

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 669 214**

51 Int. Cl.:

**C12Q 1/68** (2008.01)

**C12P 19/34** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **13.12.2012 PCT/US2012/069525**

87 Fecha y número de publicación internacional: **20.06.2013 WO13090588**

96 Fecha de presentación y número de la solicitud europea: **13.12.2012 E 12858081 (8)**

97 Fecha y número de publicación de la concesión europea: **14.02.2018 EP 2791361**

54 Título: **Procedimientos y kits para la detección de estado de metilación**

30 Prioridad:

**13.12.2011 US 201161570066 P**

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

**24.05.2018**

73 Titular/es:

**OSLO UNIVERSITETSSYKEHUS HF (100.0%)  
Postboks 450 Nydalen  
0242 Oslo, NO**

72 Inventor/es:

**DAHL, JOHN ARNE;  
ROBERTSON, ADAM BRIAN;  
KLUNGLAND, ARNE y  
ELLEVOG, LINDA**

74 Agente/Representante:

**CARPINTERO LÓPEZ, Mario**

ES 2 669 214 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

Procedimientos y kits para la detección de estado de metilación

**Campo de la invención**

5 La presente invención se refiere a procedimientos y kits para la detección de 5-hidroximetilcitosina (5hmC) y 5-metilcitosina (5meC) en ácido nucleico (p.ej., ADN, ARN). En algunas realizaciones, la presente invención se refiere a la detección de 5hmC en ADN genómico, p.ej., ADN genómico de mamífero.

**Antecedentes de la invención**

10 La modificación 5-hidroximetilcitosina (5hmC) en ADN de mamífero fue descubierta hace 30 años<sup>1</sup>. En ese momento, se señaló que la modificación 5hmC era una lesión de daño de ADN rara y no mutagénica<sup>2</sup> y, por tanto, se le dedicó poca atención. A principios de 2009, se volvió a identificar 5hmC; sin embargo, esta vez sí que se comprendió la importancia de 5hmC en epigenética, ya que dos equipos independientes comenzaron la caracterización inicial de la modificación 5hmC. Un equipo identificó una enzima capaz de catalizar la formación de 5hmC a partir de 5-metilcitosina - Tet1<sup>3</sup>. El otro equipo demostró que 5hmC era una modificación estable presente en neuronas de Purkinje especializadas<sup>4</sup>. Una posterior investigación ha demostrado que Tet1, Tet2, y Tet3 son capaces de catalizar la oxidación de 5meC creando 5hmC<sup>5-7</sup>.

15 Aún no se comprende bien la función molecular de 5hmC; sin embargo, se ha demostrado que 5hmC está relacionado con diversas transacciones de ADN; se ha demostrado que es un producto intermedio en la desmetilación de ADN<sup>3,8</sup>, que tiene una función dual en la transcripción<sup>9-11</sup> y, en el caso de patrones de 5hmC aberrantes, que está relacionado con la tumorigénesis<sup>7</sup>. Si bien aún no está del todo clara la función de la modificación 5hmC, se ha aclarado que la identificación de regiones genómicas que contienen 5hmC puede ayudar a elucidar la función de esta base. Esta necesidad de identificar regiones genómicas que contienen 5hmC ha conllevado el desarrollo de procedimientos adecuados. Actualmente, se dispone de varios procedimientos para identificar 5hmC; cada uno de dichos procedimientos presenta ciertas limitaciones, que se explican más adelante. El procedimiento descrito en el presente documento permite una resolución específica de base de (i) 5hmC y (ii) 5meC en ADN.

20 Actualmente, existen varios procedimientos que permiten la identificación de 5hmC. Dichos procedimientos incluyen anticuerpos desarrollados contra 5hmC<sup>9,21,22</sup>, anticuerpos desarrollados contra citosina 5-metilensulfonano (CMS) el producto de tratamiento de 5hmC con bisulfito<sup>7,23</sup>, secuenciación en tiempo real de molécula única basada en la cinética de ADN polimerasa<sup>24</sup>, enzimas de restricción que son resistentes o sensibles a 5hmC o  $\beta$ -glu-5hmC<sup>25-27</sup> y tres procedimientos que aprovechan la  $\beta$ -glucosiltransferasa: (i) incorporación de un marcador químico en el sustrato para  $\beta$ -gt<sup>28</sup>, (ii) procedimiento de glucosilación, oxidación de periodato y biotinilación (GLIB)<sup>23</sup> y (iii) ensayo de inmunoprecipitación de JPB1 dirigido a glu-5hmC<sup>12</sup>.

25 El uso de anticuerpos parece una opción razonable para identificar modificaciones de ADN; sin embargo, tanto los autores de la invención como otros investigadores<sup>5</sup> han observado que parte de los anticuerpos de los que se dispone actualmente dirigidos contra 5hmC parecen carecer de la capacidad para enriquecer suficientemente el ADN que contiene 5hmC; de hecho, un informe demuestra que un antisuero en particular desarrollado contra 5hmC carece de la capacidad para diferenciar 5hmC de 5meC<sup>5</sup>. Se ha notificado que los antisueros desarrollados contra 5hmC suelen presentar una preferencia por las regiones genómicas densas en contenido de 5hmC<sup>22</sup>. Además, el uso de antisueros policlonales dirigidos contra 5hmC supone un problema inherente, ya que existirá una variación de un animal a otro en la especificidad antigénica para 5hmC que afectará a la utilidad a largo plazo de dichos antisueros.

30 Tras el tratamiento con bisulfito sódico, 5hmC se convierte en CMS, que, después de la secuenciación, parece idéntico a 5meC convertido con bisulfito; por lo tanto, se ha demostrado que el uso de la secuenciación por bisulfito no puede distinguir entre 5meC y 5hmC<sup>30</sup>. Cabe destacar que un equipo ha desarrollado un antisuero dirigido contra CMS<sup>7,23</sup>.

35 La secuenciación en tiempo real de una única molécula (SMRT) aprovecha la técnica de secuenciación de Sanger original; sin embargo, este procedimiento tiene la capacidad para distinguir entre citosina, 5meC y 5hmC utilizando la firma cinética o velocidad con la que la polimerasa pasa sobre cada base<sup>24</sup>. Este procedimiento, aparte de ser prohibitivamente caro, requiere una significativa cantidad de ADN que ya está enriquecido de 5hmC antes de su uso, lo que hace que dependa del ensayo de enriquecimiento de 5hmC. Dado que dicho procedimiento emplea una secuenciación de alto rendimiento, resulta complicado para el análisis de un solo locus o unos cuantos loci.

40 Varios equipos y empresas de investigación han identificado enzimas de restricción que son sensibles o resistentes a 5hmC o  $\beta$ -glu-5hmC<sup>25-27</sup>. El principio que se esconde tras estos sistemas es que, tras el tratamiento con enzimas de restricción, se escinde ADN sin modificar, con el resultado de una reducción de señal en una reacción de qPCR. Dicha reducción de señal se compara a continuación con una muestra sin digerir y la diferencia en las señales de qPCR es proporcional a la cantidad de 5hmC presente en la muestra inicial. Estos procedimientos funcionan perfectamente para regiones genómicas que contienen cantidades significativas de 5hmC; sin embargo, como los

sitios de restricción reconocidos por estas enzimas son de 4-6 pb de longitud, estos procedimientos a base de endonucleasa de restricción, en el mejor de los casos, tan solo pueden reconocer un 1/16 de todas las modificaciones de 5hmC. La patente estadounidense US 2006/00257 desvela un procedimiento para copiar patrones de metilación de ADN genómico (MGD) durante la amplificación isotérmica del MGD. Dicho procedimiento comprende la obtención de MGD, la copia de los patrones de metilación del MGD utilizando una enzima de mantenimiento de la metilación de ADN, al mismo tiempo que se amplifica isotérmicamente MGD utilizando ADN polimerasa con actividad de desplazamiento de cadena en condiciones que promueven simultáneamente la actividad de la enzima de metilación de ADN y la ADN polimerasa con actividad de desplazamiento de cadena y, opcionalmente, la purificación de MGD, según sea necesario para permitir una posterior manipulación.

5 Tres equipos han desarrollado procedimientos que aprovechan la especificidad que tiene  $\beta$ -gt para 5hmC. El primer equipo<sup>28</sup> incorporó un grupo azida en el sustrato para  $\beta$ -gt - UDP-glucosa - creando UDP-6-N<sub>3</sub>-Glucosa. Una vez incorporada la glucosa modificada con azida en ADN que contenía 5hmC mediante  $\beta$ -gt, fue posible añadir un segundo grupo a 6-N<sub>3</sub>-glu-5hmC empleando química "clic". Este segundo grupo químico podría contener una biotina para inmunoprecipitación, una sonda fluorescente para cuantificación y, teóricamente, cualquier grupo que se pudiera acoplar a la glucosa modificada empleando química "clic". El principal inconveniente de este procedimiento es que UDP-6-N<sub>3</sub>-glucosa no se produce a nivel comercial y requiere una significativa experiencia en química orgánica para su síntesis. Por otra parte, se ha combinado esta estrategia dirigida de 5hmC con un ensayo de extensión de cebador y se ha demostrado que permite una resolución específica de base ya que un grupo químico puede unirse con 6-N<sub>3</sub>-glu-5hmC que boquea una ADN polimerasa. Al bloquear la polimerasa, se puede dar por sentado que la base terminal contenga originalmente una modificación 5hmC. El uso de este procedimiento para la resolución específica de base presenta sustanciales problemas, ya que debe dar por sentado que todos los extremos que acaban en C deben ser 5hmC. Si bien es posible promediar potencialmente este efecto con varias lecturas de secuenciación de alto rendimiento dando por sentado relaciones entre enzima y ADN altamente optimizadas, sigue siendo problemático para un análisis de un único gen.

10 Un segundo enfoque en el que se utiliza  $\beta$ -gt para identificar regiones genómicas emplea el procedimiento glucosilación, oxidación de periodato, biotilación (GLIB)<sup>23</sup>. En este procedimiento, después de transferir glucosa a 5hmC se oxida  $\beta$ -glu-5hmC resultante utilizando NaIO<sub>4</sub> que crea aldehídos reactivos en la fracción de glucosa unida a 5hmC. A continuación, es posible hacer reaccionar estas moléculas de glucosa oxidadas con sondas reactivas con aldehído que contienen modificación de biotina disponibles en el mercado. Dicha biotilación permite una inmunoprecipitación eficiente de ADN que contiene 5hmC.

15 Finalmente, el tercer enfoque en el que se utiliza  $\beta$ -gt para la identificación de 5hmC implica el reconocimiento específico de esta base modificada con una segunda proteína – proteína de unión a base J o JPB1. Dado que la única diferencia entre  $\beta$ -glucosil-5hmC y la base J es un grupo amino, se dedujo que JPB1 podía ser capaz de interactuar específicamente con  $\beta$ -glu-5hmC. De hecho, JPB1 fue capaz de interactuar específicamente con  $\beta$ -glu-5hmC<sup>12</sup>. Por lo tanto, cuando se unió covalentemente JPB1 con perlas magnéticas modificadas con epoxi, fue posible la inmunoprecipitación de ADN que contenía  $\beta$ -glu-5hmC. Una vez extraída la proteína del ADN inmunoprecipitado, se demostró por qPCR específica de gen que era posible enriquecer el ADN que contenía 5hmC<sup>12</sup>. Mecánicamente, este procedimiento proporciona dos grados de especificidad para la identificación de 5hmC en ADN genómico: en primer lugar,  $\beta$ -gt puede modificar únicamente citosinas en el ADN que están hidroximetiladas y, en segundo lugar, JPB1 interactúa específicamente con  $\beta$ -glu-5hmC. Al igual que todos los procedimientos de inmunoprecipitación de ADN, la resolución óptima en gran medida de este procedimiento puede servir para identificar una base 5hmC dentro de aproximadamente 50-100 pares de base; dicha limitación se debe a la incapacidad para identificar de forma fiable fragmentos de ADN de una longitud más corta aplicando los procedimientos de biología molecular disponibles actualmente. Otra consideración a la hora de utilizar este protocolo es que este procedimiento puede sobre-representar regiones de ADN que contienen altos niveles de 5hmC. Dicha posible sobre-representación podría tener lugar posiblemente porque en las regiones densas de 5hmC puede interactuar más JPB1 con ADN e inmunoprecipitar estas regiones más eficientemente.

20 Se necesitan procedimientos mejorados para detectar restos 5-hidroximetilcitosina en ADN. En particular, se necesitan procedimientos que puedan discriminar entre 5meC y 5hmC, así como procedimientos que puedan identificar 5meC y 5hmC en resolución de una sola base.

### **Sumario de la invención**

La presente invención se refiere a procedimientos y kits para la detección de 5-hidroximetilcitosina (5hmC) y 5-metilcitosina (5meC) en ácido nucleico (p.ej., ADN, ARN). En algunas realizaciones, la presente invención se refiere a la detección de 5hmC en ADN genómico, p.ej., ADN genómico de mamífero.

25 En algunas realizaciones, la presente invención proporciona procedimientos para detectar restos de citosina 5-metilada e hidroximetilada en una muestra de ácido nucleico que comprenden: replicación de dicha muestra de ácido nucleico en condiciones tales que se mantengan dichos restos de citosina 5-metilada y se diluyan dichos restos de citosina hidroximetilada; tratamiento de dicha muestra de ácido nucleico replicada para convertir los restos de citosina sin modificar en restos de uracilo o timidina; en los que se divide la muestra de ácido nucleico en al menos una primera y una segunda porción y se realizan dichas etapas de replicación y tratamiento sobre dicha

primera porción, y comparación de la secuencia de dicha primera porción de ácido nucleico con la secuencia de dicha segunda porción de ácido nucleico, en los que dichos restos de citosina hidroximetilada se identifican como restos que se leen por secuenciación como un resto de uracilo o timidina en dicha primera porción de ácido nucleico y como un resto citosina en la correspondiente posición en dicha segunda porción de ácido nucleico y en los que los restos de citosina 5-metilada se identifican como restos que se leen como restos de citosina en ambas porciones de ácido nucleico primera y segunda. En algunas realizaciones, la replicación de dicha primera porción comprende además: a) replicación de dicho ácido nucleico con un cebador marcado para proporcionar un ácido nucleico replicado marcado; b) tratamiento de dichas cadenas de ácido nucleico replicadas marcadas con una ADN metil transferasa para proporcionar ácido nucleico replicado modificado con 5-metilcitosina marcado; c) aislamiento de dicho ácido nucleico replicado modificado con 5-metilcitosina marcado; d) tratamiento de dicho ácido nucleico replicado modificado con 5-metilcitosina marcado aislado con bisulfito para convertir restos de citosina sin modificar en restos de uracilo; y e) replicación de dicho ácido nucleico tratado con bisulfito marcado aislado con una polimerasa para proporcionar una primera porción de ácido nucleico tratada con bisulfito. En algunas realizaciones, el cebador marcado es un cebador biotinilado. En algunas realizaciones, los demás restos de citosina modificados se seleccionan del grupo que consiste en 5-hidroximetil citosina y beta-glu-5-hidroximetil citosina. Otros ejemplos de restos de citosina modificados son alfa-glucosil-5-hidroximetilcitosina, beta-glucopiranosil-alfa-glicosil-5-hidroximetilcitosina (gentiobiosil-5-hidroximetilcitosina), 5-formilcitosina y 5-carboxicitosina.

En algunas realizaciones, la replicación de dicha primera porción en condiciones tales que se mantienen los restos de citosina 5-metilada y se diluyen los restos de citosina 5-hidroximetilada comprende la replicación de dicho ácido nucleico con una polimerasa para proporcionar ácido nucleico replicado y el tratamiento de dicho ácido nucleico replicado con una enzima para dar restos de citosina 5-metilada. En algunas realizaciones, se realizan una o más veces las etapas de replicación y tratamiento con una enzima. En algunas realizaciones, las etapas de replicación y tratamiento con una enzima se repiten 5 veces o más. En algunas realizaciones, las etapas de replicación y tratamiento con una enzima se repiten 7 veces o más. En algunas realizaciones, las etapas de replicación y tratamiento con una enzima se repiten 10 veces o más. En algunas realizaciones, las etapas de replicación y tratamiento con una enzima se realizan de aproximadamente 1 a aproximadamente 20 veces o más. En algunas realizaciones, la replicación es por reacción en cadena de la polimerasa. En algunas realizaciones, la replicación es por reacción de extensión de cebador. En algunas realizaciones, la enzima es una ADN metil transferasa. En algunas realizaciones, la ADN metil transferasa es DNMT1. En algunas realizaciones, la ADN metil transferasa es M.Sssl.

En algunas realizaciones, el tratamiento de dichas porciones primera y segunda para convertir restos de citosina sin modificar en restos de timidina comprende además el tratamiento de dichas porciones de ácido nucleico primera y segunda con bisulfito para convertir restos de citosina sin modificar en restos de uracilo y la replicación de dichas porciones de ácido nucleico primera y segunda con una polimerasa para convertir dichos restos de uracilo en restos timidina. En algunas realizaciones, la replicación se realiza 1 o más veces. En algunas realizaciones, la replicación se realiza 5 veces o más. En algunas realizaciones, la replicación se realiza 7 veces o más. En algunas realizaciones, la replicación se realiza 10 veces o más. En algunas realizaciones, la replicación se repite de aproximadamente 1 a aproximadamente 20 veces. En algunas realizaciones, la replicación es por reacción en cadena de la polimerasa. En algunas realizaciones, la replicación es por reacción de extensión de cebador.

En algunas realizaciones, la muestra de ácido nucleico se selecciona del grupo que consiste en una muestra de ácido nucleico de ser humano, planta, ratón, conejo, hámster, primate, pez, ave, vaca, oveja, cerdo, viral, bacteriano y fúngico.

En algunas realizaciones, los procedimientos comprenden además la comparación de la presencia de 5-hidroximetilcitosina y/o 5-metilcitosina en dicho ácido nucleico de dicha muestra con respecto a un nivel referencia, en la que un aumento o descenso del nivel de 5-hidroximetilcitosina y/o 5-metilcitosina en dicho ácido nucleico es indicativo de la presencia de una enfermedad o del probable curso de una enfermedad. En algunas realizaciones, los procedimientos comprenden además la etapa de proporcionar un diagnóstico o pronóstico sobre la base del aumento o descenso del nivel de 5-hidroximetilcitosina y/o 5-metilcitosina en dicho ácido nucleico en comparación con un nivel de referencia. En algunas realizaciones, la enfermedad es cáncer. En algunas realizaciones, la muestra de ácido nucleico es ADN genómico.

En algunas realizaciones, la presente invención proporciona procedimientos para detectar restos de citosina metilada e hidroximetilada en una muestra de ácido nucleico que comprende: a) dividir dicha muestra en al menos una primera y una segunda porción sin tratar; b) replicación de dicha primera porción con un cebador marcado y una polimerasa para proporcionar ácido nucleico parental y replicado marcado; c) tratamiento de dichas cadenas de ácido nucleico parental y replicado marcado con una ADN metil transferasa para proporcionar ácido nucleico replicado modificado con 5-metilcitosina marcado; d) aislamiento de dicho ácido nucleico replicado modificado con 5-metilcitosina marcado; e) tratamiento de dicho ácido nucleico replicado modificado con 5-metilcitosina marcado aislado con bisulfito para convertir restos de citosina sin modificar en restos de uracilo; f) replicación de dicho ácido nucleico tratado con bisulfito marcado aislado con una polimerasa para proporcionar una primera porción de ácido nucleico tratada con bisulfito; g) secuenciación de dicha primera porción de ácido nucleico tratada con bisulfito; h) tratamiento de dicha segunda porción de ácido nucleico con bisulfito para convertir restos de citosina sin modificar en restos de uracilo; i) replicación de dicho ácido nucleico tratado con bisulfito con una polimerasa para proporcionar

una segunda porción de ácido nucleico tratada con bisulfito; j) secuenciación de dicha segunda porción de ácido nucleico tratada con bisulfito; y k) comparación de la secuencia de dicha primera porción de ácido nucleico tratada con bisulfito con la secuencia de dicha segunda porción tratada con bisulfito, en la que se identifican los restos de citosina 5-hidroximetilada como restos que se leen por secuenciación como un resto de uracilo o timidina en dicha primera porción de ácido nucleico tratada con bisulfito y como un resto de citosina en la posición correspondiente en dicha segunda porción de ácido nucleico tratada con bisulfito y en la que se identifican los restos de citosina 5-metilada como restos que se leen como restos de citosina en dichas porciones tratadas con bisulfito primera y segunda. En algunas realizaciones, se replica dicha segunda porción con una polimerasa antes de dicha etapa de secuenciación. En algunas realizaciones, dichas etapas b, c y d se repiten de aproximadamente 2 a aproximadamente 20 veces. En algunas realizaciones, dichas etapas e y h se repiten de aproximadamente 2 a aproximadamente 20 veces. En algunas realizaciones, dicha replicación de las etapas b, e y h es por reacción en cadena de la polimerasa. En algunas realizaciones, los procedimientos comprenden además la comparación de la presencia de 5-hidroximetilcitosina y/o 5-metilcitosina en dicho ácido nucleico de dicha muestra con un nivel de referencia, en la que el aumento o descenso del nivel de 5-hidroximetilcitosina y/o 5-metilcitosina en dicho ácido nucleico es indicativo de la presencia de una enfermedad o del probable curso de una enfermedad. En algunas realizaciones, los procedimientos comprenden además la etapa de proporcionar un diagnóstico o pronóstico sobre la base del aumento o descenso del nivel de 5-hidroximetilcitosina y/o 5-metilcitosina en dicho ácido nucleico en comparación con un nivel de referencia. En algunas realizaciones, la enfermedad es cáncer. En algunas realizaciones, la muestra de ácido nucleico es ADN genómico.

