que pasó antes dicho fluido.

En general, tal como se usan aquí, los términos "*fluido*" o a veces "*microfluido*" se refieren a sistemas y disposiciones que tratan con el comportamiento, el control y la manipulación de fluidos que están geoméricamente constreñidos a una escala pequeña, típicamente submilimétrica, en al menos una o dos dimensiones (por ej., anchura y altura o un canal). Dichos fluidos de pequeño volumen se mueven, mezclan, separan o de algún otro modo procesan a una microescala que requiere pequeño tamaño y bajo consumo de energía. Los sistemas microfluidos incluyen estructuras tales como microsistemas neumáticos (fuentes de presión, bombas de líquido, microválvulas, etc.) y estructuras de microfluidos para el manejo de volúmenes de micro, nano- y picolitros (canales de microfluidos, etc.). Se describieron ejemplos de sistemas de fluidos en EP1896180, EP1904234 y EP2419705 y pueden, por consiguiente, aplicarse en ciertas realizaciones de la invención aquí presentada.

Tal como se usa aquí, el término "TOD" designa transformación de ondícula discreta; el término "coeficiente de tod" designa el coeficiente de transformación de ondícula discreta. Una transformación de ondícula significa un cálculo que utiliza un programa o subrutina sobre datos brutos. Por lo tanto, un conjunto de coeficientes tod es un conjunto de valores con ondícula transformada. Los coeficientes tod más relevantes para los análisis de ácidos nucleicos son aquellos coeficientes que capturan los sucesos significativos del experimento, por ejemplo, en caso de un experimento de fusión de una molécula de ácido nucleico de doble hélice, los coeficientes tod más relevantes pueden ser picos o desplazamientos de picos en los datos brutos de las curvas de fusión.

Tal como se usan aquí, los términos datos brutos de la curva de fusión, curva de fusión datos brutos y datos de la curva de fusión brutos son equivalentes y se usan indistintamente. Designan identificadores obtenidos siguiendo experimentos de disociación o asociación de ácidos nucleicos.

LISTA DE SECUENCIAS

	<110> Biocartis NV		
5	<120> MÉTODOS PARA EL ANÁLISIS DE LOS DATOS DE LA CURVA DE FUSIÓN DE DISOCIACIÓN		
	<130> A-1802P-EP		
	<160> 21		
10	<170> PatentIn versión 3.5		
	<210> 1		
	<211> 19		
15	<212> ADN		
	<213> Homo sapiens		
	<400> 1		
20	tagcgtgtga atcggacat	19	
	<210> 2		
	<211> 20		
	<212> ADN		
	<213> Homo sapiens		
25	<400> 2		
	ttgactgggc agatagggga	20	
	<210> 3		
30	<211> 20		
	<212> ADN		
	<213> Homo sapiens		
	<400> 3		
35	atagttcacc catggaaacc	20	
	<210> 4		
	<211> 20		
	<212> ADN		
40	<213> Homo sapiens		
	<400> 4		
	ggaggagaat ctagggaaa	20	
45	<210> 5		
	<211> 16		
	<212> ADN		
	<213> Homo sapiens		
50	<400> 5		
	actggactcc cgctgg	16	
	<210> 6		
	<211> 19		
55	<212> ADN		
	<213> Homo sapiens		
	<400> 6		
60	cgctcagcct ccataaatc	19	
	<210> 7		
	<211> 25		
	<212> ADN		
	<213> Homo sapiens		

<400> 7
 caacttcatt tcttttcagt acctt 25
 5 <210> 8
 <211> 20
 <212> ADN
 <213> Homo sapiens
 10 <400> 8
 ctgtccagat accatttctc 20
 <210> 9
 <211> 22
 15 <212> ADN
 <213> Homo sapiens
 <400> 9
 agcatccatc tcttgaagac at 22
 20 <210> 10
 <211> 23
 <212> ADN
 <213> Homo sapiens
 25 <400> 10
 gcatgtttct gccataatc tct 23
 <210> 11
 30 <211> 17
 <212> ADN
 <213> Homo sapiens
 <400> 11
 35 caacttcagc aggctgt 17
 <210> 12
 <211> 21
 <212> ADN
 40 <213> Homo sapiens
 <400> 12
 agtctgagaa gcatcaattt t 21
 45 <210> 13
 <211> 24
 <212> ADN
 <213> Homo sapiens
 50 <400> 13
 cattttctaa atgcctccct taaa 24
 <210> 14
 <211> 19
 55 <212> ADN
 <213> Homo sapiens
 <400> 14
 gtccattagg cacaaaaag 19
 60 <210> 15
 <211> 34
 <212> ADN
 <213> ADN sintético
 65 <400> 15