En algunas realizaciones, la presente invención proporciona un procedimiento para predecir una predisposición a una enfermedad en un sujeto, diagnosticar una enfermedad en un sujeto, predecir la probabilidad de la recurrencia de una enfermedad en un sujeto, proporcionar un pronóstico para un sujeto con una enfermedad, o seleccionar un sujeto con una enfermedad para su tratamiento con una terapia en particular, que comprende: a) proporcionar una muestra de ADN genómico de dicho sujeto; y b) detectar el estado de metilación de porciones predeterminadas de dicha muestra de ADN genómico a través de los procedimientos que se han descrito, en los que una alteración del nivel de metilación de 5-hidroximetilcitosina y/o 5-metilcitosina de dichas porciones predeterminadas de dicho ADN genómico con respecto a un estado de metilación de referencia proporciona una indicación seleccionada del grupo que consiste en una indicación de predisposición del sujeto a una enfermedad, una indicación de que el sujeto tiene una enfermedad, una indicación de la probabilidad de recurrencia de una enfermedad en el sujeto, una indicación de supervivencia del sujeto y una indicación de que el sujeto es candidato para tratamiento con una terapia en particular. En algunas realizaciones, la enfermedad es un cáncer. En algunas realizaciones, el sujeto es un ser humano.

En algunas realizaciones, la presente invención proporciona un kit para determinar el estado de metilación de una muestra de ácido nucleico que comprende: 1) recipiente(s) con reactivos para metilar ácido nucleico; y 2) recipiente(s) con reactivos para secuenciación por bisulfito y 3) un soporte informático que comprende un programa informático que analiza los datos de secuencia obtenidos con el kit de acuerdo con los procedimientos de la invención. En algunas realizaciones, los kits comprenden además cebadores de ácido nucleico para la amplificación y/o secuenciación de una región de dicha muestra de ácido nucleico.

Otras realizaciones se pondrán de manifiesto para las personas expertas en la materia correspondiente sobre la base de las instrucciones contenidas en el presente documento.

### **Descripción de los dibujos**

Figura 1. Representación esquemática de determinadas realizaciones de la presente invención, en las que se aplica la conversión con bisulfito y la secuenciación de "A" ADN sin tratar que se utilizará como referencia para detectar el total de 5meC y 5hmC. El procedimiento implica un ensayo de dilución de 5hmC, la dilución de 5hmC en la agrupación total de fragmentos de ADN al mismo tiempo que se mantiene 5meC. Dicha dilución se consigue a través de rondas secuenciales de un ciclo de amplificación por PCR (dilución) y tratamiento del ADN con la ADN metil transferasa de mantenimiento DNMT1 que mantiene enzimática y específicamente 5meC añadiendo un grupo metilo únicamente a la cadena no metilada de los productos de PCR hemimetilados (en la Figura 1 se hace referencia a esta muestra como "B"). Tras unas cuantas rondas de este ensayo, se aplicó la conversión y secuenciación con bisulfito de la muestra de ADN tratada, B. Las bases que se leen como citosina de esta muestra han de ser protegidas contra la conversión con bisulfito por 5meC y no por 5hmC. Al comparar "B" con la muestra de referencia "A" es posible detectar fácilmente todas las posiciones de base que contienen 5hmC.

Figura 2. La conversión de ADN con bisulfito tiene como resultado la conversión de citosina sin modificar (C) en uracilo (U) que se leerá como timina (T) tras la secuenciación del ADN amplificado por PCR. Tanto 5meC como 5hmC están protegidos contra la conversión y no se convertirán en U. Por tanto, ambas bases se leerán como C tras la secuenciación. La conversión con bisulfito es una tecnología perfectamente establecida considerada desde hace tiempo como el criterio de referencia para la detección de 5meC y tan solo recientemente (2010) se ha notificado en la bibliografía científica que la conversión con bisulfito no puede distinguir entre 5meC y 5hmC.

Figura 3. DNMT1 de ratón, DNMT1 humana y M. Sssl metilan preferentemente ADN hemi-5meC. Se incubaron 100 ng de cada sustrato de ADN con 2 unidades de DNMT1 de ratón, DNMT1 humano o metil transferasa de Sssl tal como se describe en el epígrafe “materiales y procedimientos”.

Figura 4. Validación de la viabilidad del ensayo de dilución de 5hmC. (A) El oligo de ADN de doble cadena utilizado en la validación contiene tres sitios CpG en los que uno es hemi-5meC, un segundo no tiene modificación y un tercero es hemi-5hmC. (B) La conversión y secuenciación con bisulfito de la cadena inferior sin modificar del oligo en (A), cuando el oligo no ha sido sometido a tratamiento con DNMT1, demostró que todas las C se habían convertido y se leyeron como T. (100 % de T es igual a 16 de los 16 clones individuales que se leen como T en la posición C del sitio CpG). (C) El tratamiento con DNMT1 antes de la conversión y secuenciación con bisulfito tuvo como resultado la adición de un grupo metilo a C sin metilar del sitio CpG hemi-5meC en un 87,5 % de los oligos. (La secuenciación dio una lectura de un C en la posición C del sitio CpG en 14 de los 16 clones). No se observó la adición de un grupo metilo transversalmente desde C o 5hmC.

Figura 5. Representación esquemática del procedimiento para una identificación diferenciada de 5hmC y 5meC en la resolución específica de base. (A) Un esquema de seguimiento de las bases C de los sitios CpG de un oligo ADNbc que contiene tres sitios CpG en los que uno tiene 5meC en ambas cadenas, un segundo no tiene modificación y un tercero tiene 5hmC en ambas cadenas. Se hace un seguimiento de los sitios CpG a través de una ronda de PCR (fusión, hibridación de cebador y elongación) y tratamiento con DNMT1 antes de la visualización del tratamiento con bisulfito y PCR (30 ciclos) que genera las bases que se leerán en la secuenciación. (B) Diagrama de flujo del procedimiento experimental implicado en el ensayo de dilución 5hmC.

Figura 6. Mantenimiento preferente de 5meC con respecto a 5hmC. El oligo de ADN bicatenario utilizado en este caso contiene tres sitios CpG, uno que tiene 5meC en ambas cadenas, un segundo que no tiene modificación y un tercero que tiene 5hmC en ambas cadenas. (A) La conversión y secuenciación con bisulfito del oligo sin tratar mostró que solamente estaban protegidas frente a la conversión las C modificadas (100 % tanto para 5meC como 5hmC), mientras que las citosinas sin modificar se convirtieron todas. (B) Tomando el oligo de ADN de doble cadena a través de las tres rondas del ensayo de dilución, que implicó PCR y tratamiento con DNMT1, antes de la conversión y secuenciación con bisulfito, el resultado fue un mantenimiento preferente de 5meC con respecto a 5hmC. No hubo metilación transversal desde 5hmC en ninguna de las tres rondas ya que las cadenas modificadas con 5hmC iniciales componían únicamente el 9 % de la agrupación total tras las tres rondas del ensayo de dilución. (Cabría esperar un 50 % tras una ronda, un 25 % tras dos rondas y 12,5 % tras tres rondas cuando no hay mantenimiento en absoluto). La base 5meC se mantuvo preferentemente, con el resultado de un mayor número de C protegidas en la conversión con bisulfito y una lectura significativamente superior que la base 5hmC. No se observó ninguna adición de grupo metilo transversalmente ni desde C ni desde 5hmC.

Figura 7. Representación esquemática del procedimiento para la identificación diferenciada de 5hmC y 5meC en la resolución específica de base con el uso de la evaluación específica de cadena. (A) Un esquema de seguimiento de las bases C de los sitios CpG de un oligo ADNbc que contiene tres sitios CpG, en los que uno tiene 5meC en ambas cadenas, un segundo no tiene modificación y un tercero tiene 5hmC en ambas cadenas. Se lleva un seguimiento de los sitios CpG en una ronda de PCR de extensión de cebador específico de cadena (fundido, hibridación de cebador y elongación) y tratamiento con DNMT1. El cebador utilizado puede contener marcador biotina, u otro marcador, para permitir la selección/aislamiento de la cadena recién sintetizada. La cadena recién sintetizada pasa por tratamiento con bisulfito y PCR (30 ciclos u otro número) que genera las bases que se leerán en la secuenciación. (B) Diagrama de flujo del procedimiento experimental en relación con el ensayo de dilución/pérdida de 5hmC aplicando extensión de cebador y evaluación específica de cadena.

Figura 8. Secuencia de aminoácidos para DNMT1 (*Mus musculus*) Recombinante. Número de referencia: GenBank: AAH53047.1 (SEQ ID NO: 1).

Figura 9. Secuencia de aminoácidos para DNMT1 (*Homo sapiens*) Número de referencia: GenBank: AAI44094.1 (SEQ ID NO: 2).

Figura 10. Secuencia de aminoácidos para M.Sssl (*Spiroplasma* sp. (cepa MQ1)) ADN metil transferasa específica de sitio (SEQ ID NO: 3).

Figura 11. Representación esquemática de un ensayo de pérdida de 5hmC de la presente invención utilizando cebadores biotinilados y perlas de captura de estreptavidina. El panel de la derecha, parte superior, presenta los resultados de la secuenciación representativos de 10 clones para el ensayo de bisulfito convencional, con la referencia A, en el que se leen tanto 5meC como 5hmC como citosina tras el tratamiento y los resultados de la secuenciación de 10 clones para el ensayo de transferencia de metilo/ensayo de pérdida de 5hmC, con la referencia B, en el que se leerá únicamente 5meC como citosina tras el tratamiento. Las citosinas en un contexto de secuencia CG (CpG) protegidas de la conversión con bisulfito se ilustran como círculos negros, mientras que las citosinas en un contexto de secuencia CG que experimentan desaminación en uracilo en el tratamiento con bisulfito se ilustran con círculos en blanco. La combinación de los datos del ensayo de bisulfito de referencia, A, en el que tanto 5meC como 5hmC se leen como citosina tras el tratamiento y el ensayo de transferencia de metilo, B, en el que se lee 5meC como citosina tras el tratamiento permite determinar la posición y cantidad de

5hmC, a partir de un sencillo cálculo:  $A-B = 5hmC$ . Esta cuantificación se describe en la parte inferior del panel de la derecha. Estos resultados experimentales se han reproducido en 15 experimentos independientes.

5 Figura 12. Esquema y gráfico en el que se presenta la identificación de dos 5hmC que contienen islas CpG, que es 5hmC en una secuencia CG, en el gen TRIM31 de ADN de cerebro humano aplicando el ensayo representado en la Figura 11. Las posiciones de las CpG se representan esquemáticamente (no a escala) y la cantidad de 5hmC y 5meC en esas posiciones para citosina se dan en la gráfica de barras.

Figura 13. Gráfico de barras en el que se muestran los resultados de un experimento en el que se bloquea metil transferasa mediante la adición de un grupo químico a 5hmC.

### **Definiciones**

10 A fin de facilitar la comprensión de la presente invención, se definen a continuación una serie de términos y expresiones:

Tal como se utiliza en el presente documento, el término "sensibilidad" se define como una medición estadística de rendimiento de un ensayo (p.ej., procedimiento, análisis), que se calcula dividiendo el número de verdaderos positivos por la suma de los verdaderos positivos y los falsos negativos.

15 Tal como se utiliza en el presente documento, el término "especificidad" se define como una medición estadística de rendimiento de un ensayo (p.ej., procedimiento, análisis), que se calcula dividiendo el número de verdaderos negativos por la suma de verdaderos negativos y falsos positivos.

20 Tal como se utiliza en el presente documento, el término "informativo" o la expresión "capacidad informativa" se refiere a una cualidad de un marcador o panel de marcadores y, concretamente, a la probabilidad de encontrar un marcador (p.ej., marcador epigenético; p.ej., 5hmC, en una o más localizaciones en particular) en una muestra positiva.

Tal como se utiliza en el presente documento, el término "dilución" se refiere a la reducción de restos de citosina modificados no 5-metilo (p.ej., restos 5-hidroximetil citosina) en una muestra de ácido nucleico en comparación con los restos citosina 5-metilo a través de repetidas rondas de replicación de dicha muestra de ADN.

25 Tal como se utiliza en el presente documento la expresión "restos de citosina modificada no 5-metil citosina" se refiere a restos de citosina modificada distintos a 5-metil citosina, por ejemplo, 5-hidroximetil citosina, b-glu-5-hidroximetil citosina, 5-formil-citosina y 5-carboxicitosina.

30 Tal como se utiliza en el presente documento, el término "Isla CpG" se refiere a una región de ADN genómico que contiene un alto porcentaje de sitios CpG en relación con la incidencia de CpG genómica promedio (para la misma especie, para el mismo individuo o por subpoblación (p.ej., cepa, subpoblación étnica o similares). Existen varios parámetros y definiciones para islas CpG; por ejemplo, en algunas realizaciones, islas CpG se definen por tener un porcentaje de GC por encima de 50 % y con una relación CpG observada/esperada por encima de 60 % (Gardiner-Garden y col. (1987) J Mol. Biol. 196:261-282; Baylin y col. (2006) Nat. Rev. Cáncer 6:107-116; Irizarry y col. (2009) Nat. Genetics 41:178-186; que se incorporan como referencia en su totalidad). En algunas realizaciones, las islas CpG puede tener un contenido de GC >55 % y una relación CpG observado/ CpG esperado de 0,65 (Takai y col. (2007) PNAS 99:3740-3745). Asimismo, existen varios parámetros en lo que se refiere a la longitud de las islas CpG. Tal como se utiliza en el presente documento, las islas CpG pueden ser de menos de 100 pb; 100-200 pb, 200-300 pb, 300-500 pb, 500-750 pb; 750-1000 pb; 1000 o más pb de longitud. En algunas realizaciones, las islas CpG presentan patrones de alteración de metilación (p.ej., patrones de alteración de 5hmC) en relación con los controles (p.ej., alteración de metilación 5hmC en sujetos con cáncer con respecto a los sujetos sin cáncer; patrones de alteración 5hmC específicos de tejido; patrones de alteración de 5hmC en muestras biológicas de sujetos con neoplasia o tumor con respecto a sujetos sin neoplasia ni tumor. En algunas realizaciones, la alteración de la metilación implica un aumento de la incidencia de 5hmC. En algunas realizaciones, la alteración de la metilación implica un descenso de la incidencia de 5hmC.

45 Tal como se utiliza en el presente documento, las expresiones "orilla CpG " u "orilla de Isla CpG " se refieren a una región genómica externa a una Isla CpG que tiene o tiene potencial para tener patrones de alteración de metilación (p.ej., 5hmC) (véase, p.ej., Irizarry y col. (2009) Nat. Genetics 41:178-186). Las orillas de isla CpG pueden presentar también patrones de alteración (p.ej., 5hmC) con respecto a los controles (p.ej., alteración de 5hmC en sujetos con cáncer con respecto a los sujetos sin cáncer; patrones de alteración de 5hmC específicos de tejido; alteración de 5hmC en muestras biológicas de sujetos con neoplasia o tumor con respecto a sujetos sin neoplasia ni tumor. En algunas realizaciones, alteración de la metilación implica un aumento o descenso de 5hmC. En algunas realizaciones, alteración de la metilación implica una menor incidencia de 5hmC. Las orillas de isla CpG pueden estar localizadas en diversas regiones en relación con las islas CpG (véase, p.ej., Irizarry y col. (2009) Nat. Genetics 41; 178-186). En consecuencia, en algunas realizaciones, las orillas de isla CpG están localizadas a una distancia de menos de 100 pb; 100-250 pb; 250-500 pb; 500-1000 pb; 1000-1500 pb; 1500-2000 pb; 2000-3000 pb; 3000 pb o más de una Isla CpG.

Tal como se utiliza en el presente documento, el término "metástasis" sirve para referirse a un proceso en el que las células cancerosas originadas en un órgano o una parte del cuerpo se relocalizan en otra parte del cuerpo y continúan replicándose. Las células metastásicas forman posteriormente tumores que pueden seguir metastatizando. La metástasis se refiere pues a la propagación del cáncer desde la parte del cuerpo en la que se produce originalmente a otras partes del cuerpo.

Tal como se utiliza en el presente documento, "se sospecha que un individuo es susceptible de metastatizar cáncer" sirve para referirse a un individuo que está en riesgo por encima de la media de desarrollar un cáncer metastatizado. Entre los ejemplos de individuos en particular riesgo de desarrollar cáncer de un tipo en particular (p.ej., cáncer colorrectal, cáncer de vesícula, cáncer de mama, cáncer de próstata) se incluyen aquellos cuya historia médica familiar indica una incidencia por encima de la media de dicho tipo de cáncer entre los miembros de la familia y/o aquellos que han desarrollado ya el cáncer y han sido tratados eficazmente, que por tanto se enfrentan al riesgo de recaída o recurrencia. Otros factores que pueden contribuir a un riesgo por encima de la media de desarrollar cáncer metastatizado que llevan a clasificar a un individuo como sospechoso de ser susceptible de cáncer metastatizado pueden basarse en las características y antecedentes genéticos, médicos y/o conductuales específicos del individuo.

El término "neoplasma" tal como se utiliza en el presente documento se refiere a cualquier crecimiento anormal y nuevo del tejido. Por tanto, un neoplasma puede ser un neoplasma premaligno o un neoplasma maligno. La expresión "marcador específico de neoplasma" se refiere a cualquier material biológico que se pueda utilizar para indicar la presencia de un neoplasma. Entre los ejemplos de materiales biológicos se incluyen, sin limitación, ácido nucleicos, polipéptidos, hidratos de carbono, ácidos grasos, componentes celulares (p.ej., membranas celulares y mitocondrias) y células enteras.

Tal como se utiliza en el presente documento, el término "amplicón" se refiere a un ácido nucleico generado mediante el uso de pares de cebador. El amplicón es normalmente ADN monocatenario (p.ej., el resultado de amplificación asimétrica), si bien, puede ser ARN o ADNbc.

El término "amplificar" o "amplificación" en el contexto de ácido nucleicos se refiere a la producción de múltiples copias de un polinucleótido o una porción del polinucleótido, normalmente partiendo de una pequeña cantidad del polinucleótido (p.ej., una única molécula de polinucleótido), en la que son detectables generalmente los productos de amplificación o amplicones. La amplificación de polinucleótidos abarca diversos procesos químicos y enzimáticos. La generación de múltiples copias de ADN a partir de una o unas cuantas copias de una molécula de ADN diana o matriz durante una reacción en cadena de la polimerasa (PCR) o una reacción en cadena de la ligasa (LCR; véase, p.ej., patente estadounidense No. 5.494.810) son formas de amplificación. Otros tipos de amplificación incluyen, pero sin limitarse a ellos, PCR específica de alelo (véase, p.ej., patente estadounidense No. 5.639.611), PCR de ensamblaje (véase, p.ej., patente estadounidense No. 5.965.408), amplificación dependiente de helicasa (véase, p.ej., patente estadounidense No. 7.662.594), amplificación con arranque en caliente (véase, p.ej., patentes estadounidenses Nos. 5.773.258 y 5.338.671), PCR específica de intersecuencia, PCR inversa (véase, p.ej., Triglia, y col. (1988) *Nucleic Acids Res.*, 16:8186), PCR mediada por ligación (véase, p.ej., Guilfoyle, R. y col., *Nucleic Acids Research*, 25:1854-1858 (1997); patente estadounidense No. 5.508.169), PCR específica de metilación (véase, p.ej., Herman, y col., (1996) *PNAS* 93(13) 9821-9826), PCR minicebador, amplificación de sonda dependiente de ligación multiplex (véase, p.ej., Schouten, y col., (2002) *Nucleic Acids Research* 30(12): e57), PCR multiplex (véase, p.ej., Chamberlain, y col., (1988) *Nucleic Acids Research* 16(23) 11141-11156; Ballabio, y col., (1990) *Human Genetics* 84(6) 571-573; Hayden, y col., (2008) *BMC Genetics* 9: 80), PCR anidada, PCR de extensión solapada (véase, p.ej., Higuchi, y col., (1988) *Nucleic Acids Research* 16(15) 7351-7367), PCR tiempo real (véase, p.ej., Higuchi, y col., (1992) *Biotechnology* 10:413-417; Higuchi, y col., (1993) *Biotechnology* 11: 1026-1030), PCR de transcripción inversa (véase, p.ej., Bustin, S.A. (2000) *J. Molecular Endocrinology* 25: 169-193), PCR en fase sólida, PCR térmica de entrelazado asimétrico y PCR de Touchdown (véase, p.ej., Don, y col., *Nucleic Acids Research* (1991) 19(14) 4008; Roux, K. (1994) *Biotechniques* 16(5) 812-814; Hecker, y col., (1996) *Biotechniques* 20(3) 478-485). La amplificación de polinucleótidos también se puede llevar a cabo utilizando PCR digital (véase, p.ej., Kalinina, y col., *Nucleic Acids Research*. 25; 1999-2004, (1997); Vogelstein y Kinzler, *Proc Natl Acad Sci Estados Unidos*. 96; 9236-41, (1999); publicación de patente internacional No. WO05023091A2; publicación de solicitud de patente estadounidense No. 20070202525). Tal como se utiliza en el presente documento, los términos "complementario" o "complementariedad" se utilizan haciendo referencia polinucleótidos (es decir, una secuencia de nucleótidos) relacionados por las reglas de apareamiento de bases. Por ejemplo, la secuencia "5'-A-G-T-3'," es complementaria de la secuencia "3'-T-CA-5'." La complementariedad puede ser "parcial," en la que solamente algunos de las bases de los ácido nucleicos están apareados de acuerdo con las reglas de apareamiento de bases. O, puede haber una complementariedad "completa" o "total" entre los ácidos nucleicos. El grado de complementariedad entre las cadenas de ácido nucleico tiene un significativo efecto en la eficacia e intensidad de la hibridación entre las cadenas de ácido nucleico. Esto reviste una particular importancia en las reacciones de amplificación, así como los procedimientos de detección que dependen de la unión entre ácidos nucleicos.

Tal como se utiliza en el presente documento, el término "cebador" se refiere a un oligonucleótido, ya sea de origen natural, como en un digesto de restricción purificado, o se haya producido sintéticamente, que sea capaz de actuar como punto de inicio de la síntesis cuando se establecen las condiciones en las que se induce la síntesis de un producto de extensión de cebador que es complementario con una cadena de ácido nucleico (p.ej., en presencia de nucleótidos y un agente de inducción como pueda ser un biocatalizador (p.ej., una ADN polimerasa o similares) y a



una temperatura y pH adecuados). El cebador es normalmente monocatenario para una máxima eficiencia en la amplificación, si bien, alternativamente, puede ser bicatenario. Si es bicatenario, generalmente, se trata primero el cebador para separar sus cadenas antes de utilizarlo para preparar productos de extensión. En algunas realizaciones, el cebador es un oligodesoxirribonucleótido. El cebador es suficientemente largo como para estimular la síntesis de los productos de extensión en presencia del agente de inducción. Las longitudes exactas de los cebadores dependerán de muchos factores, incluyendo la temperatura, la fuente del cebador y el uso del procedimiento. En determinadas realizaciones, el cebador es un cebador de captura.

Tal como se utiliza en el presente documento, la expresión "molécula de ácido nucleico" se refiere a cualquier molécula que contiene ácido nucleico, incluyendo, pero sin limitarse a ellos, ADN o ARN. El término abarca secuencias que incluyen cualquiera de los análogos de base de ADN y ARN, incluyendo, pero sin limitarse a ellos, 4 acetilcitosina, 8-hidroxi-N6-metiladenosina, acridinilcitosina, pseudoisocitosina, 5-(carboxihidroxi-metil) uracilo, 5-fluorouracilo, 5-bromouracilo, 5-carboximetilaminometil-2-tiouracilo, 5-carboximetil-aminometiluracilo, dihidouracilo, inosina, N6-isopenteniladenina, 1-metiladenina, 1-metilpseudouracilo, 1-metilguanina, 1-metilinosina, 2,2-dimetilguanina, 2-metiladenina, 2-metilguanina, 3-metil-citosina, 5-metilcitosina, 5-hidroximetilcitosina, b-glucosil-5-hidroximetilcitosina, 5-formilcitosina y 5-carboxicitosina, N6-metiladenina, 7-metilguanina, 5-metilaminometiluracilo, 5-metoxi-amino-metil-2-tiouracilo, beta-D-manosilqueosina, 5'-metoxycarbonilmetiluracilo, 5-metoxiuracilo, 2-metil-N-isopenteniladenina, éster metílico de ácido uracil-5-oxiacético, ácido uracil-5-oxiacético, oxibutoxosina, pseudouracilo, queosina, 2-tiocitosina, 5-metil-2-tiouracilo, 2-tiouracilo, 4-tiouracilo, 5-metiluracilo, éster metílico de ácido N-uracil-5-oxiacético, ácido uracil-5-oxiacético, pseudouracilo, queosina, 2-tiocitosina y 2,6-diaminopurina.

Tal como se utiliza en el presente documento, el término "nucleobase" es sinónimo de otros términos utilizados en la técnica, que incluyen "nucleótido," "desoxinucleótido," "resto nucleótido," "resto desoxinucleótido," "nucleótido trifosfato (NTP)," o desoxinucleótido trifosfato (dNTP).

Un "oligonucleótido" se refiere a un ácido nucleico que incluye al menos dos unidades de monómero de ácido nucleico (p.ej., nucleótidos), normalmente más de tres unidades de monómero y más normalmente más de diez unidades de monómero. El tamaño exacto de un oligonucleótido depende por lo general de varios factores, entre los que se incluyen la función o el uso en última instancia del oligonucleótido. Para mayor ilustración, los oligonucleótidos tienen normalmente una longitud de menos de 200 restos (p.ej., entre 15 y 100), sin embargo, tal como se utiliza en el presente documento, se pretende que el término abarque también cadenas de polinucleótidos más largas. Se suele hacer referencia a los oligonucleótidos según su longitud. Por ejemplo se hace referencia a un oligonucleótido de 24 como aquel de "24-meros". Normalmente, los monómeros de nucleósido están unidos por enlaces fosfodiéster o análogos de los mismos, incluyendo fosforotioato, fosforoditioato, fosforoselenoato, fosforodiselenoato, fosforoanilotoato, fosforanilidato, fosforamidato y similares, incluyendo contra-iones asociados, p.ej., H<sup>+</sup>, NH<sub>4</sub><sup>+</sup>, Na<sup>+</sup> y similares, si están presentes dichos contra-iones. Asimismo, los oligonucleótidos son normalmente monocatenarios. Opcionalmente, se preparan los oligonucleótidos a través de cualquier procedimiento adecuado, incluyendo, pero sin limitarse a ellos, aislamiento de una secuencia existente o natural, replicación o amplificación de ADN, transcripción inversa, clonación y digestión de restricción de secuencias apropiadas, o síntesis química directa a través de un procedimiento como pueda ser el procedimiento de fosfortriéster de Narang y col. (1979) Meth Enzymol. 68: 90-99; el procedimiento de fosfodiéster de Brown y col. (1979) Meth Enzymol. 68: 109-151; el procedimiento de dietilfosforamidato de Beaucage y col. (1981) Tetrahedron Lett. 22: 1859-1862; el procedimiento de triéster de Matteucci y col. (1981) J Am Chem Soc. 103:3185-3191; procedimientos de síntesis automática; o el procedimiento de soporte sólido de la patente estadounidense No. 4.458.066, titulada "PROCEDIMIENTO PARA LA PREPARACION DE POLINUCLEÓTIDOS," publicada el 3 de julio de 1984 para Caruthers y col., u otros procedimientos conocidos entre las personas especializadas en la técnica.

Una "secuencia" de un biopolímero se refiere al orden e identidad de unidades de polímero (p.ej., nucleótidos, etc.) en el biopolímero. La secuencia (p.ej., secuencia base) de un ácido nucleico se lee normalmente en la dirección 5' a 3'.

Tal como se utiliza en el presente documento, el término "sujeto" se refiere a cualquier animal (p.ej., un mamífero), incluyendo, pero sin limitarse a ellos, seres humanos, primates no humanos, roedores y similares, que será el receptor de un tratamiento en particular. Normalmente, los términos "sujeto" y "paciente" se utilizan indistintamente en el presente documento para referirse a un sujeto humano.

El término "gen" se refiere a una secuencia de ácidos nucleicos (p.ej., ADN) que comprende secuencias de codificación necesarias para la producción de un polipéptido, ARN (p.ej. incluyendo, pero sin limitarse a ellos, ARNm, ARNt y ARNr) o precursor. El polipéptido, ARN, o precursor puede estar codificado por una secuencia de codificación de longitud completa o por una porción de la secuencia de codificación siempre y cuando se retengan la actividad deseada o las propiedades funcionales (p.ej., actividad enzimática, unión de ligando, transducción de señal, etc.) de la longitud completa o el fragmento. El término abarca asimismo la región de codificación de un gen estructural y las secuencias incluidas localizadas adyacentes a la región de codificación en ambos extremos 5' y 3' para una distancia de aproximadamente 1 kb en cada extremo, de manera que el gen corresponde a la longitud del ARNm de longitud completa. Las secuencias que están localizadas en dirección 5' de la región codificante y que están presentes en el ARNm se denominan secuencias 5' sin traducir. Las secuencias que están localizadas en dirección 3' o aguas abajo de la región codificante y que están presentes en el ARNm se denominan secuencias 3'

sin traducir. El término “gen” abarca tanto ADNc como las formas genómicas de un gen. Una forma genética, o un clon de un gen, contienen la región codificante interrumpida con secuencias no codificantes que reciben el nombre de “intrones” o “regiones intervinientes” o “secuencias intervinientes.” Los intrones son segmentos de un gen que se transcriben en ARN nuclear (ARNhn); los intrones pueden contener elementos reguladores, tales como potenciadores. Los intrones se separan o se “desempalman” del transcrito nuclear o primario, los intrones, por tanto, están ausentes en el transcrito procesado por el ARN mensajero (ARNm). El ARNm funciona durante la traducción para especificar la secuencia o el orden de aminoácidos en un polipéptido naciente.

### **Descripción detallada de la invención**

La presente invención se refiere a procedimientos y kits para la detección de 5-hidroximetilcitosina (5hmC) y 5-metilcitosina (5meC). En algunas realizaciones, la presente invención se refiere a la detección de 5hmC en ADN genómico, p.ej., ADN genómico de mamífero. Los procedimientos disponibles actualmente para identificar 5hmC tienen un límite de resolución de aproximadamente 50-200 pares de base. Muchos de procedimientos actuales están limitados por la etapa de conversión con bisulfito que no puede distinguir entre 5-metilcitosina (5meC) y 5hmC. La presente invención aborda ambos problemas. En primer lugar, la presente invención permite discriminar entre modificaciones de ADN 5meC y 5hmC. En segundo lugar, la presente invención permite la detección tanto de 5meC como 5hmC en resolución de una sola base.

El procedimiento descrito en el presente documento identifica 5-hidroximetilcitosina (5hmC) en ADN con resolución de una sola base. Adicionalmente, dicho procedimiento puede identificar 5meC en una resolución específica de base de forma simultánea con 5hmC. El procedimiento empleado aprovecha el hecho de que la DNMT1 metil transferasa no puede metilar transversalmente desde una 5hmC (o 5hmC modificada; como es el caso con  $\beta$ -glucosil-5-hidroximetilcitosina) y preferentemente metila transversal desde 5-metilcitosina (5meC). Tras rondas secuenciales de un ciclo de amplificación por PCR y tratamiento del ADN con DNMT1, se diluye la población de ADN que contiene 5hmC en un factor de dos mientras que la población que contiene 5meC permanece estable. Esta dilución acompañada de la conversión con bisulfito permite la identificación específica de base de restos de ADN que contienen 5hmC (Figura 1).

La conversión con bisulfito de ADN tiene como resultado la conversión de citosina sin modificar (C) en uracilo (U) que se leerá como timina (T) tras la secuenciación de ADN amplificado por PCR. Tanto 5meC como 5hmC están protegidos contra la conversión y no se convertirán en U. Por lo tanto, ambos se leerán como C tras la secuenciación (véase Figura 2). La conversión con bisulfito es una tecnología perfectamente establecida considerada durante mucho tiempo como el criterio de referencia para la detección de 5meC y tan solo recientemente (2010) se ha notificado en la bibliografía científica que no se puede distinguir entre 5meC y 5hmC<sup>30</sup>.

Sin embargo, el procedimiento que se describe en el presente documento aprovecha este hecho para crear un conjunto de datos de referencia (con la referencia "A" en la Figura 1).

En la presente invención, se diluye 5hmC en la agrupación total de ADN al mismo tiempo que se mantiene 5meC. Se consigue esta dilución a través de rondas secuenciales de un ciclo de amplificación por PCR y tratamiento del ADN con la ADN metil transferasa de mantenimiento, DNMT1, que mantiene enzimática y específicamente 5meC solamente añadiendo un grupo metilo a la cadena no metilada de los productos de PCR hemimetilados (en la Figura 1 se hace referencia a esta muestra como "B"). Después de una o más rondas de este ensayo, se realiza la conversión con bisulfito, seguida de secuenciación de la muestra de ADN tratada, en la que 5meC es ahora la modificación predominante. Se contempla que todas o la mayoría de las bases leídas como C de esta muestra hayan sido protegidas contra la conversión por 5meC y no por 5mC. Al comparar con la muestra de referencia "A" es posible detectar todas las posiciones base que contienen 5hmC. Se puede conseguir la dilución en una base tan amplia como el genoma o con respecto a un locus génico en particular o una porción de un gen. En las realizaciones preferentes, la región de dilución se define mediante los cebadores utilizados para replicación y/o amplificación de la región de interés diana.

En consecuencia, en algunas realizaciones, la presente invención proporciona procedimientos para detectar o determinar el estado de 5meC y 5hmC en una región predeterminada de una muestra de ADN genómico. En algunas realizaciones preferentes, la región predeterminada (o región de interés diana) corresponde a un locus del gen de interés o una porción de un gen. En algunas realizaciones, la región predeterminada se define por los cebadores de ácido nucleico utilizados para la replicación o amplificación de la región predeterminada. La muestra de ácido nucleico se divide en al menos dos porciones para posterior análisis. En algunas realizaciones, se replica la primera porción en condiciones tales que se mantienen los restos de citosina 5-metilada y se diluyen los restos de citosina 5-hidroximetilada. La presente invención no está limitada a ningún nivel de dilución en particular. Por ejemplo, se pueden diluir los restos de citosina 5-hidroximetilada en un factor de 1,5, 2, 5, 10, 20, 40, 100, 200, 400, 800, 1600 o más.

En algunas realizaciones, la dilución de restos de citosina 5-hidroximetilada se lleva a cabo por replicación del ácido nucleico (preferentemente replicación de la región predeterminada) con una polimerasa para proporcionar ácido nucleico replicado y después el tratamiento del ácido nucleico replicado con una enzima que añade un grupo metilo a la cadena sin metilar del ácido nucleico hemimetilado, pero no añade un grupo hidroximetilo a la cadena sin

hidroximetilar del ácido nucleico hemihidroximetilado. La presente invención no está limitada con el uso de una enzima en particular. En algunas realizaciones, la enzima es una enzima que mantiene el estado de metilación del ADN de un ácido nucleico, por ejemplo una ADN metil transferasa (DNMT). Entre los ejemplos de ADN metil transferasas se incluyen, pero sin limitarse a ellas, DNMT1 de ratón (SEQ ID NO: 1; Figura 7), DNMT1 humana (SEQ ID NO: 2, Figura 8) o DNMT de *M.Sssl* (*Spiroplasma* sp.) (SEQ ID NO: 3, Figura 9) o un homólogo o variante de la misma. En algunas realizaciones, los homólogos o variantes tienen la actividad de añadir un grupo metilo a la cadena no metilada de un ácido nucleico hemimetilado. En algunas realizaciones, los homólogos o variantes tienen al menos un 50 %, 60 %, 70 %, 80 %, 90 %, 95 % o 99 % de identidad con las SEQ ID NO: 1, 2 o 3 y/o tienen la actividad de añadir un grupo metilo a la cadena no metilada de un ácido nucleico hemimetilado.

En algunas realizaciones, la etapa de replicación se realiza a través de una o más rondas de reacción en cadena de la polimerasa. En realizaciones preferentes, se replica una región predeterminada por extensión a partir de cebadores de ácido nucleico que definen los extremos 5' y 3' de la región predeterminada. A continuación se trata el ácido nucleico replicado con una enzima de metilación de ADN, tal como se ha descrito, para mantener la metilación 5-metilcitosina de la región predeterminada y después se repite el procedimiento hasta conseguir el nivel deseado de dilución de restos de citosina 5-hidroximetilada en comparación con los restos 5-metilados. En algunas realizaciones, el nivel de dilución por ciclo es preferentemente aproximadamente el doble, pero puede ser de un factor tan bajo como 1,1. En algunas realizaciones, el nivel de mantenimiento de restos de 5-metil citosina es aproximadamente 100 %, pero puede ser de hasta 10 % y seguir proporcionando una determinación y discriminación eficaz entre los restos de 5mC y 5hmC en la región predeterminada. En algunas realizaciones, el número de ciclos de replicación y tratamiento con enzima de metilación de ADN puede ser 1, 2, 3, 5, 7, 10 o 20 ciclos o más, o entre aproximadamente 1 y 20 ciclos.

En algunas realizaciones, se utilizan cebadores marcados en la etapa de replicación a fin de poder aislar productos de extensión marcados de la etapa de replicación utilizando un reactivo de unión a marcador y utilizarlos en las etapas posteriores, como por ejemplo para el tratamiento con ADN metil transferasa. En realizaciones preferentes, solamente se utilizan las cadenas recién sintetizadas (es decir, cadenas marcadas por el cebador marcado) y se analizan en las etapas posteriores. La Figura 11 proporciona una representación esquemática del uso de cebadores marcados en el procedimiento. En dicha Figura, "A" presenta el ensayo de conversión y secuenciación con bisulfito convencional y "B" presenta el ensayo dependiente de metil transferasa. Tal como se muestra en el panel de la izquierda para el ensayo "B" el uso de extensión de cebador a partir del cebador biotinilado y el posterior aislamiento con perlas de estreptavidina asegura que todas las cadenas inferiores en el análisis sean las recién sintetizadas. Por lo tanto, al realizar el tratamiento con DNMT1 (u otra metil transferasa) y analizar después las cadenas inferiores aisladas con biotina-estreptavidina se conseguirá una cuantificación directa y precisa del nivel de 5mC de la cadena complementaria. En el panel de la derecha, en la parte superior, se presenta los resultados de secuenciación representativos de 10 clones para el ensayo de bisulfito convencional, "A", en el que se leerá tanto 5mC como 5hmC como citosina tras el tratamiento, y los resultados de secuenciación representativos de 10 clones para el ensayo de transferencia de metilo "B" en el que solamente se leerá 5mC como citosina tras el tratamiento. La combinación de los datos del ensayo de bisulfito convencionales "A", en el que se leen tanto 5mC como 5hmC como citosina tras el tratamiento y el ensayo de transferencia de metilo "B" en el que se lee solamente 5mC como citosina tras el tratamiento permite determinar la posición y cantidad de 5hmC (a partir de un sencillo cálculo: A-B = 5hmC). Esta cuantificación se indica en la parte inferior del panel de la derecha. Para replicados experimentales con este mismo resultado cuantitativo se obtuvo n = 15.

La presente invención no está limitada al uso de ningún cebador marcado ni reactivo de unión a marcador en particular para el aislamiento del cebador marcado. En algunas realizaciones preferentes, el cebador está biotinilado y el reactivo de unión a marcador es un reactivo estreptavidina, como por ejemplo una perla de estreptavidina. Las cadenas de ácido nucleico replicadas que comprenden el cebador biotinilado (es decir, el producto de extensión de cebador que resulta de la extensión del cebador biotinilado) se aíslan poniendo en contacto las cadenas con perlas de estreptavidina. Es posible utilizar cualquier combinación de cebador marcado y reactivo de unión a marcador. Otros ejemplos adecuados incluyen cebadores haptenilados y perlas u otros reactivos que comprenden un anticuerpo u otra proteína de unión a antígeno que se une al hapteno. Entre los haptenos adecuados se incluyen, pero sin limitarse a ellos, pirazoles, en particular nitropirazoles; compuestos de nitrofenilo; benzofurazanos; triterpenos; ureas y tioureas, en particular fenil ureas e incluso más particularmente fenil tioureas; rotenona y derivados de rotenona, a los que se hace referencia también en el presente documento como rotenoides; oxazol y tiazoles, particularmente oxazol y tiazol sulfonamidas; cumarina y derivados de cumarina; ciclolignanos, como por ejemplo podofilotoxina y derivados de podofilotoxina; y combinaciones de los mismos. Entre los ejemplos específicos de haptenos se incluyen, pero sin limitarse a ellos, 2,4-dinitrofenil (DNP), biotina, derivados de fluoresceína (FITC, TAMRA, Texas Red, etc.), digoxigenina (DIG), 5-nitro-3-pirozolcarbamida (nitropirazol, NP), 4,5,-dimetoxi-2-nitrocina (nitrocina, NCA), 2-(3,4-dimetoxifenil)-quinolina-4-carbamida (fenilquinolona, DPQ), 2,1,3-benzoxadiazol-5-carbamida (benzofurazano, BF), 3-hidroxi-2-quinoxalincarbamida (hidroxiquinolina, HQ), 4-(dimetilamino)azobenceno-4'-sulfonamida (DABSIL), rotenona isoxazolona (Rot), (E)-2-(2-(2-oxo-2,3-dihidro-1H-benzo[b][1,4]diazepin-4-il)fenoci)acetamida (benzodiazepina, BD), ácido 7-(dietilamino)-2-oxo-2H-cromeno-3-carboxílico (coumarina 343, CDO), 2-acetamido-4- metil-5-tiazolsulfonamida (tiazolsulfonamida, TS) y p-metoxifenilpirazolopodofilamida (Podo).