	cgcacgacat ggaaaaaaaa aatccgtgcg taaa	34
	<210> 16	
	<211> 33	
5	<212> ADN	
	<213> ADN sintético	
	<400> 16	
10	cgtcgaacct taaaaaaaaa agttaccgac gaa	33
	<210> 17	
	<211> 36	
	<212> ADN	
	<213> ADN sintético	
15	<400> 17	
	cgcacgactt attaaaaaaaa aatgacagtg cgtaaa	36
	<210> 18	
20	<211> 33	
	<212> ADN	
	<213> ADN sintético	
	<400> 18	
25	cgtcggtacc taaaaaaaaa acatcacgac gaa	33
	<210> 19	
	<211> 23	
	<212> ADN	
30	<213> ADN sintético	
	<400> 19	
	gtgcataaaa aaagagcact aaa	23
35	<210> 20	
	<211> 33	
	<212> ADN	
	<213> ADN sintético	
40	<400> 20	
	cgcacttgcc aaaaaaaatt gatggtgctg aaa	33
	<210> 21	
	<211> 30	
45	<212> ADN	
	<213> ADN sintético	
	<400> 21	
50	cgtcgccctt aaaaaaaaaac tgccgacgaa	30

REIVINDICACIONES

1. Un método para analizar el estado de IMS de un cáncer colorrectal, endometrial o gástrico analizando los loci de IMS en una muestra biológica, comprendiendo el método la etapa de:
 - 5 - determinar el número de nucleótidos en las siguientes repeticiones homopoliméricas mapeadas en el genoma de referencia humano GRCh38/hg38:

repetición homopolimérica que comprende 11 adeninas consecutivas localizadas en el gen DIDO1 humano y que comienza en la posición chr20:62,905,340;

repetición homopolimérica que comprende 11 adeninas consecutivas localizadas en el gen MRE11 humano y que comienza en la posición chr11:94,479,765;

repetición homopolimérica que comprende 10 adeninas consecutivas localizadas en el gen SULF2 humano y que comienza en la posición chr20:47,657,577; y

repetición homopolimérica que comprende 8 adeninas consecutivas localizadas en el gen ACVR2A humano y que comienza en la posición chr2:147,926,117.
 - 15 2. El método según la reivindicación 1, que además comprende determinar el número de nucleótidos en una cualquiera, dos o la totalidad de las siguientes regiones de repeticiones homopoliméricas mapeadas en GRCh38/hg38:

repetición homopolimérica que comprende 9 timinas consecutivas localizadas en el gen SEC31A humano y que comienza en la posición chr4:82,864,412;

repetición homopolimérica que comprende 10 adeninas consecutivas localizadas en el gen BTBD7 humano y que comienza en la posición chr14:93,241,685;

repetición homopolimérica que comprende 10 adeninas consecutivas localizadas en el gen RYR3 humano y que comienza en la posición chr15:33,865,341.
 3. El método según cualquiera de las reivindicaciones 1-2, que además comprende la etapa de
 - 25 - amplificar regiones de ácidos nucleicos que comprenden las repeticiones homopoliméricas o sus formas mutadas, enumeradas en las reivindicaciones 1-2.
 4. El método según la reivindicación 3, en donde la etapa de amplificación da como resultado la generación de datos de la curva de fusión.
 5. El método según la reivindicación 4, que además comprende las etapas de:
 - (a) aplicar transformación de ondículas a los datos de la curva de fusión; y
 - 30 (b) usar los resultados obtenidos de (a) en la determinación del número de nucleótidos en cualquiera de las repeticiones homopoliméricas enumeradas en las reivindicaciones 1-2.
 6. El método según cualquiera de las reivindicaciones 3-5, en donde la etapa de amplificación comprende el uso de al menos una sonda de baliza molecular.
 7. El método según cualquiera de las reivindicaciones 3-6, en donde la etapa de amplificación comprende el uso de al menos un cebador que tiene la secuencia identificada por cualquiera de las siguientes SEQ ID NO. 1-14.
 8. El método según cualquiera de las reivindicaciones 3-7, en donde la etapa de amplificación comprende al menos una amplificación dúplex de un par de repeticiones homopoliméricas, o sus formas mutadas, siendo seleccionado dicho par de las siguientes combinaciones:
 - amplificación dúplex de
 - 40 la repetición homopolimérica que comprende 11 adeninas consecutivas localizadas en el gen DIDO1 humano y que comienza en la posición chr20:62,905,340
 - junto con
 - la repetición homopolimérica que comprende 11 adeninas consecutivas localizadas en el gen MRE11 humano y que comienza en la posición chr11:94,479,765;
 - 45 - amplificación dúplex de

la repetición homopolimérica que comprende 8 adeninas consecutivas localizadas en el gen ACVR2A humano y que comienza en la posición chr2:147,926,117

junto con

5 la repetición homopolimérica que comprende 9 timinas consecutivas localizadas en el gen SEC31A humano y que comienza en la posición chr4:82,864,412; y