En algunas realizaciones, se modifican los grupos 5hmC de la muestra con a grupo bloqueante para aumentar la relación de eficiencia de metil transferasa entre 5meC y 5hmC. Tal como se utiliza en el presente documento, un "grupo bloqueante" es cualquier grupo químico que se puede añadir a 5hmC (o citosina en la posición 5-carbono) que haga que el grupo total sea demasiado grande, o esté cargado de forma desfavorable, para el bolsillo de ADN metil transferasa y bloquee así la actividad de ADN metil transferasa en el resto 5hmC. Se contempla que el uso de grupos bloqueantes aumente la relación de especificidad y/o eficiencia de metil transferasa DNMT1 para catalizar la transferencia de un grupo metilo transversalmente desde 5meC y 5hmC en ADNbc. La presente invención no está limitada al uso de ningún grupo bloqueante en particular. Entre los grupos bloqueantes adecuados se incluyen sin limitarse a ellos Glucosa (beta-glucosa y alfa-glucosa); Gentiobiosa (6-O-β-D-glucopiranosil-D-glucosa) (y cualquier otro estereoisómero, también es posible la unión alfa: 6-O-alfa-D-glucopiranosil-D-glucosa); ceto-glucosa; azida-glucosa (p.ej. N3Glucosa); un grupo químico unido a la glucosa o azida-glucosa, p.ej., por química clic, por ejemplo biotina (biotina-N3Glucosa-5hmC); JPB1 (proteína 1 de unión J) unido a glu-5hmC (longitud completa y versiones truncadas); proteínas TET (p.ej. TET1, TET2 y TET3) (longitud completa y versiones truncadas) de unión a 5hmC; otras proteínas de unión a 5hmC o Glu-5hmC y/o dominios de unión a proteína; (versiones nativas y reticuladas de proteínas); cualquier producto de oxidación de glucosa o glucosa modificada, p.ej., glucosa oxidada con periodato; cualquier grupo químico que reaccione con glucosa oxidada para unir o modificar la glucosa; y cualquier proteína o complejo de proteínas que pueda identificar específicamente 5meC, 5hmC y variantes modificadas de estas bases (p.ej., JPB1 y proteínas de la clase MPB (p.ej., MPB1 y MeCP2)).

Sin el bloqueo es posible conseguir 100 % frente al 0 % frente al 0 % de transferencia de metilo transversalmente desde 5meC, C y 5hmC respectivamente, si bien el procedimiento es también aplicable a menos de 100 % transferencia de metilo transversalmente desde 5meC y más de 0 % transferencia transversalmente desde C y 5hmC. En estos casos, se puede obtener una mayor precisión en la cuantificación cuando se añade un control conocido a la muestra para poder determinar la eficiencia en la muestra. Con el bloqueo es posible conseguir un 100 % frente a 0 % frente a 0 % de transferencia de metilo transversalmente desde 5meC, C y 5hmC respectivamente, si bien se puede aplicar el procedimiento a menos de 100 % de transferencia de metilo transversalmente desde 5meC y más de 0 % de transferencia transversalmente desde C y 5hmC. El bloqueo puede ser útil para el ensayo "convencional" ya que permitirá que obtener con mayor solidez 100 % frente a 0% frente a 0% con el ensayo DNMT1 con un índice de éxito superior en comparación con la falta de bloqueo.

Con el bloqueo, se puede conseguir un 100 % frente a 100 % frente a 0% de transferencia de metilo transversalmente desde 5meC, C, 5hmC respectivamente para M.Sssl (o DNMT1, preferentemente un gran exceso molar de DNMT1). La metilación transversalmente desde 5meC y C es una forma alternativa de transferir la información del estado de modificación desde la cadena parental a la cadena replicada/extendida con cebador para ayudar a identificar las posiciones y cantidades de 5meC, 5hmC y C. Esto puede permitir la lectura directa de 5hmC como citosinas sin modificar que no están protegidas de la conversión con bisulfito o, en comparación con la conversión y secuenciación con bisulfito convencional, puede revelar información cuantitativa para 5meC, C y 5hmC en la secuencia de ácidos nucleicos. Esto permitirá revelar a través de un sencillo cálculo la posición y cantidad de 5meC, C y 5hmC.

Es probable que se pueda conseguir la identificación de 5meC, 5hmC y C si se realiza el bloqueo en los restos 5hmC y C. Los agentes bloqueantes en los restos de citosina podrían ser por ejemplo proteínas que contienen el motivo CXXC o cualquier proteína o fragmento de las mismas que se pueda unir a CpG sin modificar.

En algunas realizaciones, se trata la muestra de ácido nucleico de 5hmC diluido y la porción sin diluir para convertir restos de citosina sin modificar en restos timidina. En realizaciones preferentes, se tratan las porciones con bisulfito para convertir restos de citosina sin modificar en restos de uracilo. El ácido nucleico tratado con bisulfito se replica después con una polimerasa para convertir dichos restos de uracilo en restos timidina. En algunas realizaciones, la etapa de replicación se realiza a través de una o más rondas de reacción en cadena de la polimerasa (véase, p.ej., Figuras 1 y 5) o reacción de extensión de cebador (véase, p.ej., Fig. 5). En realizaciones preferentes, se replica una región predeterminada por extensión desde cebadores de ácido nucleico que definen los extremos 5' y 3' de la región predeterminada. En algunas realizaciones, el número de ciclos de replicación puede ser superior a 2, 3, 5, 7, 10 o 20 ciclos o entre aproximadamente 2 y 20 ciclos.

El procedimiento descrito en los párrafos anteriores proporciona dos porciones de ácido nucleico diferentes. En la primera porción, se han diluido los restos 5-hidroximetilados en comparación con los restos de citosina 5-metilada, que se han mantenido. En la segunda porción, no se han diluido los restos 5-hidroximetilados. Cuando se tratan las porciones con bisulfito, todos los restos de citosina sin modificar se convierten en restos de uracilo y después en restos timidina tras 1 o más rondas de replicación o extensión de cebador. En realizaciones preferentes, se secuencian ambas porciones, preferentemente utilizando cebadores que permiten la secuenciación de la región predeterminada. En realizaciones preferentes, la comparación de las secuencias de las porciones primera y segunda permite la identificación de restos de 5meC y 5hmC en la región predeterminada. Se identifican restos de 5hmC como restos que se leen por secuenciación como un resto de timidina en la primera porción (es decir, la porción en la que se han diluido los restos de 5hmC restos) y como resto citosina en la posición correspondiente en la segunda porción de ácido nucleico y se identifican restos de 5meC restos como restos que se leen como citosina tanto en la porción primera como la segunda del ácido nucleico.

La secuenciación de las muestras de ácido nucleico se puede realizar a través de cualquiera de los procedimientos conocidos en la técnica. Entre los procedimientos de secuenciación adecuados se incluyen, pero sin limitarse a ellos, procedimientos de secuenciación de terminación de cadena (p.ej., procedimientos de secuenciación de Sanger) y procedimientos de secuenciación de ADN de segunda generación en los que se utilizan los sistemas provistos por Illumina (San Diego CA), Pacific Biosciences (Menlo Park, CA) y otros. En las realizaciones en las que se utilizan procedimientos de secuenciación de segunda generación, la etapa de replicación con la polimerasa antes de la secuenciación (que convierte el resto uracilo en un resto timidina) es opcional y se puede leer el resto uracilo directamente.

En algunas realizaciones, se utilizan los procedimientos descritos para predecir una predisposición de un sujeto a contraer una enfermedad, diagnosticar una enfermedad en un sujeto, predecir la probabilidad de recurrencia de una enfermedad en un sujeto, proporcionar un diagnóstico para un sujeto que tiene una enfermedad o seleccionar a un sujeto que sufre una enfermedad para su tratamiento con una terapia en particular. Dichos procedimientos comprenden preferentemente proporcionar una muestra de ADN genómico de un sujeto; y detectar el estado de metilación de regiones predeterminadas de la muestra de ADN genómico a través de los procedimientos que se han descrito. En algunas realizaciones, una alteración del nivel de metilación de 5-hidroximetilcitosina y/o 5-metilcitosina (es decir, un mayor o un menor nivel) de las regiones predeterminadas del ADN genómico con respecto a un estado de metilación de referencia proporciona una indicación seleccionada del grupo que consiste en una indicación de una predisposición del sujeto a una enfermedad, una indicación de que el sujeto tiene una enfermedad, una indicación de la probabilidad de recurrencia de una enfermedad en el sujeto, una indicación de supervivencia del sujeto y una indicación de que el sujeto es candidato para tratamiento con una terapia en particular.

En consecuencia, en algunas realizaciones, los procedimientos de la presente invención implican la determinación (p.ej., evaluación, valoración, cuantificación) del nivel de modificación 5meC y/o 5hmC de un indicador de una afección de interés, como pueda ser neoplasma, en una muestra. Las personas expertas en la materia comprenden que un nivel de modificación 5meC y/o 5hmC mayor, menor, informativo o diferente de otra forma claramente distinguible se articula con respecto a una referencia (p.ej., un nivel de referencia, un nivel de control, un nivel umbral o similar). Por ejemplo, la expresión "nivel elevado de 5hmC o 5meC", tal como se utiliza en el presente documento con respecto al estado de 5hmC o 5meC de un locus de gen es cualquier nivel de 5hmC y/o 5meC que está por encima de la mediana del nivel de 5hmC o 5meC en una muestra de una población aleatoria de mamíferos (p.ej., una población aleatoria de 10, 20, 30, 40, 50, 100 o 500 mamíferos) que no tienen un neoplasma (p.ej., un cáncer) u otra afección de interés. Niveles elevados de modificación 5meC y/o 5hmC puede ser cualquier nivel siempre y cuando el nivel sea mayor que el nivel de referencia correspondiente. Por ejemplo, un nivel elevado de 5meC y/o 5hmC de un locus de interés puede ser 0,5, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, veces más que el nivel de referencia de 5meC y/o 5hmC observado en una muestra normal. Debe advertirse que el nivel de referencia puede ser cualquier cantidad. La expresión "puntuación elevada de 5meC y/o 5hmC", tal como se utiliza en el presente documento con respecto a los eventos de 5meC y/o 5hmC detectados en un panel de matrices de marcadores de ácido nucleico en particular es cualquier puntuación de 5meC y/o 5hmC que está por encima de la mediana de la puntuación de 5meC y/o 5hmC en una muestra de una población aleatoria de mamíferos (p.ej., una población aleatoria de 10, 20, 30, 40, 50, 100 o 500 mamíferos) que no tienen un neoplasma (p.ej., un cáncer). Una puntuación elevada de 5hmC en un panel de matrices de marcadores de ácido nucleico en particular puede ser cualquier puntuación siempre y cuando la puntuación sea superior a la de la puntuación de referencia correspondiente. Por ejemplo, una puntuación elevada de 5meC y/o 5hmC en un locus de interés puede ser 0,5, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 o más veces por encima de la puntuación de referencia de 5meC y/o 5hmC en una muestra normal. Debe advertirse que la puntuación de referencia puede ser cualquier cantidad utilizada con fines comparativos.

Se aplican consideraciones similares para los ensayos del descenso de los niveles de modificaciones de 5meC y/o 5hmC en una muestra, locus diana, región genómica diana y similares. Por ejemplo, la expresión "descenso del nivel de 5meC y/o 5hmC" tal como se utiliza en el presente documento con respecto al estado de 5meC y/o 5hmC de un locus de gen es cualquier nivel de 5meC y/o 5hmC que esté por debajo de la mediana del nivel de 5meC y/o 5hmC en una muestra de una población aleatoria de mamíferos (p.ej., una población aleatoria de 10, 20, 30, 40, 50, 100 o 500 mamíferos) que no tiene un neoplasma (p.ej., un cáncer). El descenso de los niveles de modificación 5meC y/o 5hmC puede ser cualquier nivel siempre y cuando el nivel sea menor que el nivel de referencia correspondiente. Por ejemplo, un descenso del nivel de 5meC y/o 5hmC de un locus de interés puede ser 0,5, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 o más veces por debajo del nivel de referencia de 5meC y/o 5hmC observado en una muestra normal. Debe advertirse que un nivel de referencia puede ser cualquier cantidad. La expresión "menor puntuación de 5hmC", tal como se utiliza en el presente documento con respecto a los eventos de 5meC y/o 5hmC detectados en un panel de matrices de marcadores de ácido nucleico en particular es cualquier puntuación de 5meC y/o 5hmC que esté por debajo de la mediana de la puntuación de 5meC y/o 5hmC en una muestra de una población aleatoria de mamíferos (p.ej., una población aleatoria de 10, 20, 30, 40, 50, 100 o 500 mamíferos) que no tienen un neoplasma (p.ej., un cáncer). Una menor puntuación de 5meC y/o 5hmC en un panel de matrices de marcadores de ácido nucleico en particular puede ser cualquier puntuación siempre y cuando dicha puntuación sea mayor que la puntuación de referencia correspondiente. Por ejemplo, una puntuación menor de 5meC y/o 5hmC en un locus de interés puede ser 0,5, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, o más veces por debajo de la puntuación de referencia de 5meC y/o 5hmC observada en una muestra normal. Debe advertirse que la puntuación de referencia puede ser cualquier cantidad que se utilice con fines comparativos.

Los procedimientos no están limitados a un tipo de mamífero en particular. En algunas realizaciones, el mamífero es un ser humano. En algunas realizaciones, el neoplasma es premaligno. En algunas realizaciones, el neoplasma es maligno. En algunas realizaciones, el neoplasma es cáncer sin tener en cuenta el estadio (p.ej., estadio I, II, III o IV).

5 La presente divulgación proporciona también procedimientos y materiales para ayudar a los profesionales de la medicina y la investigación a determinar si un mamífero padece o no un neoplasma (p.ej., cáncer). Los profesionales de la medicina pueden ser por ejemplo médicos, enfermeras, técnicos médicos y farmacéuticos de laboratorio. Los profesionales de investigación pueden ser por ejemplo investigadores principales, técnicos de investigación, posdoctorados en prácticas y estudiantes de posgrado. Se puede ayudar a un profesional (1) determinando la relación de 5hmC y/u otros marcadores en una muestra y /2) comunicando la información sobre la relación a dicho profesional, por ejemplo.

10 Una vez notificado el nivel (p.ej., la puntuación o frecuencia) de una modificación 5meC y/o 5hmC en particular en una muestra, el profesional de la medicina puede adoptar una o más acciones que afecten a la atención médica del paciente. Por ejemplo, un profesional de la medicina puede registrar los resultados en la historia clínica del paciente. En algunos casos, un profesional de la medicina puede registrar un diagnóstico de neoplasia o transformar de otra forma la historia clínica del paciente para que refleje la patología del paciente. En algunos casos un profesional de la medicina puede revisar y evaluar la historia clínica del paciente completa y evaluar varias estrategias de tratamiento, para la intervención clínica de una afección del paciente. En algunos casos, un profesional de la medicina puede registrar la predicción de que se produzca un tumor con otros indicadores notificados. En algunos casos, un profesional de la medicina puede revisar y evaluar la historia clínica del paciente en su totalidad y evaluar varias estrategias de tratamiento, para la intervención clínica de una afección del paciente.

15 Un profesional de la medicina puede iniciar o modificar el tratamiento de un neoplasma una vez recibida la información en lo que respecta al nivel (puntuación, frecuencia) asociada con el nivel de 5meC y/o 5hmC en una muestra de orina del paciente. En algunos casos, un profesional de la medicina puede comparar informes anteriores y el nivel (puntuación, frecuencia) de modificación de 5meC y/o 5hmC que se ha comunicado recientemente y recomendar un cambio en la terapia. En algunos casos, el profesional de la medicina puede reclutar a un paciente para un ensayo clínico para una nueva intervención terapéutica de neoplasma. En algunos casos, un profesional de la medicina puede optar por esperar a comenzar la terapia hasta que los síntomas del paciente requieran la intervención clínica.

20 Un profesional de la medicina puede comunicar los resultados del ensayo al paciente o a un familiar del paciente. En algunos casos, un profesional de la medicina puede proporcionar información relacionada con la neoplasia al paciente y/o un familiar del paciente, incluyendo opciones de tratamiento, un pronóstico o la derivación a especialistas, p.ej., oncólogos y/o radiólogos. En algunos casos, un profesional de la medicina puede proporcionar una copia de la historia clínica del paciente para comunicar los resultados del ensayo al especialista. Un profesional de investigación puede solicitar información sobre los resultados del ensayo del sujeto para avanzar en la investigación del neoplasma. Por ejemplo, un investigador puede recopilar los datos sobre los resultados del ensayo con información sobre la eficacia de un fármaco para el tratamiento de neoplasia para identificar un tratamiento eficaz. En algunos casos, un profesional de investigación puede obtener los resultados de ensayo para evaluar el reclutamiento de un sujeto o la participación continuada en un estudio de investigación o un ensayo clínico. En algunos casos, un profesional de investigación puede clasificar la gravedad de una afección del sujeto, sobre la base de los resultados de ensayo. En algunos casos, un profesional de investigación puede comunicar los resultados del ensayo del sujeto a un profesional de la medicina. En algunos casos, un profesional de investigación puede derivar un sujeto a un profesional de la medicina para la evaluación clínica de una neoplasia y su tratamiento. Se puede emplear un procedimiento apropiado para comunicar la información a otra persona (p.ej., un profesional). Por ejemplo, la información puede proporcionarse directa o indirectamente a un profesional. Por ejemplo, un técnico de laboratorio puede introducir los resultados del ensayo en un registro informático. En algunos casos, se comunica la información introduciendo una alteración física en los registros clínicos o de investigación. Por ejemplo, un profesional de la medicina puede realizar una anotación permanente médica o marcar una historia médica para comunicar un diagnóstico a otros profesionales de la medicina que revisen dicha historia. Por otra parte, se puede emplear cualquier tipo de comunicación para transmitir la información. Por ejemplo, correo físico, e-mail, teléfono e interacciones cara a cara. Asimismo, es posible transmitir la información a un profesional facilitando a dicho profesional la información por vía electrónica. Por ejemplo, es posible transmitir la información a un profesional introduciendo la información en una base de datos informática de manera que el profesional pueda acceder a dicha información. Por otra parte, es posible comunicar la información a un hospital, clínica o centro de investigación como agente para el profesional.

25 30 35 40 45 50 55 60 Debe advertirse es posible analizar la muestra en cuanto a un marcador específico de neoplasma o en cuanto a varios marcadores específicos de neoplasma. En realizaciones preferentes, se analiza una sola muestra en cuanto a varios marcadores específicos de neoplasma, por ejemplo, empleando ensayos multi-marcador. Por otra parte, es posible extraer varias muestras de un solo mamífero y analizarlas tal como se describe en el presente documento. En algunas realizaciones, se divide una muestra en una primera y una segunda porción, en las que se somete la primera porción a análisis citológico y la segunda porción a purificación u otros tratamientos (p.ej., etapa(s) de captura específica de secuencia (p.ej. para el aislamiento de loci específicos para el análisis de los niveles de 5hmC). En algunas realizaciones, se somete la muestra a una o más etapas del procedimiento antes de dividirla en

porciones. En algunas realizaciones, se trata, manipula o preserva la muestra de modo que se promueva la integridad del ADN y/o se inhiba la degradación de ADN (p.ej. a través del uso de tampones de almacenamiento con agentes de estabilización (p.ej., agentes quelantes, inhibidores de ADNasa) o técnicas de manipulación o tratamiento que promuevan la integridad de ADN (p.ej., tratamiento inmediato o almacenamiento a baja temperatura (p.ej., -80 grados C)).

En algunas realizaciones, se ensamblan en un kit todos los materiales esenciales y reactivos básicos para detectar neoplasia a través de la detección del nivel (presencia, ausencia, puntuación, frecuencia) de marcadores en una muestra obtenida de un mamífero, tal como se define en la reivindicación 14. Dichos kit comprenden generalmente, por ejemplo, reactivos útiles, suficientes o necesarios para detectar y/o caracterizar uno o más marcadores (p.ej., marcadores epigenéticos; modificaciones de 5hmC) específicos para un neoplasma. En algunas realizaciones, los kits contienen enzimas adecuadas para amplificar los ácidos nucleicos incluyendo diversas polimerasas, desoxinucleótidos y tampones para proporcionar la mezcla de reacción necesaria para la amplificación. En algunas realizaciones, los kits de la presente invención incluyen un medio para contener los reactivos confinados en el espacio para su venta comercial, como pueda ser p.ej., envases plásticos moldeados por soplado o por inyección en los que se retiene el reactivo deseado. Se proporciona asimismo otros envases adecuados para llevar a cabo las etapas concretas de los procedimientos desvelados.

En algunas realizaciones, los procedimientos desvelados en el presente documento son útiles para llevar un seguimiento del tratamiento de neoplasia (p.ej., cáncer). Por ejemplo, en algunas realizaciones, los procedimientos pueden llevarse a cabo inmediatamente antes del tratamiento, durante el mismo y/o después del mismo para llevar un seguimiento del éxito del tratamiento. En algunas realizaciones, se llevan a cabo los procedimientos a intervalos en pacientes sin la enfermedad para asegurar el éxito del tratamiento.

La presente divulgación proporciona además una serie de realizaciones con soporte informático. Concretamente, en algunas realizaciones, la invención proporciona un programa informático para analizar y comparar los resultados de un patrón de detección de marcador específico de neoplasma en una muestra obtenida de un sujeto, por ejemplo, con una biblioteca de dichos patrones de marcador reconocidos como indicativos de la presencia o ausencia de un neoplasma o un estadio en particular o neoplasma. La presente divulgación proporciona un programa informático para analizar y comparar un primer y un segundo patrón de resultados de detección de un marcador específico de neoplasma a partir de una muestra tomada al menos en dos puntos temporales diferentes. En algunas partes de la divulgación, el primer patrón puede ser indicativo de una afección pre-cancerosa y/o un estado de bajo riesgo para el cáncer y/o avance desde un estado pre-canceroso en un estado canceroso. En dichas partes, la comparación proporciona un seguimiento del avance de la afección desde un primer punto temporal hasta el segundo punto temporal.

En otra parte más, la divulgación proporciona un programa informático para analizar y comparar un patrón de los resultados de detección el marcador específico de neoplasma a partir de una muestra con una biblioteca de patrones demarcadores específicos de neoplasma reconocidos como indicativos de la presencia o ausencia de un cáncer, en el que la comparación proporciona por ejemplo un diagnóstico diferencial entre un neoplasma benigno y un neoplasma agresivamente maligno (p.ej., el patrón de marcador proporciona la estadificación y/o graduación del estado canceroso).

Los procedimientos y sistemas descritos en el presente documento pueden implementarse de diversas formas. En una parte de la divulgación, los procedimientos implican el uso de una infraestructura de comunicación, como pueda ser internet. Varias partes de la divulgación se explican más adelante. Debe entenderse asimismo que la presente divulgación puede implementarse en diversas formas de hardware, software, firmware, procesadores, servidores (p.ej., tal como se utiliza en computación en la nube) o una combinación de los mismos. Los procedimientos y sistemas descritos en el presente documento se implementan como una combinación de hardware y software. El software puede implementarse como un programa de aplicación, concretado tangiblemente en un dispositivo de almacenamiento informático, o diferentes partes de un software implementado en el entorno informático del usuario (p.ej., un applet) y en el entorno informático del revisor, pudiendo estar localizado el revisor en un emplazamiento alejado (p.ej., un centro de servicio del proveedor).

Por ejemplo, es posible procesar los datos en el entorno informático de la parte del usuario mientras el usuario introduce los datos o después de introducirlos. Por ejemplo, puede programarse el entorno informático del usuario para proporcionar códigos de prueba definidos para representar una plataforma, prueba de portador/diagnóstico o ambos; procesamiento de datos utilizando indicadores definidos y/o generación de configuración de indicadores, transmitiéndose las respuestas como procesadas o parcialmente procesadas en el entorno informático del revisor en forma de código de prueba o configuraciones de indicador para la posterior ejecución de uno o más algoritmos para proporcionar un resultado y/o generar un informe del entorno informático del revisor.

El programa de aplicación para ejecutar los algoritmos descritos en el presente documento se puede cargar y ejecutar a través de una máquina que comprende cualquier arquitectura adecuada. En general, la máquina implica una plataforma informática que tiene un hardware, como por ejemplo una o más unidades centrales de procesamiento (CPU), una memoria de acceso aleatorio (RAM) e interfaces de entrada/salida (I/O). La plataforma informática incluye también un sistema operativo y un código de microinstrucción. Los distintos procedimientos y

funciones que se describen en el presente documento pueden formar parte del código de microinstrucción o pueden formar parte del programa de aplicación (o una combinación de los mismos) que se ejecutan a través del sistema operativo. Por otra parte, se pueden conectar otros dispositivos periféricos diversos a la plataforma informática, como puedan ser un dispositivo de almacenamiento de datos adicional o una impresora.

5 Como sistema informático, el sistema incluye generalmente una unidad de procesamiento. La unidad de procesamiento funciona para recibir información, que por lo general incluye datos del ensayo (p.ej. productos génicos específicos ensayados) y datos de los resultados del ensayo (p.ej., el patrón de resultados de detección de marcador específico de neoplasma (p.ej. marcador epigenético, modificador de 5hmC) de una muestra). La información recibida se puede almacenar al menos temporalmente en una base de datos y analizarse los datos comparándolos con una biblioteca de patrones de marcador reconocidos como indicativos de la presencia o ausencia de un estado pre-canceroso, o reconocidos como indicativos de un estadio y/o grado de cáncer.

10 Asimismo, es posible enviar electrónicamente parte o todos los datos de entrada y salida; es posible enviar electrónicamente o por vía telefónica (p.e., por fax, p.ej., utilizando dispositivos como pueda ser un fax) virtuales ciertos datos de salida. Entre los ejemplos de dispositivos receptores de salida se incluyen un elemento de pantalla, una impresora, un dispositivo de fax y similares. Entre las formas electrónicas de transmisión y/o despliegue en pantalla se incluyen email, televisión interactiva, y similares. En algunas partes de la divulgación, se mantienen en un servidor todos o una porción de los datos de entrada y/o todos o una porción de los datos de salida (p.ej., normalmente, al menos la biblioteca de patrones de resultados de detección de marcador específico de neoplasma reconocidos como indicativos de la presencia o ausencia de un estado pre-canceroso) para su acceso, p.ej., acceso confidencial. Los resultados se pueden enviar o hacer accesibles a los profesionales, según se desee.

15 Un sistema para su uso en los procedimientos descritos en el presente documento incluye por lo general al menos un procesador informático (p.ej., en los casos en los que el procedimiento se lleva a cabo en su totalidad en un solo emplazamiento) o al menos dos procesadores informáticos en red (p.ej., en los casos en los que un usuario introduce los datos del marcador detectado para una muestra obtenida de un sujeto (p.ej., un técnico o la persona que realice los ensayos) y se transmite a un emplazamiento a distancia a un segundo procesador informático para su análisis (p.ej., cuando se comparan los resultados de detección del patrón del marcador específico de neoplasma con una biblioteca de patrones reconocidos como indicativos de la presencia o ausencia de un estado pre-canceroso), estando los procesadores informáticos primero y segundo conectados por una red (p.ej., por intranet o internet). El sistema puede incluir un componente(s) de usuario para entrada; y un componente(s) de revisor para revisar los datos y generar informes, incluyendo detección de un estado pre-canceroso, estadificación y/o graduación de un neoplasma o seguimiento del avance de un estado pre-canceroso o un neoplasma. Otros componentes del sistema pueden incluir un componente(s) de servidor; y una base(s) de datos para almacenar datos (p.ej., una base de datos de elementos de informe, p.ej. una biblioteca de patrones de marcador reconocidos como indicativos de la presencia o ausencia de un estado pre-canceroso y/o reconocidos como indicativos de un grado y/o estadio de un neoplasma o una base de datos relacional (BDR) que puede incluir entrada de datos del usuario y datos de salida. Los procesadores informáticos pueden ser procesadores que se encuentran normalmente en ordenadores personales (p.ej., IBM, Dell, Macintosh), ordenadores portátiles, ordenadores centrales, miniordenadores u otros dispositivos informáticos.

20 Los componentes de entrada pueden ser ordenadores personales completos, independientes que ofrecen toda la gama de potencia y características para ejecutar aplicaciones. El componente de usuario funciona por lo general en cualquier sistema operativo deseado e incluye un elemento de comunicación (p.ej., un modem u otro hardware para conexión con la red), uno o más dispositivos de entrada (p.ej., un teclado, un ratón, un teclado numérico u otro dispositivo utilizado para transferir información o comandos), un elemento de almacenamiento (p.ej., un disco duro o un medio de almacenamiento con soporte informático para escritura y lectura) y un elemento de pantalla (p.ej., un monitor, una televisión, LCD, LED u otro dispositivo de pantalla que convierte la información para el usuario). El usuario introduce los comandos de entrada en el procesador informático a través de un dispositivo de entrada. Generalmente, la interfaz de usuario es una interfaz gráfica de usuario (GUI) escrita para aplicaciones del navegador.

25 El(los) componente(s) de servidor pueden consistir en un ordenador personal, un microordenador y un ordenador central o se pueden distribuir a través de varios servidores (p.ej., como aplicaciones informáticas de la nube) y ofrece gestión de datos, información compartida entre clientes, administración y seguridad de la red. La aplicación y las bases de datos que se utilicen pueden estar en los mismos servidores o diferentes. Se contemplan también otras disposiciones informáticas para el usuario y el(los) servidor(es) incluyendo procesamiento en una sola máquina, como pueda ser un ordenador central, una colección de máquinas, u otra configuración adecuada. En general, las máquinas del usuario y el servidor funcionan en conjunto para llevar a efecto el procesamiento de la presente invención.

30 Cuando se utilizan, la(s) base(s) de datos está(n) conectadas normalmente con el componente de servidor de base de datos y pueden ser cualquier dispositivo que retiene datos. Por ejemplo, la base de datos puede ser cualquier dispositivo de almacenamiento óptico o magnético para un ordenador (p.ej., CDROM, disco duro interno, una unidad de cinta). La base de datos puede estar localizada a distancia del componente del servidor (con acceso por red, modem, etc.) o localmente con respecto al componente del servidor.



5 Cuando se utiliza en el sistema y los procedimientos, la base de datos puede ser una base de datos relacional que está organizada y a la que se accede de acuerdo con las relaciones entre los ítems de los datos. La base de datos relacional está compuesta generalmente de varias tablas (entidades). Las filas de una tabla representan los registros (colecciones de información sobre ítems por separado) y las columnas representan los campos (atributos en particular de un registro). En su concepción más sencilla, la base de datos relacional es una colección de entradas de datos que "guardan una relación" entre sí a través de al menos un campo común.

10 Se pueden utilizar puestos de trabajo adicionales equipados con ordenadores e impresoras en el punto de servicio para introducir los datos y, en algunas partes de la divulgación, generar informes apropiados, si se desea. El(los) ordenador(es) pueden tener un acceso directo (p.ej., en la mesa de escritorio) para lanzar la aplicación para facilitar el inicio de la entrada de datos, transmisión, análisis, recepción de informe, etc., si se desea.

15 La presente divulgación es útil tanto para diagnosticar enfermedades y trastornos en un sujeto como para determinar el pronóstico del sujeto. Los procedimientos, reactivos y sistemas de la presente divulgación se puede aplicar a una amplia variedad de enfermedades y trastornos. Determinadas partes de la divulgación proporcionan procedimientos para obtener el perfil de riesgo de un sujeto de desarrollar un neoplasma (p.ej., cáncer). En algunas partes de la divulgación, dichos procedimientos implican la obtención de una muestra de un sujeto (p.ej., un ser humano en riesgo de desarrollar un cáncer; un ser humano al que se le realiza un examen físico de rutina), la detección de la presencia, ausencia o nivel (p.ej., frecuencia o puntuación o puntuación de modificación 5hmC) de uno o más marcadores específicos para un neoplasma en la muestra o asociado con ella (p.ej., específico para un neoplasma) en la muestra y la generación de un perfil de riesgo de desarrollar neoplasma (p.ej., cáncer) sobre la base del nivel detectado (puntuación, frecuencia) de la presencia o ausencia de indicadores de neoplasia. Por ejemplo, en algunas partes de la divulgación, el perfil de riesgo generado cambiará dependiendo de los marcadores específicos y que se detecten como presentes o ausentes o en niveles de umbral definido. La presente divulgación no está limitada de forma particular en la manera de generar el perfil de riesgo. En algunas partes de la divulgación, se utiliza un procesador (p.ej., ordenador) para generar dicho perfil de riesgo. En algunas partes de la divulgación, el procesador utiliza un algoritmo (p.ej., software) específico para interpretar la presencia o ausencia de modificaciones 5hmC específicas, según se determina con los procedimientos de la presente invención. En algunas partes de la divulgación, la presencia y ausencia de marcadores específicos, según se determina con los procedimientos de la presente divulgación se introducen en dicho algoritmo y se registra el perfil de riesgo sobre la base de una comparación de dicha entrada con las normas establecidas (p.ej., norma establecida de estado pre-canceroso, norma establecida para varios niveles de riesgo para un cáncer en desarrollo, norma establecida para sujetos diagnosticados con diferentes estadios de cáncer). En algunas partes de la divulgación, el perfil de riesgo indica un riesgo del sujeto de desarrollar cáncer o un riesgo del sujeto de volver a desarrollar cáncer. En algunas partes de la divulgación, el perfil de riesgo indica por ejemplo un sujeto que tiene una posibilidad muy baja, baja, moderada, alta o muy alta de desarrollar o volver a desarrollar cáncer. En algunas partes de la divulgación, el profesional sanitario (p.ej., un oncólogo) utilizará dicho perfil de riesgo para determinar el curso de tratamiento o la intervención (p.ej., biopsia, la espera, la derivación a un oncólogo, la derivación al cirujano, etc.).

40 Otras enfermedades y trastornos que se pueden diagnosticar o pronosticar con los procedimientos, reactivos y sistemas de la presente divulgación incluyen, pero sin limitarse a ellas, síndrome de Prader-Willi, síndrome de Angelman, síndrome de Beckwith-Wiedemann, pseudohipoparatiroidismo, síndrome de Russell-Silver, síndrome ICF, síndrome de Rett,  $\alpha$ -talasemia/ retraso mental, X-ligado (ATR-X), displasia inmuno ósea, de tipo Schimke, síndrome de Rubinstein-Taybi, deficiencia de MTHFR, mola hidatidiforme recurrente, síndrome de retraso mental X frágil, delección de LCP  $\gamma\delta\beta$  y  $\delta\beta$ -talasemia, distrofia FSH, trastorno de XIC, displasia inmuno ósea de Schimke (SIOD), síndrome de Sotos, atriquia, distrofia muscular de Emery-Dreifuss ligada a X (EDMD), EDMD autosómica, CMT2B1, displasia mandibuloacral, distrofia muscular del anillo óseo tipo 1B, lipodistrofia parcial familiar, cardiomiopatía dilatada 1ª, síndrome de progeria de Hutchinson-Gilford y anomalía de Pelger-Huet.

Los siguientes ejemplos sirven para demostrar e ilustrar mejor ciertas realizaciones y aspectos preferentes de la presente invención y no deben interpretarse como limitativos del ámbito de la misma.

**Ejemplo 1**

50 El procedimiento descrito en el presente documento se basa en la sucesiva dilución por PCR de la modificación 5hmC junto con el mantenimiento de la modificación 5meC dado que DNMT1 no puede metilar transversalmente desde 5hmC y citosina; sin embargo, DNMT1 puede metilar ADN transversalmente desde 5meC. Las Figuras 3 y 4 demuestran que DNMT1 no puede catalizar la transferencia de un grupo metilo desde S-adenosil-metilmetionina cuando el sustrato de ADN es una citosina, 5hmC o  $\beta$ -glucosil-5hmC. Por tanto, es posible diluir la modificación 5hmC por PCR seguido del tratamiento con DNMT1 al mismo tiempo que se mantiene la modificación 5meC a través de múltiples rondas de PCR y tratamiento con DNMT1.

60 El procedimiento de los autores de la invención aplica la conversión y secuenciación con bisulfito de muestra "A", ADN sin tratar, que se utilizará como referencia para detectar el total de 5meC y 5hmC. El procedimiento implica un ensayo de dilución de 5hmC, dilución de 5hmC en la agrupación total de fragmentos de ADN al mismo tiempo que mantiene 5meC. La dilución se consigue a través de rondas secuenciales de un ciclo de amplificación por PCR (dilución) y tratamiento del ADN con ADN metil transferasa de mantenimiento, DNMT1, que mantiene enzimática y

específicamente 5meC añadiendo un grupo metilo únicamente a la cadena no metilada de los productos de PCR hemimetilados (en la Figura 1, se hace referencia a esta muestra como muestra "B"). Después de unas cuantas rondas de este ensayo, se aplicó la conversión con bisulfito y la secuenciación de la muestra de ADN tratada, muestra B, en la que 5meC es ahora la única modificación presente (o la única modificación altamente mantenida).  
 5 Por lo tanto, todas las bases (o la mayoría de ellas) que se leen como C desde la muestra han debido ser protegidas contra la conversión, por 5meC y no por 5hmC. Al comparar "B" con esta muestra de referencia "A" es posible detectar fácilmente todas las posiciones de base que contienen 5hmC.

Debe advertirse que este procedimiento al mismo tiempo que diluye eficazmente 5hmC, mantiene la señal 5meC. Por lo tanto, este procedimiento puede servir a dos fines (i) la identificación de 5hmC en ADN y (ii) la identificación de 5meC en ADN. La prueba de viabilidad del ensayo descrito queda demostrada en la Figura 4.

### Diseño experimental

El procedimiento que se describe en el presente documento se basa en la dilución sucesiva por PCR de la modificación 5hmC al mismo tiempo que se mantiene la modificación 5meC dado que DNMT1 no puede metilar transversalmente desde 5hmC y citosina; sin embargo, DNMT1 puede metilar ADN transversalmente desde 5meC.  
 15 Las Figuras 3. A y 4 demuestran que DNMT1 no puede catalizar la transferencia de un grupo metilo desde S-adenosil-metilmetionina cuando el sustrato de ADN es una citosina, 5hmC o b-glucosil-5hmC. Por lo tanto, es posible diluir la modificación 5hmC por PCR seguido del tratamiento con DNMT1 al mismo tiempo que se mantiene la modificación 5meC a través de múltiples rondas de PCR y el tratamiento con DNMT1. Cabe destacar que podrían utilizarse las enzimas de DNMT1, así como otras metil transferasas, para distinguir entre 5meC y 5hmC incluso en  
 20 los casos en los que estas enzimas sí que metilen transversalmente desde 5hmC y citosina siempre y cuando dichas enzimas tengan preferencia por hemi 5meC con respecto a 5hmC, b-glucosil-5hmC o citosina en los sitios CpG (Figura 3. B y C). Por otra parte, DNMT1 o la enzima metil transferasa pueden permitir la identificación tanto de 5meC como 5hmC en el ensayo descrito en el presente documento incluso aunque la transferencia de un grupo metilo desde S-adenosil-metilmetionina sea de una tasa mucho menor que 100%. El requisito para distinguir 5meC de 5hmC es que exista una preferencia de 5meC con respecto a 5hmC, b-glucosil-5hmC o citosina en sitios CpG  
 25 (Figura 3B y C).

El procedimiento de los autores de la invención aplica la conversión y secuenciación con bisulfito de la muestra "A", ADN sin tratar (Figura 1), que se utilizará como referencia ya que detectará el total tanto de 5meC como 5hmC. El procedimiento implica un ensayo de dilución de 5hmC, dilución de 5hmC en la agrupación total de fragmentos de ADN al mismo tiempo que mantiene 5meC. La dilución se consigue a través de rondas secuenciales de un ciclo de  
 30 amplificación por PCR (dilución) y tratamiento del ADN con ADN metil transferasa de mantenimiento, DNMT1, que mantiene enzimática y específicamente 5meC añadiendo un grupo metilo únicamente a la cadena sin metilar de los productos de PCR hemimetilados (en la Figura 1, se hace referencia a esta muestra como muestra "B"). Después de unas cuantas rondas de este ensayo, se aplicó la conversión y secuenciación con bisulfito de la muestra de ADN tratada, muestra B, en la que 5meC es ahora la única modificación presente (o la única modificación que altamente mantenida). Por lo tanto, todas las bases (o la mayoría de ellas) que se leen como C desde la muestra han debido ser protegidas contra la conversión, por 5meC y no por 5hmC. Al comparar "B" con la muestra de referencia "A" es posible detectar fácilmente todas las posiciones de base que contienen 5hmC.

Debe advertirse que este procedimiento al mismo tiempo que diluye eficazmente 5hmC, mantiene la señal 5meC. Por lo tanto, este procedimiento puede servir a dos fines (i) la identificación de 5hmC en ADN y (ii) la identificación de 5meC en ADN. La prueba de viabilidad del ensayo descrito queda demostrada en la Figura 4.

Por otra parte, debe advertirse que en el estado actual de la técnica los kit de conversión con bisulfito tienen limitaciones en cuanto a la sensibilidad. Por ejemplo, el kit MetilEasy Xceed (Human Genetic Signatures, cat. no. ME002) permite el análisis de 5meC de tan solo 8 células, pero no permite un análisis de una sola célula. El procedimiento descrito en el presente documento al mismo tiempo que diluye 5hmC y mantiene la señal de 5meC, permite una mayor sensibilidad de detección tanto para 5meC como 5hmC, con un claro potencial para el análisis de una sola célula como resultado de la amplificación por PCR de la muestra de ADN (con amplificación específica de gen o de todo el genoma).

## MATERIALES y PROCEDIMIENTOS

### 50 Sustratos

Sustratos de ADN creados por hibridación de los oligonucleótidos complementarios apropiados (véase Tabla 1 suplementaria) por calentamiento a 95 °C y refrigeración a 1 °C/min hasta que la reacción alcanzó 25 °C. En el ensayo de especificidad de metil transferasa se utilizaron oligonucleótidos creados por hibridación de 5hmC superior, 5meC superior o citosina superior con citosina inferior. Se creó el sustrato utilizado para simular una ronda de PCR seguido de tratamiento con DNMT1 por hibridación de 5hmC:C:5meC superior con sustrato inferior sin modificar. El sustrato utilizado para el ensayo completo se creó por hibridación de 5hmC:C:5meC superior con 5hmC:C:5meC inferior.

**Ensayo de especificidad de metil transferasa**

Se incubaron reacciones (50 µl) que contenían 100 ng de sustrato de ADN (citosina, hemi-5meC o hemi-5hmC), 50 mM Tris-HCl, 1 mM ditiotreitól, 1 mM EDTA pH 8,0, 5 % (v/v) glicerina, S-[metil-<sup>14</sup>C]-adenosil-L-metionina a 37 °C durante 30 minutos. Se terminaron las reacciones por adición de 200 µl de tampón TE. Se hizo precipitar el ADN de las reacciones con etanol y se lavó tres veces con etanol al 70 % enfriado con hielo. Se secó el aglomerado de ADN y se suspendió en 20 µl de tampón TE. Se transfirió toda la reacción a un vial de centelleo de 5 ml que contenía 2 ml de Ecosinct A. Se hizo el recuento por centelleo de las fracciones insolubles en ácido utilizando una ventana abierta durante 10 minutos.