- amplificación dúplex de

la repetición homopolimérica que comprende 10 adeninas consecutivas localizadas en el gen BTBD7 humano y que comienza en la posición chr14:93,241,685

junto con

10 repetición homopolimérica que comprende 10 adeninas consecutivas localizadas en el gen SULF2 humano y que comienza en la posición chr20:47,657,577.

9. El método según cualquiera de las reivindicaciones 1-8, en donde la determinación del número de nucleótidos en cualquiera de las repeticiones homopoliméricas enumeradas en las reivindicaciones 1-8 se realiza además también en una muestra biológica de control que comprende material derivado de la línea celular HTC116 cl.110268743.

15

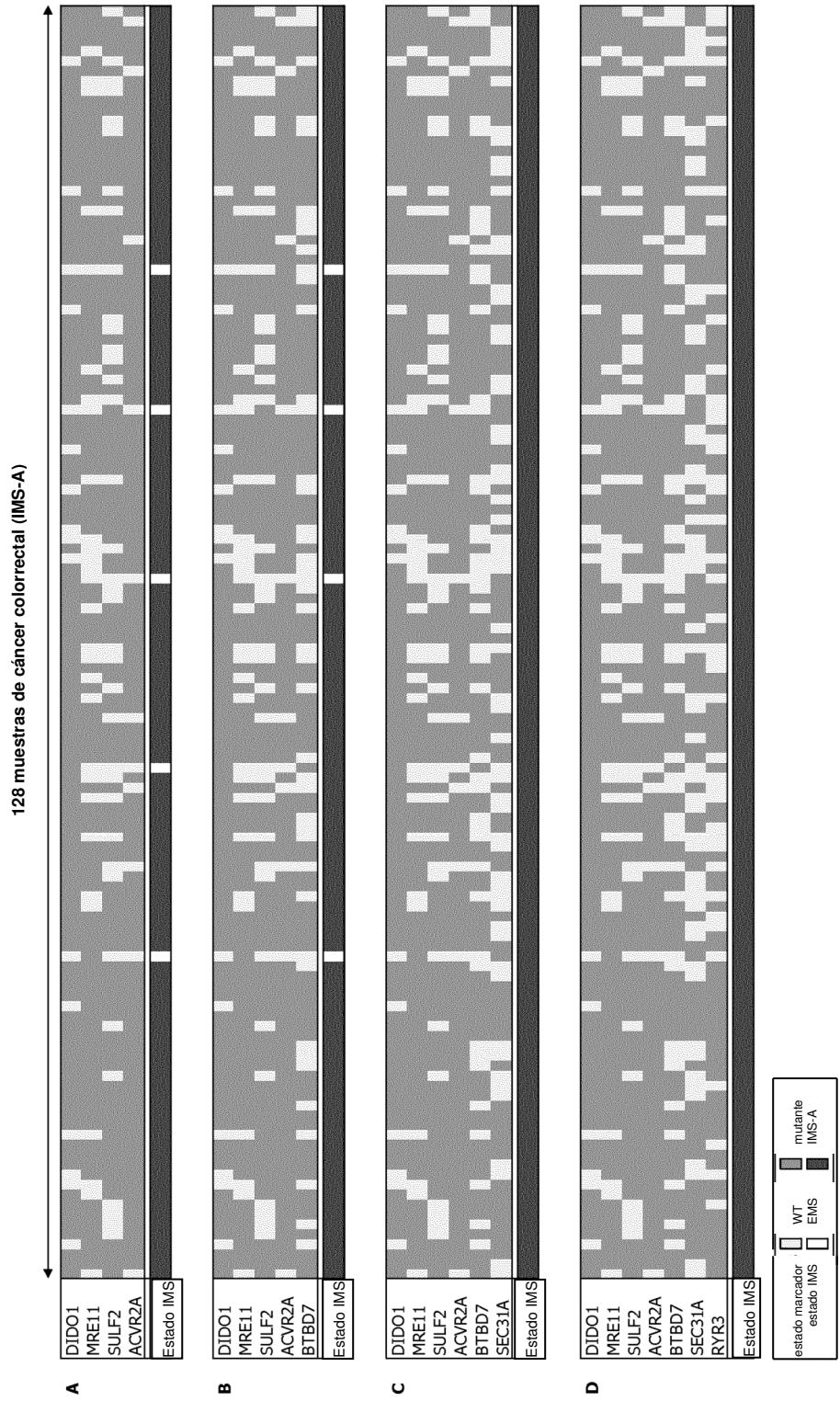


Fig. 1

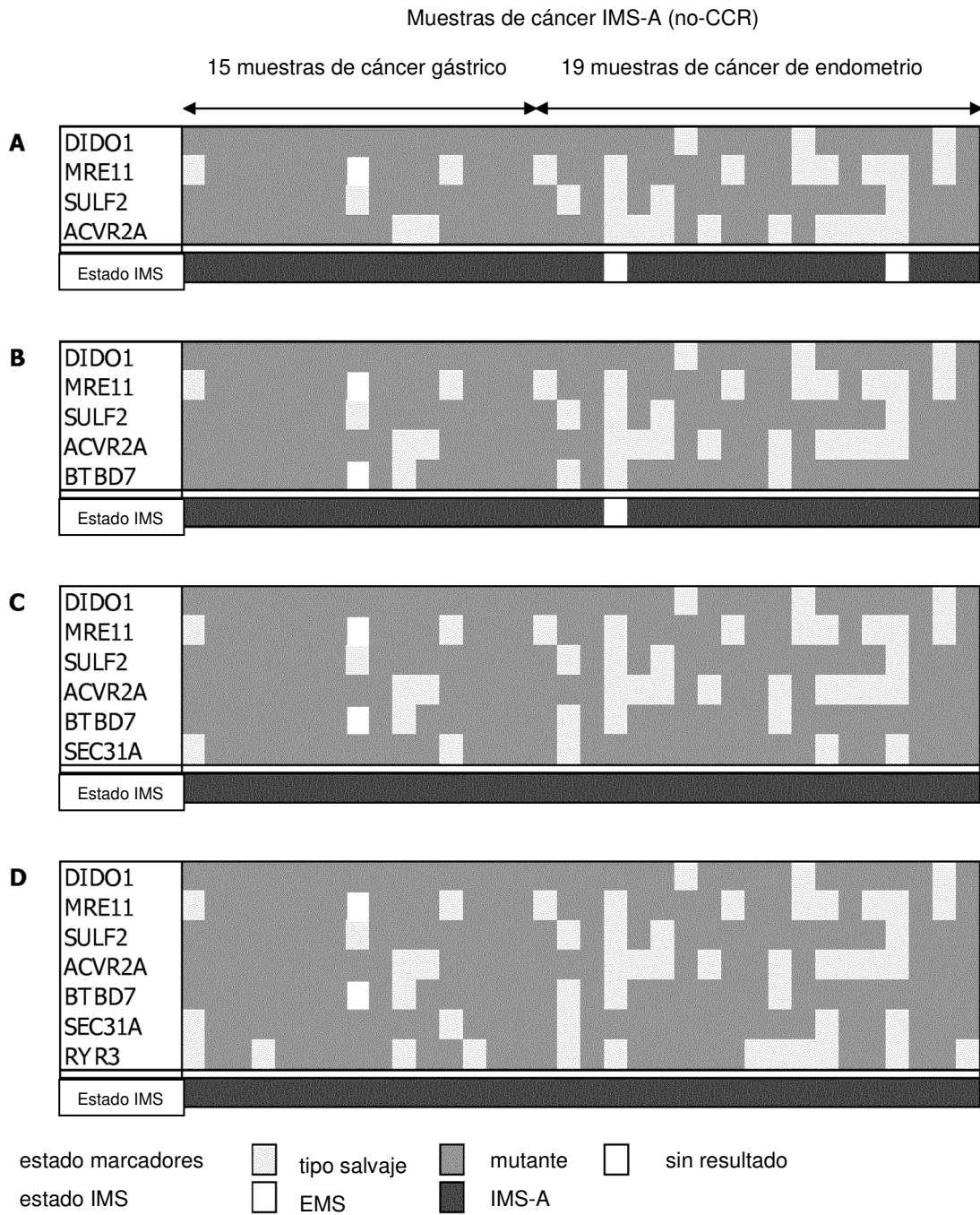


Fig. 2