**10 Conversión con bisulfito, clonación y secuenciación**

Se llevó a cabo la conversión con bisulfito de acuerdo con la guía de usuario del kit MetilEasy Xceed (Human Genetic Signatures, cat. no. ME002). Se realizó la clonación utilizando un kit TOPO TUn (Invitrogen, cat. no. K4595-40). Se llevó a cabo la secuenciación aplicando el procedimiento descrito por Sanger.

**Prueba del principio de ensayo de dilución de 5hmC**

15 Se utilizó un oligo de ADNbc de 112 pb que contenía tres sitios CpG en los que uno era hemi-5meC, el segundo CpG no contenía modificación y el tercer CpG era hemi-5hmC, como prueba del principio de ensayo de dilución de 5hmC. Se añadió el oligo (65 ng) a una mezcla de 5,0 µl 10x DNMT1-tampón (NEB), 2,5 µl de 3,2 mM SAM, 0,5 µl BSA (NEB, cat. no. B9001S) y 10 unidades de DNMT1 de ratón en un volumen total de 50 µl ajustado con MqH<sub>2</sub>O. Se incubaron las reacciones de DNMT1 en una termomezcladora a 37 °C, 600 rpm durante 4 h. A continuación, se purificaron los oligos de ADN con un equipo MinElute Reaction Cleanup Kit. Se llevó a cabo la conversión y  
20 secuenciación con bisulfito de la cadena inferior sin modificar del oligo antes y después del tratamiento con DNMT1.

**Ensayo de dilución de 5hmC**

Se utilizó un oligo de ADNbc de 112 pb que contenía tres sitios CpG, uno que tenía 5meC en ambas cadenas, un segundo que no tenía modificación y un tercero que tenía 5hmC en ambas cadenas para demostrar el ensayo de  
25 dilución de 5hmC. Para preparar oligonucleótidos hemi-modificados, se ajustó y se puso en marcha la PCR del siguiente modo: Se añadió el oligonucleótido (65 ng) a una mezcla de 4,0 µl de 5x Phusion HF-tampón, 1,6 µl de 2,5 mM dNTPs, 1 µl de cada uno de los cebadores directo e inverso 10 µM, 0,2 µl de polimerasa Phusion en un volumen total de 20 µl ajustado con MqH<sub>2</sub>O. Se llevó a cabo la fusión de las cadenas de ADN durante 3 min a 98 °C, seguido de la hibridación del cebador durante 2 min a 56 °C y la elongación durante 8 min a 72 °C. A continuación, se purificó el ADN con un kit MinElute Reaction Cleanup, se midió la concentración por fluorometría en un Qubit instrument y se llevó a cabo el tratamiento con DNMT1 de acuerdo con el siguiente protocolo: se añadió la cantidad total de oligo recuperado a una mezcla de 5,0 µl 10x DNMT1-tampón (NEB), 2,5 µl de 3,2 mM SAM, 0,5 µl BSA (NEB, cat. no. B9001S) y 10 unidades de DNMT1 de ratón en un volumen total de 50 µl ajustado con MqH<sub>2</sub>O. Se incubaron las reacciones de DNMT1 en una termomezcladora a 37 °C, 600 rpm durante 4 h. A continuación, se  
30 añadió 1 µl de Proteinase K (14-22 mg/ml) (Roche) y se continuó la incubación a 50 °C en una termomezcladora, 600 rpm durante 1 h. Después, se hicieron precipitar los oligos de ADN con etanol y se purificaron más con un kit MinElute Reaction Cleanup Kit. Se volvió a medir la concentración de ADN por fluorometría en un Qubit instrument. El protocolo descrito en esta sección se puede llevar a cabo una o más veces para dar como resultado un intervalo de dilución de 5hmC y conservación de 5meC.

**40 Ensayo de dilución/pérdida de 5hmC que permite la evaluación específica de cadena**

Se utilizó un oligo ADNbc de 112 pb ADNbc que contenía tres sitios CpG, uno que tenía 5meC en ambas cadenas, uno segundo que no tenía modificación y un tercero que tenía 5hmC en ambas cadenas para demostrar el ensayo de dilución de 5hmC (también denominado ensayo de pérdida de 5hmC y ensayo de extensión de cebador-(biotina)) valiéndose de la evaluación específica de cadena. Para preparar oligonucleótidos hemi-modificados, se ajustó la  
45 PCR de extensión de cebador específica de cadena y se puso en marcha del siguiente modo: se añadió el oligonucleótido (65 ng) a una mezcla de 4,0 µl de 5x Phusion HF-tampón, 1,6 µl de 2,5 mM dNTPs, 1 µl de únicamente uno entre los cebadores directo e inverso 10 µM que contenían una molécula biotina 5' /marcador, 0,2 µl de polimerasa Phusion en un volumen total de 20 µl ajustado con MqH<sub>2</sub>O. Se llevó a cabo el fundido de las cadenas de ADN durante 3 min a 98 °C, seguido de hibridación del cebador durante 2 min a 56 °C y elongación durante 8 min a 72 °C. A continuación, se purificó el ADN con perlas magnéticas revestidas con estreptavidina y se llevó a cabo el tratamiento con DNMT1 de acuerdo con el siguiente protocolo: se añadió la cantidad total de oligo recuperado a una mezcla de 5,0 µl 10x DNMT1-tampón (NEB), 2,5 µl de 3,2 mM SAM, 0,5 ml BSA (NEB, cat. no. B9001S) y 10 unidades de DNMT1 de ratón en un volumen total de 50 µl ajustado con MqH<sub>2</sub>O. Se incubaron las reacciones de DNMT1 en una termomezcladora a 37 °C, 600 rpm durante 4 h. A continuación, se purificaron los oligonucleótidos biotinizados utilizando perlas magnéticas de estreptavidina y se convirtieron con bisulfato, se utilizaron como  
55 matrices en la PCR y se secuenciaron.

## RESULTADOS

En la Figura 1 se representa un esbozo del procedimiento. Para demostrar la viabilidad y el éxito del procedimiento, se tratará de demostrar que (i) metil transferasas específicas modifican preferentemente sustratos de ADN hemi-5meC, (ii) que esta preferencia puede identificarse por secuenciación con bisulfito tras el tratamiento con la metil transferasa apropiada y (iii) la modificación 5hmC puede diluirse con sucesivas rondas de amplificación de ADN seguido del tratamiento con la ADN metilasa apropiada.

### DNMT1 de ratón, DNMT1 humana y metil transferasa de *M. Sssl* metilan preferentemente sustratos hemi-5meC

Se incubaron DNMT1 de ratón, DNMT1 de ser humano y *M. Sssl* metil transferasa con 100 ng sin modificar, hemi-5meC, hemi-5hmC o hemi-beta-glucosil-5hmC. DNMT1 de ratón fue capaz de catalizar la transferencia de un grupo metilo exclusivamente para el sustrato hemi-5meC, no presentando actividad en los demás sustratos (Figura 3A). DNMT1 humana presentó una preferencia enzimática para hemi-5meC al mismo tiempo que presentó una actividad limitada en los demás sustratos (Figure 3B). Finalmente, *M. Sssl* metilasa (*Spiroplasma sp.*) también presentó una preferencia para ADN que contenía hemi-5meC; (Figura 3C). De este resultado se llevó a la conclusión de que cualquiera de estas metil transferasas podía ser suficiente para el ensayo de dilución descrito en la Figura 1.

### DNMT1 de ratón tiene una sólida preferencia por hemi-5meC como sustrato

Se incubó un sustrato de ADNbc que contenía una hemi-5meC, citosina sin modificar y hemi-5hmC con DNMT1 de ratón en presencia de S-adenosil metilmationina. Se limpió el ADN y se sometió a secuenciación con bisulfito tal como se describe en "Materiales y Procedimientos." Después de la secuenciación con bisulfito fue posible demostrar que prácticamente toda (87,5 %) la hemi-5meC quedó metilada mientras que CpG sin modificar y hemi-5hmC quedaron sin modificar con DNMT1 de ratón (Figura 4). Dado que el sustrato utilizado para este ensayo simula la 5hmC total o ADN de 5hmC total tras una ronda de amplificación, se determinó que el ensayo podría funcionar utilizándolo con múltiples rondas de amplificación de ADN y tratamiento con DNMT1 de ratón.

### Sucesivas rondas de tratamiento con DNMT1 y amplificación por PCR puede diluir 5hmC al tiempo que se mantiene 5meC

Se amplificó un sustrato de ADNbc que contenía 5meC completo, CpG y 5hmC completo utilizando Taq o Phusion polimerasa seguido del tratamiento con DNMT1 de ratón. Se llevó a cabo este procedimiento tres veces, tal como se ha descrito en "materiales y procedimientos." En la Figura 6B se demuestra la dilución eficaz de 5hmC al tiempo que se mantiene 5meC. Se puede observar que antes de la dilución (Figura 6A) no se puede distinguir la identidad de 5hmC y 5meC; sin embargo, tras el tratamiento de dilución (Figura 6B); 5hmC y 5meC se pueden distinguir claramente ya que 5hmC está presente en una cantidad reducida en gran medida en comparación con 5meC.

### PCR de extensión de cebador específica de cadena combinada con el uso de cebadores biotinilizados permite la evaluación de la tasa de transferencia de DNMT1 de grupos metilo en sitios CpG en una cadena recién sintetizada en sitios transversalmente desde 5hmC, 5meC o C.

Se utilizó un sustrato de ADNbc que contenía 5meC completa, CpG y -5hmC completa como matriz para la PCR de extensión de cebador específica de cadena con cebadores que contenían un marcador de biotina 5' y seguido del tratamiento con DNMT1 de ratón. Se aislaron los oligonucleótidos recién sintetizados para asegurarse de que ninguna metilación /señal en los tres sitios CpG de análisis de la cadena seleccionada para el estudio provenía de la copia parental de la misma cadena. Esta estrategia permite la detección directa y la cuantificación del nivel de 5meC y 5hmC sin posteriores rondas de ensayo. Se puede utilizar el mismo ensayo que contiene el oligonucleótido descrito en el epígrafe materiales y procedimientos como referencia interna y control para ayudar a calcular el contenido de bases C modificadas en muestras de ADN genómico.

Un protocolo representativo del ensayo dependiente de metil transferasa (ensayo "B" en la Figura 11) es el siguiente:

- Extensión de cebador-Biotina: "Una ronda " PCR w/ cebador biotinilizado
- Productos de PCR agrupados (del oligo control muestra genómica)
- Purificación PCR MinElute
- Purificación Biotina-estreptavidina con perlas MyOne™ Streptavidin T1
- Tratamiento con DNMT1 sobre perlas, 0,6-1 µl de 0,5 mg/ml DNMT1, 1,6 mM SAM, 37 °C en 30 min.
- Lavados; y elución en 50 µl MQ-H<sub>2</sub>O 95 °C en 10 min.
- MinElute Reaction Cleanup (opcional)
- Tratamiento con bisulfito
- PCR Bisulfito
- Clonación TOPO TA, transformación, selección en placas LB amp X-gal
- Secuenciación

Tal como se muestra en la Figura 12, el ensayo HyLo dependiente de metil transferasa/DNMT1 identificó dos 5hmC CpGs en el gen *TRIM31* gene en ADN de cerebro humano. Se utilizó el ensayo esbozado en la Figura 11 con ADN

genómico con adiciones con un oligo de control que contenía sitios CpG conocidos para cada 5mC, C y 5hmC. Según esto, se pudo asegurar una cuantificación precisa de ADN genómico, ya que se llevó un seguimiento con el uso del oligo de la eficiencia de metil transferasa dentro de la muestra.

## Ejemplo 2

5 Se puede llevar a cabo la adición de un grupo químico a 5hmC, como glucosa para mejorar la relación de eficiencia de metil transferasa entre 5mC y 5hmC. Véase Figura 13. El bloqueo estérico de metil transferasa en la posición 5hmC modificada puede aprovechar el aumento de la solidez del ensayo dependiente de metil transferasa. En el presente documento se demuestra el efecto bloqueante de la adición de glucosa para 5hmC en un ensayo de metil transferasa radioactiva. Es posible bloquear tanto DNMT1 como M.SssI mediante la adición de un grupo químico con un mayor tamaño que el que pueda encajar en un bolsillo de metil transferasa, por ejemplo, mediante la adición de glucosa. Tal como se puede deducir de los datos y la demostración anterior de que el grupo carbono-5 de citosina de -CCCH<sub>3</sub> (tamaño de 6,1 Å) no encaja en el bolsillo de metil transferasa (Valinkluck y Sovers, *Cáncer Res*, 2007) se puede dar por supuesto que la adición de cualquier grupo químico a 5hmC que haga el grupo total demasiado grande para el bolsillo de metil transferasa es útil para aumentar la tasa de eficiencia de metil transferasa entre 5mC y 5hmC.

## REFERENCIAS

1. Penn, N.W. Modification of brain deoxyribonucleic acid base content with maturation in normal and malnourished rats. *Biochem J* 155, 709-712 (1976).
- 20 2. Cannon-Carlson, S.V., Gokhale, H. & Teebor, G.W. Purification and characterization of 5-hydroxymethyluracil-DNA glycosylase from calf thymus. Its possible role in the maintenance of methylated cytosine residues. *J Biol Chem* 264, 13306-13312 (1989).
3. Tahiliani, M. y col. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* (Nueva York, N. Y 324, 930-935 (2009).
- 25 4. Kriaucionis, S. & Heintz, N. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* (Nueva York, N.Y 324, 929-930 (2009).
5. Ito, S. y col. Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature* 466, 1129-1133 (2010).
6. Szwagierczak, A., Bultmann, S., Schmidt, C.S., Spada, F. & Leonhardt, H. Sensitive enzymatic quantification of 5-hydroxymethylcytosine in genomic DNA. *Nucleic acids research* 38, e181 (2010).
- 30 7. Ko, M. y col. Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. *Nature* 468, 839-843 (2010).
8. Guo, J.U., Su, Y., Zhong, C., Ming, G.L. & Song, H. Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. *Cell* 145, 423-434 (2011).
9. Wu, H. y col. Genome-wide analysis of 5-hydroxymethylcytosine distribution reveals its dual function in transcriptional regulation in mouse embryonic stem cells. *Genes & development* 25, 679-684 (2011).
- 35 10. Wu, H. y col. Dual functions of Tet1 in transcriptional regulation in mouse embryonic stem cells. *Nature* (2011).
11. Robertson, J., Robertson, A.B. & Klungland, A. The Presence of 5-hydroxymethylcytosine at the gene promotor and not in the gene body negatively regulates gene expression. *Biochem Biophys Res Comm* (2011).
- 40 12. Robertson, A.B. y col. A novel method for the efficient and selective identification of 5-hydroxymethylcytosine in genomic DNA. *Nucleic acids research* 39, e55 (2011).
13. Georgopoulos, C.P. & Revel, H.R. Studies with glucosyl transferase mutants of the T-even bacteriophages. *Virology* 44, 271-285 (1971).
- 45 14. Kornberg, S.R., Zimmerman, S.B. & Kornberg, A. Glucosylation of deoxyribonucleic acid by enzymes from bacteriophage-infected *Escherichia coli*. *J Biol Chem* 236, 1487-1493 (1961).
15. Gommers-Ampt, J.H. y col. beta-D-glucosyl-hydroxymethyluracil: a novel modified base present in the DNA of the parasitic protozoan *T. brucei*. *Cell* 75, 1129-1136 (1993).
16. Borst, P. & Sabatini, R. Base J: discovery, biosynthesis, and possible functions. *Annu Rev Microbiol* 62, 235-251 (2008).
- 50 17. van Leeuwen, F. y col. beta-D-glucosyl-hydroxymethyluracil is a conserved DNA modification in kinetoplastid protozoans and is abundant in their telomeres. *Proc Natl Acad Sci U S A* 95, 2366-2371 (1998).
18. Sabatini, R., Meeuwenoord, N., van Boom, J.H. & Borst, P. Recognition of base J in duplex DNA by J-binding protein. *J Biol Chem* 277, 958-966 (2002).
19. Cross, M. y col. The modified base J is the target for a novel DNA-binding protein in kinetoplastid protozoans. *EMBO J* 18, 6573-6581 (1999).
- 55 20. Grover, R.K. y col. O-glycoside orientation is an essential aspect of base J recognition by the kinetoplastid DNA binding protein JBP1. *Angewandte Chemie (International ed)* 46, 2839-2843 (2007).
21. Ficiz, G. y col. Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. *Nature* (2011).
- 60 22. Stroud, H., Feng, S., Morey Kinney, S., Pradhan, S. & Jacobsen, S.E. 5-hydroxymethylcytosine is associated with enhancers and gene bodies in human embryonic stem cells. *Genome Biol* 12, R54 (2011).
23. Pastor, W.A. y col. Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. *Nature* 473, 394-397 (2011).

24. Flusberg, B.A. y col. Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nat Methods* 7, 461-465 (2010).
25. Song, C.X., Yu, M., Dai, Q. & He, C. Detection of 5-hydroxymethylcytosine in a combined glycosylation restriction analysis (CGRA) using restriction enzyme Taq(alpha)I. *Bioorg Med Chem Lett* (2011).
- 5 26. Xu, S.Y., Corvaglia, A.R., Chan, S.H., Zheng, Y. & Linder, P. A type IV modification-dependent restriction enzyme SauUSI from *Staphylococcus aureus* subsp. *aureus* USA300. *Nucleic acids research* (2011).
27. Szwagierczak, A. y col. Characterization of PvuRts1I endonuclease as a tool to investigate genomic 5-hydroxymethylcytosine. *Nucleic acids research* (2011).
- 10 28. Song, C.X. y col. Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. *Nature biotechnology* 29, 68-72 (2011).
29. Ficiz, G. y col. Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. *Nature* 473, 398-402 (2011).
30. Nestor, C., Ruzov, A., Meehan, R. & Dunican, D. Enzymatic approaches and bisulfite sequencing cannot distinguish between 5-methylcytosine and 5-hydroxymethylcytosine in DNA. *Biotechniques* 48, 317-319 (2010).
- 15 31. Studier, F.W. Protein production by auto-induction in high density shaking cultures. *Protein Expr Purif* 41, 207-234 (2005).
32. Zhang, Y. y col. Model-based analysis of ChIP-Seq (MACS). *Genome Biol* 9, R137 (2008).

Si bien se ha descrito la invención en conexión con realizaciones preferentes concretas, debe entenderse que la invención tal como se reivindica no debe quedar indebidamente limitada con dichas realizaciones concretas.

20 LISTADO DE SECUENCIAS

- <110> Oslo universitetssykehus HF Robertson, Adam B. Dahl, John A. Klungland, Arne Ellevag, Linda
- <120> Métodos y kits para la detección de estado de metilación
- 25 <130> INVEN-32320/WO-1/ORD
- <150> US 61/570.066
- 30 <151> 13-12-2011
- <160> 9
- <170> PatentIn versión 3.5
- 35 <210> 1
- <211> 1627
- <212> PRT
- <213> *Mus musculus*
- 40 <400> 1

ES 2 669 214 T3

Met Pro Ala Arg Thr Ala Pro Ala Arg Val Pro Ala Leu Ala Ser Pro  
 1 5 10 15

Ala Gly Ser Leu Pro Asp His Val Arg Arg Arg Leu Lys Asp Leu Glu  
 20 25 30

Arg Asp Gly Leu Thr Glu Lys Glu Cys Val Arg Glu Lys Leu Asn Leu  
 35 40 45

Leu His Glu Phe Leu Gln Thr Glu Ile Lys Ser Gln Leu Cys Asp Leu  
 50 55 60

Glu Thr Lys Leu His Lys Glu Glu Leu Ser Glu Glu Gly Tyr Leu Ala  
 65 70 75 80

Lys Val Lys Ser Leu Leu Asn Lys Asp Leu Ser Leu Glu Asn Gly Thr  
 85 90 95

His Thr Leu Thr Gln Lys Ala Asn Gly Cys Pro Ala Asn Gly Ser Arg  
 100 105 110

Pro Thr Trp Arg Ala Glu Met Ala Asp Ser Asn Arg Ser Pro Arg Ser  
 115 120 125

Arg Pro Lys Pro Arg Gly Pro Arg Arg Ser Lys Ser Asp Ser Asp Thr  
 130 135 140

Leu Cys Lys Asp Thr Arg His Thr Ala Val Glu Thr Ser Pro Ser Ser





ES 2 669 214 T3

Ala Met Lys Arg Arg Arg Cys Gly Val Cys Glu Val Cys Gln Gln Pro  
660 665 670

Glu Cys Gly Lys Cys Lys Ala Cys Lys Asp Met Val Lys Phe Gly Gly  
675 680 685

Thr Gly Arg Ser Lys Gln Ala Cys Leu Lys Arg Arg Cys Pro Asn Leu  
690 695 700

Ala Val Lys Glu Ala Asp Asp Asp Glu Glu Ala Asp Asp Asp Val Ser  
705 710 715 720

Glu Met Pro Ser Pro Lys Lys Leu His Gln Gly Lys Lys Lys Lys Gln  
725 730 735

Asn Lys Asp Arg Ile Ser Trp Leu Gly Gln Pro Met Lys Ile Glu Glu  
740 745 750

Asn Arg Thr Tyr Tyr Gln Lys Val Ser Ile Asp Glu Glu Met Leu Glu  
755 760 765

Val Gly Asp Cys Val Ser Val Ile Pro Asp Asp Ser Ser Lys Pro Leu  
770 775 780

Tyr Leu Ala Arg Val Thr Ala Leu Trp Glu Asp Lys Asn Gly Gln Met  
785 790 795 800

Met Phe His Ala His Trp Phe Cys Ala Gly Thr Asp Thr Val Leu Gly  
805 810 815

Ala Thr Ser Asp Pro Leu Glu Leu Phe Leu Val Gly Glu Cys Glu Asn  
820 825 830

Met Gln Leu Ser Tyr Ile His Ser Lys Val Lys Val Ile Tyr Lys Ala  
835 840 845

Pro Ser Glu Asn Trp Ala Met Glu Gly Gly Thr Asp Pro Glu Thr Thr  
850 855 860

Leu Pro Gly Ala Glu Asp Gly Lys Thr Tyr Phe Phe Gln Leu Trp Tyr  
865 870 875 880

Asn Gln Glu Tyr Ala Arg Phe Glu Ser Pro Pro Lys Thr Gln Pro Thr  
885 890 895

Glu Asp Asn Lys His Lys Phe Cys Leu Ser Cys Ile Arg Leu Ala Glu  
900 905 910

ES 2 669 214 T3

Leu Arg Gln Lys Glu Met Pro Lys Val Leu Glu Gln Ile Glu Glu Val  
 915 920 925  
 Asp Gly Arg Val Tyr Cys Ser Ser Ile Thr Lys Asn Gly Val Val Tyr  
 930 935 940  
 Arg Leu Gly Asp Ser Val Tyr Leu Pro Pro Glu Ala Phe Thr Phe Asn  
 945 950 955 960  
 Ile Lys Val Ala Ser Pro Val Lys Arg Pro Lys Lys Asp Pro Val Asn  
 965 970 975  
 Glu Thr Leu Tyr Pro Glu His Tyr Arg Lys Tyr Ser Asp Tyr Ile Lys  
 980 985 990  
 Gly Ser Asn Leu Asp Ala Pro Glu Pro Tyr Arg Ile Gly Arg Ile Lys  
 995 1000 1005  
 Glu Ile His Cys Gly Lys Lys Lys Gly Lys Val Asn Glu Ala Asp  
 1010 1015 1020  
 Ile Lys Leu Arg Leu Tyr Lys Phe Tyr Arg Pro Glu Asn Thr His  
 1025 1030 1035  
 Arg Ser Tyr Asn Gly Ser Tyr His Thr Asp Ile Asn Met Leu Tyr  
 1040 1045 1050  
 Trp Ser Asp Glu Glu Ala Val Val Asn Phe Ser Asp Val Gln Gly  
 1055 1060 1065  
 Arg Cys Thr Val Glu Tyr Gly Glu Asp Leu Leu Glu Ser Ile Gln  
 1070 1075 1080  
 Asp Tyr Ser Gln Gly Gly Pro Asp Arg Phe Tyr Phe Leu Glu Ala  
 1085 1090 1095  
 Tyr Asn Ser Lys Thr Lys Asn Phe Glu Asp Pro Pro Asn His Ala  
 1100 1105 1110  
 Arg Ser Pro Gly Asn Lys Gly Lys Gly Lys Gly Lys Gly Lys Gly  
 1115 1120 1125  
 Lys Gly Lys His Gln Val Ser Glu Pro Lys Glu Pro Glu Ala Ala  
 1130 1135 1140  
 Ile Lys Leu Pro Lys Leu Arg Thr Leu Asp Val Phe Ser Gly Cys

ES 2 669 214 T3

1145						1150						1155			
Gly	Gly	Leu	Ser	Glu	Gly	Phe	His	Gln	Ala	Gly	Ile	Ser	Glu	Thr	
1160						1165					1170				
Leu	Trp	Ala	Ile	Glu	Met	Trp	Asp	Pro	Ala	Ala	Gln	Ala	Phe	Arg	
1175						1180					1185				
Leu	Asn	Asn	Pro	Gly	Thr	Thr	Val	Phe	Thr	Glu	Asp	Cys	Asn	Val	
1190						1195					1200				
Leu	Leu	Lys	Leu	Val	Met	Ala	Gly	Glu	Val	Thr	Asn	Ser	Leu	Gly	
1205						1210					1215				
Gln	Arg	Leu	Pro	Gln	Lys	Gly	Asp	Val	Glu	Met	Leu	Cys	Gly	Gly	
1220						1225					1230				
Pro	Pro	Cys	Gln	Gly	Phe	Ser	Gly	Met	Asn	Arg	Phe	Asn	Ser	Arg	
1235						1240					1245				
Thr	Tyr	Ser	Lys	Phe	Lys	Asn	Ser	Leu	Val	Val	Ser	Phe	Leu	Ser	
1250						1255					1260				
Tyr	Cys	Asp	Tyr	Tyr	Arg	Pro	Arg	Phe	Phe	Leu	Leu	Glu	Asn	Val	
1265						1270					1275				
Arg	Asn	Phe	Val	Ser	Tyr	Arg	Arg	Ser	Met	Val	Leu	Lys	Leu	Thr	
1280						1285					1290				
Leu	Arg	Cys	Leu	Val	Arg	Met	Gly	Tyr	Gln	Cys	Thr	Phe	Gly	Val	
1295						1300					1305				
Leu	Gln	Ala	Gly	Gln	Tyr	Gly	Val	Ala	Gln	Thr	Arg	Arg	Arg	Ala	
1310						1315					1320				
Ile	Ile	Leu	Ala	Ala	Ala	Pro	Gly	Glu	Lys	Leu	Pro	Leu	Phe	Pro	
1325						1330					1335				
Glu	Pro	Leu	His	Val	Phe	Ala	Pro	Arg	Ala	Cys	Gln	Leu	Ser	Val	
1340						1345					1350				
Val	Val	Asp	Asp	Lys	Lys	Phe	Val	Ser	Asn	Ile	Thr	Arg	Leu	Ser	
1355						1360					1365				
Ser	Gly	Pro	Phe	Arg	Thr	Ile	Thr	Val	Arg	Asp	Thr	Met	Ser	Asp	
1370						1375					1380				

ES 2 669 214 T3

Leu Pro Glu Ile Gln Asn Gly Ala Ser Asn Ser Glu Ile Pro Tyr  
1385 1390 1395

Asn Gly Glu Pro Leu Ser Trp Phe Gln Arg Gln Leu Arg Gly Ser  
1400 1405 1410

His Tyr Gln Pro Ile Leu Arg Asp His Ile Cys Lys Asp Met Ser  
1415 1420 1425

Pro Leu Val Ala Ala Arg Met Arg His Ile Pro Leu Phe Pro Gly  
1430 1435 1440

Ser Asp Trp Arg Asp Leu Pro Asn Ile Gln Val Arg Leu Gly Asp  
1445 1450 1455

Gly Val Ile Ala His Lys Leu Gln Tyr Thr Phe His Asp Val Lys  
1460 1465 1470

Asn Gly Tyr Ser Ser Thr Gly Ala Leu Arg Gly Val Cys Ser Cys  
1475 1480 1485

Ala Glu Gly Lys Ala Cys Asp Pro Glu Ser Arg Gln Phe Ser Thr  
1490 1495 1500

Leu Ile Pro Trp Cys Leu Pro His Thr Gly Asn Arg His Asn His  
1505 1510 1515

Trp Ala Gly Leu Tyr Gly Arg Leu Glu Trp Asp Gly Phe Phe Ser  
1520 1525 1530

Thr Thr Val Thr Asn Pro Glu Pro Met Gly Lys Gln Gly Arg Val  
1535 1540 1545

Leu His Pro Glu Gln His Arg Val Val Ser Val Arg Glu Cys Ala  
1550 1555 1560

Arg Ser Gln Gly Phe Pro Asp Ser Tyr Arg Phe Phe Gly Asn Ile  
1565 1570 1575

Leu Asp Arg His Arg Gln Val Gly Asn Ala Val Pro Pro Pro Leu  
1580 1585 1590

Ala Lys Ala Ile Gly Leu Glu Ile Lys Leu Cys Leu Leu Ser Ser  
1595 1600 1605

Ala Arg Glu Ser Ala Ser Ala Ala Val Lys Ala Lys Glu Glu Ala  
1610 1615 1620

Ala Thr Lys Asp  
1625

ES 2 669 214 T3

<210> 2  
 <211> 1632  
 <212> PRT  
 <213> *Homo sapiens*

5

<400> 2

```

Met  Pro  Ala  Arg  Thr  Ala  Pro  Ala  Arg  Val  Pro  Thr  Leu  Ala  Val  Pro
 1          5          10          15

Ala  Ile  Ser  Leu  Pro  Asp  Asp  Val  Arg  Arg  Arg  Leu  Lys  Asp  Leu  Glu
 20          25          30

Arg  Asp  Ser  Leu  Thr  Glu  Lys  Glu  Cys  Val  Lys  Glu  Lys  Leu  Asn  Leu
 35          40          45

Leu  His  Glu  Phe  Leu  Gln  Thr  Glu  Ile  Lys  Asn  Gln  Leu  Cys  Asp  Leu
 50          55          60

Glu  Thr  Lys  Leu  Arg  Lys  Glu  Glu  Leu  Ser  Glu  Glu  Gly  Tyr  Leu  Ala
 65          70          75          80

Lys  Val  Lys  Ser  Leu  Leu  Asn  Lys  Asp  Leu  Ser  Leu  Glu  Asn  Gly  Ala
 85          90          95

His  Ala  Tyr  Asn  Arg  Glu  Val  Asn  Gly  Arg  Leu  Glu  Asn  Gly  Asn  Gln
 100         105         110

Ala  Arg  Ser  Glu  Ala  Arg  Arg  Val  Gly  Met  Ala  Asp  Ala  Asn  Ser  Pro
 115         120         125

Pro  Lys  Pro  Leu  Ser  Lys  Pro  Arg  Thr  Pro  Arg  Arg  Ser  Lys  Ser  Asp
 130         135         140

Gly  Glu  Ala  Lys  Arg  Ser  Arg  Asp  Pro  Pro  Ala  Ser  Ala  Ser  Gln  Val
 145         150         155         160

Thr  Gly  Ile  Arg  Ala  Glu  Pro  Ser  Pro  Ser  Pro  Arg  Ile  Thr  Arg  Lys
 165         170         175

Ser  Thr  Arg  Gln  Thr  Thr  Ile  Thr  Ser  His  Phe  Ala  Lys  Gly  Pro  Ala
 180         185         190

Lys  Arg  Lys  Pro  Gln  Glu  Glu  Ser  Glu  Arg  Ala  Lys  Ser  Asp  Glu  Ser
 195         200         205
    
```

ES 2 669 214 T3

Ile Lys Glu Glu Asp Lys Asp Gln Asp Glu Lys Arg Arg Arg Val Thr  
 210 215 220

Ser Arg Glu Arg Val Ala Arg Pro Leu Pro Ala Glu Glu Pro Glu Arg  
 225 230 235 240

Ala Lys Ser Gly Thr Arg Thr Glu Lys Glu Glu Glu Arg Asp Glu Lys  
 245 250 255

Glu Glu Lys Arg Leu Arg Ser Gln Thr Lys Glu Pro Thr Pro Lys Gln  
 260 265 270

Lys Leu Lys Glu Glu Pro Asp Arg Glu Ala Arg Ala Gly Val Gln Ala  
 275 280 285

Asp Glu Asp Glu Asp Gly Asp Glu Lys Asp Glu Lys Lys His Arg Ser  
 290 295 300

Gln Pro Lys Asp Leu Ala Ala Lys Arg Arg Pro Glu Glu Lys Glu Pro  
 305 310 315 320

Glu Lys Val Asn Pro Gln Ile Ser Asp Glu Lys Asp Glu Asp Glu Lys  
 325 330 335

Glu Glu Lys Arg Arg Lys Thr Thr Pro Lys Glu Pro Thr Glu Lys Lys  
 340 345 350

Met Ala Arg Ala Lys Thr Val Met Asn Ser Lys Thr His Pro Pro Lys  
 355 360 365

Cys Ile Gln Cys Gly Gln Tyr Leu Asp Asp Pro Asp Leu Lys Tyr Gly  
 370 375 380

Gln His Pro Pro Asp Ala Val Asp Glu Pro Gln Met Leu Thr Asn Glu  
 385 390 395 400

Lys Leu Ser Ile Phe Asp Ala Asn Glu Ser Gly Phe Glu Ser Tyr Glu  
 405 410 415

Ala Leu Pro Gln His Lys Leu Thr Cys Phe Ser Val Tyr Cys Lys His  
 420 425 430

Gly His Leu Cys Pro Ile Asp Thr Gly Leu Ile Glu Lys Asn Ile Glu  
 435 440 445

Leu Phe Phe Ser Gly Ser Ala Lys Pro Ile Tyr Asp Asp Asp Pro Ser  
 450 455 460

ES 2 669 214 T3

Leu Glu Gly Gly Val Asn Gly Lys Asn Leu Gly Pro Ile Asn Glu Trp  
 465 470 475 480  
 Trp Ile Thr Gly Phe Asp Gly Gly Glu Lys Ala Leu Ile Gly Phe Ser  
 485 490 495  
 Thr Ser Phe Ala Glu Tyr Ile Leu Met Asp Pro Ser Pro Glu Tyr Ala  
 500 505 510  
 Pro Ile Phe Gly Leu Met Gln Glu Lys Ile Tyr Ile Ser Lys Ile Val  
 515 520 525  
 Val Glu Phe Leu Gln Ser Asn Ser Asp Ser Thr Tyr Glu Asp Leu Ile  
 530 535 540  
 Asn Lys Ile Glu Thr Thr Val Pro Pro Ser Gly Leu Asn Leu Asn Arg  
 545 550 555 560  
 Phe Thr Glu Asp Ser Leu Leu Arg His Ala Gln Phe Val Val Glu Gln  
 565 570 575  
 Val Glu Ser Tyr Asp Glu Ala Gly Asp Ser Asp Glu Gln Pro Ile Phe  
 580 585 590  
 Leu Thr Pro Cys Met Arg Asp Leu Ile Lys Leu Ala Gly Val Thr Leu  
 595 600 605  
 Gly Gln Arg Arg Ala Gln Ala Arg Arg Gln Thr Ile Arg His Ser Thr  
 610 615 620  
 Arg Glu Lys Asp Arg Gly Pro Thr Lys Ala Thr Thr Thr Lys Leu Val  
 625 630 635 640  
 Tyr Gln Ile Phe Asp Thr Phe Phe Ala Glu Gln Ile Glu Lys Asp Asp  
 645 650 655  
 Arg Glu Asp Lys Glu Asn Ala Phe Lys Arg Arg Arg Cys Gly Val Cys  
 660 665 670  
 Glu Val Cys Gln Gln Pro Glu Cys Gly Lys Cys Lys Ala Cys Lys Asp  
 675 680 685  
 Met Val Lys Phe Gly Gly Ser Gly Arg Ser Lys Gln Ala Cys Gln Glu  
 690 695 700  
 Arg Arg Cys Pro Asn Met Ala Met Lys Glu Ala Asp Asp Asp Glu Glu





ES 2 669 214 T3

Ala Phe Thr Phe Asn Ile Lys Leu Ser Ser Pro Val Lys Arg Pro Arg  
 965 970 975

Lys Glu Pro Val Asp Glu Asp Leu Tyr Pro Glu His Tyr Arg Lys Tyr  
 980 985 990

Ser Asp Tyr Ile Lys Gly Ser Asn Leu Asp Ala Pro Glu Pro Tyr Arg  
 995 1000 1005

Ile Gly Arg Ile Lys Glu Ile Phe Cys Pro Lys Lys Ser Asn Gly  
 1010 1015 1020

Arg Pro Asn Glu Thr Asp Ile Lys Ile Arg Val Asn Lys Phe Tyr  
 1025 1030 1035

Arg Pro Glu Asn Thr His Lys Ser Thr Pro Ala Ser Tyr His Ala  
 1040 1045 1050

Asp Ile Asn Leu Leu Tyr Trp Ser Asp Glu Glu Ala Val Val Asp  
 1055 1060 1065

Phe Lys Ala Val Gln Gly Arg Cys Thr Val Glu Tyr Gly Glu Asp  
 1070 1075 1080

Leu Pro Glu Cys Val Gln Val Tyr Ser Met Gly Gly Pro Asn Arg  
 1085 1090 1095

Phe Tyr Phe Leu Glu Ala Tyr Asn Ala Lys Ser Lys Ser Phe Glu  
 1100 1105 1110

Asp Pro Pro Asn His Ala Arg Ser Pro Gly Asn Lys Gly Lys Gly  
 1115 1120 1125

Lys Gly Lys Gly Lys Gly Lys Pro Lys Ser Gln Ala Cys Glu Pro  
 1130 1135 1140

Ser Glu Pro Glu Ile Glu Ile Lys Leu Pro Lys Leu Arg Thr Leu  
 1145 1150 1155

Asp Val Phe Ser Gly Cys Gly Gly Leu Ser Glu Gly Phe His Gln  
 1160 1165 1170

Ala Gly Ile Ser Asp Thr Leu Trp Ala Ile Glu Met Trp Asp Pro  
 1175 1180 1185

Ala Ala Gln Ala Phe Arg Leu Asn Asn Pro Gly Ser Thr Val Phe  
 1190 1195 1200

ES 2 669 214 T3

Thr Glu Asp Cys Asn Ile Leu Leu Lys Leu Val Met Ala Gly Glu  
 1205 1210 1215  
 Thr Thr Asn Ser Arg Gly Gln Arg Leu Pro Gln Lys Gly Asp Val  
 1220 1225 1230  
 Glu Met Leu Cys Gly Gly Pro Pro Cys Gln Gly Phe Ser Gly Met  
 1235 1240 1245  
 Asn Arg Phe Asn Ser Arg Thr Tyr Ser Lys Phe Lys Asn Ser Leu  
 1250 1255 1260  
 Val Val Ser Phe Leu Ser Tyr Cys Asp Tyr Tyr Arg Pro Arg Phe  
 1265 1270 1275  
 Phe Leu Leu Glu Asn Val Arg Asn Phe Val Ser Phe Lys Arg Ser  
 1280 1285 1290  
 Met Val Leu Lys Leu Thr Leu Arg Cys Leu Val Arg Met Gly Tyr  
 1295 1300 1305  
 Gln Cys Thr Phe Gly Val Leu Gln Ala Gly Gln Tyr Gly Val Ala  
 1310 1315 1320  
 Gln Thr Arg Arg Arg Ala Ile Ile Leu Ala Ala Ala Pro Gly Glu  
 1325 1330 1335  
 Lys Leu Pro Leu Phe Pro Glu Pro Leu His Val Phe Ala Pro Arg  
 1340 1345 1350  
 Ala Cys Gln Leu Ser Val Val Val Asp Asp Lys Lys Phe Val Ser  
 1355 1360 1365  
 Asn Ile Thr Arg Leu Ser Ser Gly Pro Phe Arg Thr Ile Thr Val  
 1370 1375 1380  
 Arg Asp Thr Met Ser Asp Leu Pro Glu Val Arg Asn Gly Ala Ser  
 1385 1390 1395  
 Ala Leu Glu Ile Ser Tyr Asn Gly Glu Pro Gln Ser Trp Phe Gln  
 1400 1405 1410  
 Arg Gln Leu Arg Gly Ala Gln Tyr Gln Pro Ile Leu Arg Asp His  
 1415 1420 1425  
 Ile Cys Lys Asp Met Ser Ala Leu Val Ala Ala Arg Met Arg His  
 1430 1435 1440

ES 2 669 214 T3

Ile Pro Leu Ala Pro Gly Ser Asp Trp Arg Asp Leu Pro Asn Ile  
 1445 1450 1455

Glu Val Arg Leu Ser Asp Gly Thr Met Ala Arg Lys Leu Arg Tyr  
 1460 1465 1470

Thr His His Asp Arg Lys Asn Gly Arg Ser Ser Ser Gly Ala Leu  
 1475 1480 1485

Arg Gly Val Cys Ser Cys Val Glu Ala Gly Lys Ala Cys Asp Pro  
 1490 1495 1500

Ala Ala Arg Gln Phe Asn Thr Leu Ile Pro Trp Cys Leu Pro His  
 1505 1510 1515

Thr Gly Asn Arg His Asn His Trp Ala Gly Leu Tyr Gly Arg Leu  
 1520 1525 1530

Glu Trp Asp Gly Phe Phe Ser Thr Thr Val Thr Asn Pro Glu Pro  
 1535 1540 1545

Met Gly Lys Gln Gly Arg Val Leu His Pro Glu Gln His Arg Val  
 1550 1555 1560

Val Ser Val Arg Glu Cys Ala Arg Ser Gln Gly Phe Pro Asp Thr  
 1565 1570 1575

Tyr Arg Leu Phe Gly Asn Ile Leu Asp Lys His Arg Gln Val Gly  
 1580 1585 1590

Asn Ala Val Pro Pro Pro Leu Ala Lys Ala Ile Gly Leu Glu Ile  
 1595 1600 1605

Lys Leu Cys Met Leu Ala Lys Ala Arg Glu Ser Ala Ser Ala Lys  
 1610 1615 1620

Ile Lys Glu Glu Glu Ala Ala Lys Asp  
 1625 1630

<210> 3  
 <211> 386  
 <212> PRT  
 <213> Spiroplasma MQ-1

<400> 3

Met Ser Lys Val Glu Asn Lys Thr Lys Lys Leu Arg Val Phe Glu Ala  
 1 5 10 15

5

10

ES 2 669 214 T3

Phe Ala Gly Ile Gly Ala Gln Arg Lys Ala Leu Glu Lys Val Arg Lys  
 20 25 30  
 Asp Glu Tyr Glu Ile Val Gly Leu Ala Glu Trp Tyr Val Pro Ala Ile  
 35 40 45  
 Val Met Tyr Gln Ala Ile His Asn Asn Phe His Thr Lys Leu Glu Tyr  
 50 55 60  
 Lys Ser Val Ser Arg Glu Glu Met Ile Asp Tyr Leu Glu Asn Lys Thr  
 65 70 75 80  
 Leu Ser Trp Asn Ser Lys Asn Pro Val Ser Asn Gly Tyr Trp Lys Arg  
 85 90 95  
 Lys Lys Asp Asp Glu Leu Lys Ile Ile Tyr Asn Ala Ile Lys Leu Ser  
 100 105 110  
 Glu Lys Glu Gly Asn Ile Phe Asp Ile Arg Asp Leu Tyr Lys Arg Thr  
 115 120 125  
 Leu Lys Asn Ile Asp Leu Leu Thr Tyr Ser Phe Pro Cys Gln Asp Leu  
 130 135 140  
 Ser Gln Gln Gly Ile Gln Lys Gly Met Lys Arg Gly Ser Gly Thr Arg  
 145 150 155 160  
 Ser Gly Leu Leu Trp Glu Ile Glu Arg Ala Leu Asp Ser Thr Glu Lys  
 165 170 175  
 Asn Asp Leu Pro Lys Tyr Leu Leu Met Glu Asn Val Gly Ala Leu Leu  
 180 185 190  
 His Lys Lys Asn Glu Glu Glu Leu Asn Gln Trp Lys Gln Lys Leu Glu  
 195 200 205  
 Ser Leu Gly Tyr Gln Asn Ser Ile Glu Val Leu Asn Ala Ala Asp Phe  
 210 215 220  
 Gly Ser Ser Gln Ala Arg Arg Arg Val Phe Met Ile Ser Thr Leu Asn  
 225 230 235 240  
 Glu Phe Val Glu Leu Pro Lys Gly Asp Lys Lys Pro Lys Ser Ile Lys  
 245 250 255  
 Lys Val Leu Asn Lys Ile Val Ser Glu Lys Asp Ile Leu Asn Asn Leu

ES 2 669 214 T3

```

                260                      265                      270

Leu Lys Tyr Asn Leu Thr Glu Phe Lys Lys Thr Lys Ser Asn Ile Asn
   275                      280                      285

Lys Ala Ser Leu Ile Gly Tyr Ser Lys Phe Asn Ser Glu Gly Tyr Val
   290                      295                      300

Tyr Asp Pro Glu Phe Thr Gly Pro Thr Leu Thr Ala Ser Gly Ala Asn
   305                      310                      315                      320

Ser Arg Ile Lys Ile Lys Asp Gly Ser Asn Ile Arg Lys Met Asn Ser
   325                      330                      335

Asp Glu Thr Phe Leu Tyr Ile Gly Phe Asp Ser Gln Asp Gly Lys Arg
   340                      345                      350

Val Asn Glu Ile Glu Phe Leu Thr Glu Asn Gln Lys Ile Phe Val Cys
   355                      360                      365

Gly Asn Ser Ile Ser Val Glu Val Leu Glu Ala Ile Ile Asp Lys Ile
   370                      375                      380

Gly Gly
   385

```

```

5  <210> 4
   <211> 12
   <212> ADN
   <213> Secuencia artificial

10 <220>
   <223> Sintético

15 <220>
   <221> característica_mis
   <222> (6)..(6)
   <223> El resto en esta posición es 5-metilcitosine

20 <220>
   <221> característica_mis
   <222> (10)..(10)
   <223> El resto en esta posición es 5-hidroximetilcitosine

   <400> 4
   tcgatcgatc gt      12

25 <210> 5
   <211> 12
   <212> ADN
   <213> Secuencia artificial

30 <220>
   <223> Sintético

```

5  
 <220>  
 <221> característica\_mis  
 <222> (2)..(2)  
 <223> El resto en esta posición es 5-hidroximetilcitosine

10  
 <220>  
 <221> característica\_mis  
 <222> (6)..(6)  
 <223> El resto en esta posición es 5-metilcitosine

<400> 5  
 acgatcgatc ga 12

15  
 <210> 6  
 <211> 12  
 <212> ADN  
 <213> Secuencia artificial

20  
 <220>  
 <223> Sintético

<400> 6  
 acgatcgatc ga 12

25  
 <210> 7  
 <211> 12  
 <212> ADN  
 <213> Secuencia artificial

30  
 <220>  
 <223> Sintético

35  
 <220>  
 <221> característica\_mis  
 <222> (6)..(6)  
 <223> El resto en esta posición es 5-metilcitosine

<400> 7  
 acgatcgatc ga 12

40

45  
 <210> 8  
 <211> 12  
 <212> ADN  
 <213> Secuencia artificial

<220>  
 <223> Sintético

50  
 <220>  
 <221> característica\_mis  
 <222> (6)..(6)  
 <223> El resto en esta posición es 5-metilcitosine

55  
 <220>  
 <221> característica\_mis  
 <222> (10)..(10)  
 <223> El resto en esta posición es 5-hidroximetilcitosine

<400> 8  
 tugatcgatc gt 12

60  
 <210> 9  
 <211> 12

# ES 2 669 214 T3

<212> ADN  
<213> Secuencia artificial

5 <220>  
<223> Sintético

<400> 9  
ttgatcgatc gt 12

## REIVINDICACIONES

1. Procedimiento de detección de restos de citosina 5-metilada y 5-hidroximetilada en una muestra de ácido nucleico que comprende:

- 5 a) replicación de dicha muestra de ácido nucleico en condiciones tales que se mantienen los restos de citosina 5-metilada y se diluyen dichos restos de citosina hidroximetilada;  
 b) tratamiento de dicha muestra de ácido nucleico replicada para convertir restos de citosina sin modificar en restos de uracilo o timidina;

10 en el que dicha muestra de ácido nucleico se divide en al menos una primera y una segunda porción y dichas etapas de replicación y tratamiento se realizan en dicha primera porción, y la comparación de la secuencia de dicha primera porción de ácido nucleico con la secuencia de dicha segunda porción de ácido nucleico, en la que se identifican dichos restos de citosina hidroximetilada como restos que son leídos por secuenciación como restos de uracilo o timidina en dicha primera porción de ácido nucleico y como restos de citosina en la posición correspondiente en dicha segunda porción de ácido nucleico y en la que se identifican restos de citosina 5-metilada como restos que se leen como restos de citosina en ambas porciones mencionadas de ácido nucleico primera y segunda.

15 2. Procedimiento de la reivindicación 1, en el que dicho tratamiento de dichas porciones primera y segunda para convertir restos de citosina sin modificar en restos timidina comprende además el tratamiento de dichas porciones de ácido nucleico primera y segunda con bisulfito para convertir restos de citosina sin modificar en restos de uracilo y la replicación de dichas porciones de ácido nucleico primera y segunda con una polimerasa para convertir dichos restos de uracilo en restos timidina.

20 3. Procedimiento de la reivindicación 1, en el que dicha replicación de dicha primera porción comprende además:

- a1) replicación de dicho ácido nucleico con un cebador marcado para proporcionar ácido nucleico replicado marcado;  
 a2) tratamiento de dichas cadenas de ácido nucleico replicadas marcadas con una ADN metil transferasa para proporcionar ácido nucleico replicado modificado con 5- metilcitosina marcado;  
 25 a3) aislamiento de dicho ácido nucleico replicado modificado con 5-metilcitosina marcado;  
 a4) tratamiento de dicho ácido nucleico replicado modificado con 5-metilcitosina marcado aislado con bisulfito para convertir restos de citosina sin modificar en restos de uracilo; y  
 a5) replicación de dicho ácido nucleico tratado con bisulfito marcado aislado con una polimerasa para proporcionar una primera porción de ácido nucleico tratada con bisulfito.

30 4. Procedimiento de cualquiera de las reivindicaciones 1 a 3, en el que dichos restos de citosina hidroximetilada se seleccionan del grupo que consiste en 5-hidroximetil citosina y b-glu-5-hidroximetil citosina.

5. Procedimiento de la reivindicación 1 de detección de restos de citosina metilada e hidroximetilada en una muestra de ácido nucleico que comprende:

- 35 a) división de dicha muestra en al menos una primera y una segunda porciones sin tratar;  
 b) replicación de dicha primera porción con un cebador marcado y una polimerasa para proporcionar ácido nucleico parental y replicado marcado;  
 c) tratamiento de dichas cadenas de ácido nucleico parental y replicado marcado con una ADN metil transferasa para proporcionar ácido nucleico replicado modificado con 5-metilcitosina marcado;  
 40 d) aislamiento de dicho ácido nucleico replicado modificado con 5-metilcitosina marcado;  
 e) tratamiento de dicho ácido nucleico replicado modificado con 5-metilcitosina marcado aislado con bisulfito para convertir restos de citosina sin modificar en restos de uracilo;  
 f) replicación de dicho ácido nucleico tratado con bisulfito marcado aislado con una polimerasa para proporcionar una primera porción de ácido nucleico tratada con bisulfito;  
 45 g) secuenciación de dicha primera porción de ácido nucleico tratada con bisulfito;  
 h) tratamiento de dicha segunda porción con bisulfito para convertir restos de citosina sin modificar en restos de uracilo;  
 i) replicación de dicho ácido nucleico tratado con bisulfito con una polimerasa para proporcionar una segunda porción de ácido nucleico tratada con bisulfito;  
 50 j) secuenciación de dicha segunda porción de ácido nucleico tratada con bisulfito; y  
 k) comparación de la secuencia de dicha primera porción de ácido nucleico tratada con bisulfito con la secuencia de dicha segunda porción tratada con bisulfito, en la que se identifican los restos de citosina 5-hidroximetilada como restos que se leen por secuenciación como un resto de uracilo o timidina en dicha primera porción de ácido nucleico tratada con bisulfito y como un resto citosina en la posición correspondiente en dicha segunda porción  
 55 de ácido nucleico tratada con bisulfito y en la que se identifican los restos de citosina 5-metilada como restos que se leen como restos de citosina en dichas porciones primera y segunda tratadas con bisulfito.

6. Procedimiento de la reivindicación 4 o 5, en el que dicho cebador marcado es un a cebador biotinilado.



7. Procedimiento de cualquiera de las reivindicaciones 1 a 6, en el que dicha replicación en la etapa a) del procedimiento de la reivindicación 1 o la replicación en las etapas b), f) e i) del procedimiento de la reivindicación 5 es mediante una reacción en cadena de la polimerasa, preferentemente una reacción de extensión de cebador, utilizando dicha replicación preferentemente un cebador biotinilado.
- 5 8. Procedimiento de cualquiera de las reivindicaciones anteriores, en el que:
- dichas etapas de replicación a) y tratamiento b) en el procedimiento de la reivindicación 1 se llevan a cabo una o más veces, repitiéndose dichas etapas preferentemente 5 veces o más, 7 veces o más, 10 veces o más o de aproximadamente 1 a aproximadamente 20 veces o más o
- 10 dichas etapas b), c) y d) en el procedimiento de cualquiera de las reivindicaciones 5 se repiten de aproximadamente 2 a aproximadamente 20 veces, y/o en el que dichas etapas e y h se repiten de aproximadamente 2 a aproximadamente 20 veces.
9. Procedimiento de cualquiera de las reivindicaciones anteriores, en el que, en el procedimiento de la reivindicación 1, las condiciones en la etapa de replicación a), en la que se mantienen dichos restos de citosina 5-metilada y se diluyen los restos de citosina 5-hidroximetilada, comprende la replicación de dicho ácido nucleico con una
- 15 polimerasa para proporcionar ácido nucleico replicado y el tratamiento de dicho ácido nucleico replicado con una enzima para 5-metilar los restos de citosina, siendo dicha enzima preferentemente una ADN metil transferasa, seleccionándose la ADN metil transferasa en el procedimiento preferentemente del grupo que consiste en DNMT1 de ratón, DNMT1 humana y DNMT de M.Sssl.
- 20 10. Procedimiento de cualquiera de las reivindicaciones anteriores, que comprende además la etapa de modificar restos de citosina 5-hidroximetilada en dichas muestras para impedir la metilación de dichos restos de citosina 5-hidroximetilada, modificándose dichos restos de citosina 5-hidroximetilada preferentemente mediante la adición de un grupo bloqueante, seleccionándose dicho grupo bloqueante preferentemente del grupo que consiste en beta-glucosa, alfa-glucosa, 6-O-β-D-glucopiranosil-D-glucosa, 6-O-alfa-D-glucopiranosil-D-glucosa; ceto-glucosa; azida-glucosa; y versiones modificadas de los mismos.
- 25 11. Procedimiento de cualquiera de las reivindicaciones anteriores, en el que dicha muestra de ácido nucleico se selecciona del grupo que consiste en muestras de ácido nucleico de ser humano, planta, ratón, conejo, hámster, primate, pez, ave, vaca, oveja, cerdo, viral, bacteriano y fúngico; y o en el que dicha muestra de ácido nucleico es ADN genómico.
- 30 12. Procedimiento de cualquiera de las reivindicaciones anteriores, que comprende además la comparación de la presencia de 5-hidroximetilcitosina y/o 5-metilcitosina en dicho ácido nucleico en dicha muestra con un nivel de referencia estándar, en la que un aumento o descenso de 5-hidroximetilcitosina y/o 5-metilcitosina en dicho ácido nucleico es indicativo de la presencia de una enfermedad o del probable curso de una enfermedad, y/o que comprende además la etapa de proporcionar un diagnóstico o pronóstico de una enfermedad en base a un aumento o descenso del nivel de 5-hidroximetilcitosina y/o 5-metilcitosina en dicho ácido nucleico en comparación con un
- 35 estándar de referencia, siendo dicha enfermedad preferentemente cáncer.
13. Procedimiento para predecir una predisposición a una enfermedad en un sujeto, diagnosticar una enfermedad en un sujeto, predecir la probabilidad de recurrencia de enfermedad en un sujeto, proporcionar un pronóstico para un sujeto con una enfermedad o seleccionar un sujeto con una enfermedad para tratamiento con una terapia en particular, que comprende:
- 40 a) proporcionar una muestra de ADN genómico de dicho sujeto; y  
b) detectar el estado de metilación e hidroximetilación de porciones predeterminadas de dicha muestra de ADN genómico a través del procedimiento de cualquiera de las reivindicaciones 1 a 13,
- en el que una alteración del nivel de metilación de 5-hidroximetilcitosina y/o 5-metilcitosina de dichas porciones predeterminadas de dicho ADN genómico a un estado de metilación y/o hidroximetilación de referencia proporciona
- 45 una indicación seleccionada del grupo que consiste en una indicación de una predisposición del sujeto a una enfermedad, una indicación de que el sujeto tiene una enfermedad, una indicación de la probabilidad de recurrencia de una enfermedad en el sujeto, una indicación de la supervivencia del sujeto y una indicación de que el sujeto es un candidato para tratamiento con una terapia en particular, siendo dicha enfermedad preferentemente un cáncer, siendo dicho sujeto preferentemente un ser humano.
- 50 14. Un kit para la determinación del estado de metilación e hidroximetilación de una muestra de ácido nucleico que comprende:
- 1) recipiente(s) con reactivos para metilar ácido nucleico; y  
2) recipiente(s) con reactivos para secuenciación con bisulfito;  
3) un medio legible por ordenador que comprende un producto de programa informático que analiza los datos de
- 55 la secuencia obtenida utilizando dicho kit de acuerdo con el procedimiento de las reivindicaciones 1-13; y comprendiendo además dicho kit preferentemente cebadores de ácido nucleico para amplificación y/o secuenciación de una región de dicha muestra de ácido nucleico.

### Ensayo de dilución de 5hmC

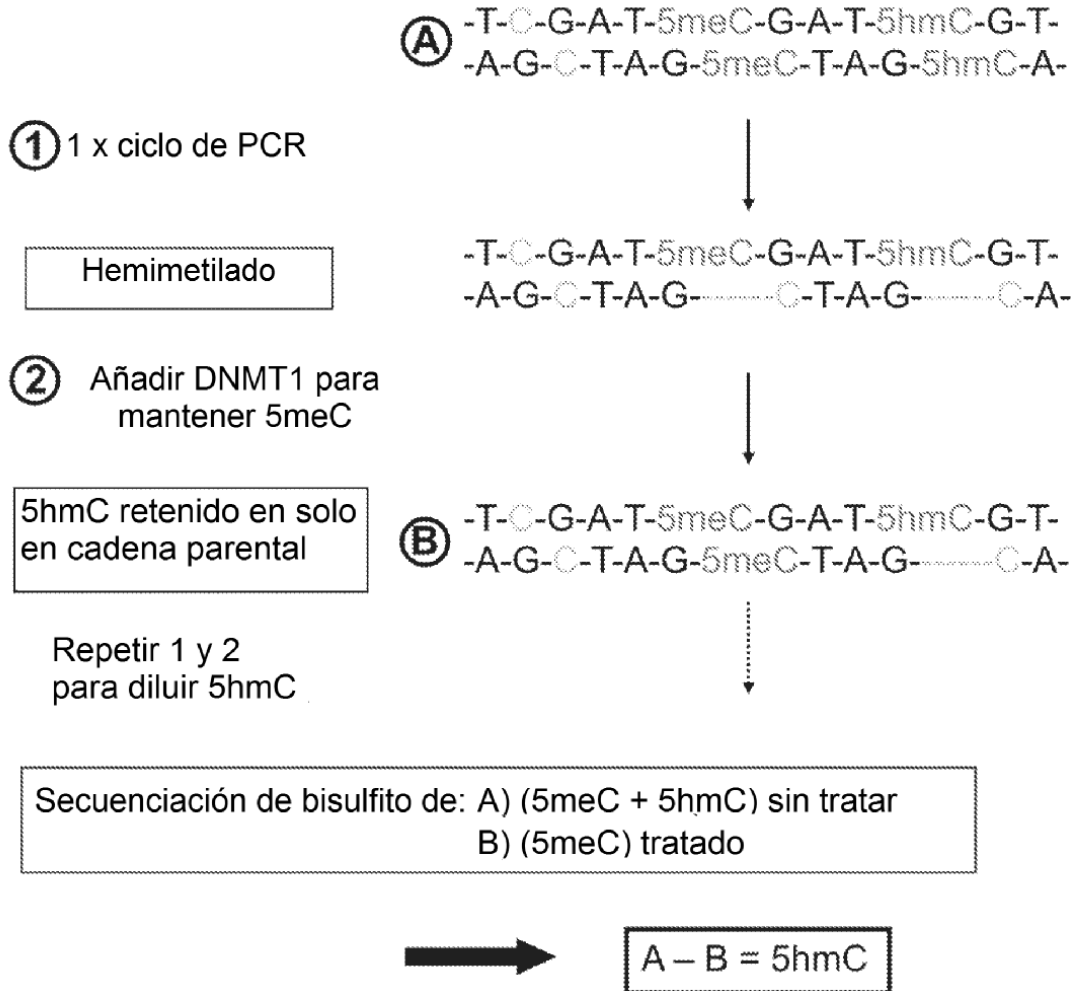


Figura 1

## La conversión de bisulfito y la secuenciación no pueden distinguir entre 5meC y 5hmC

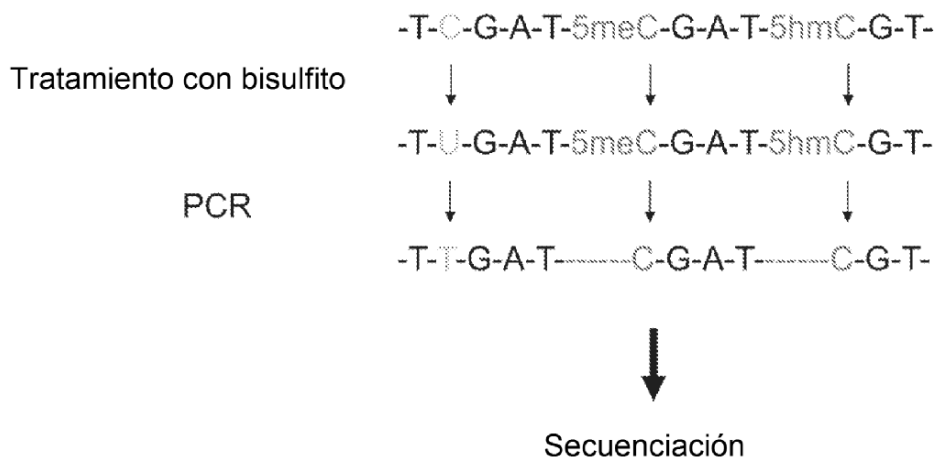


Figura 2

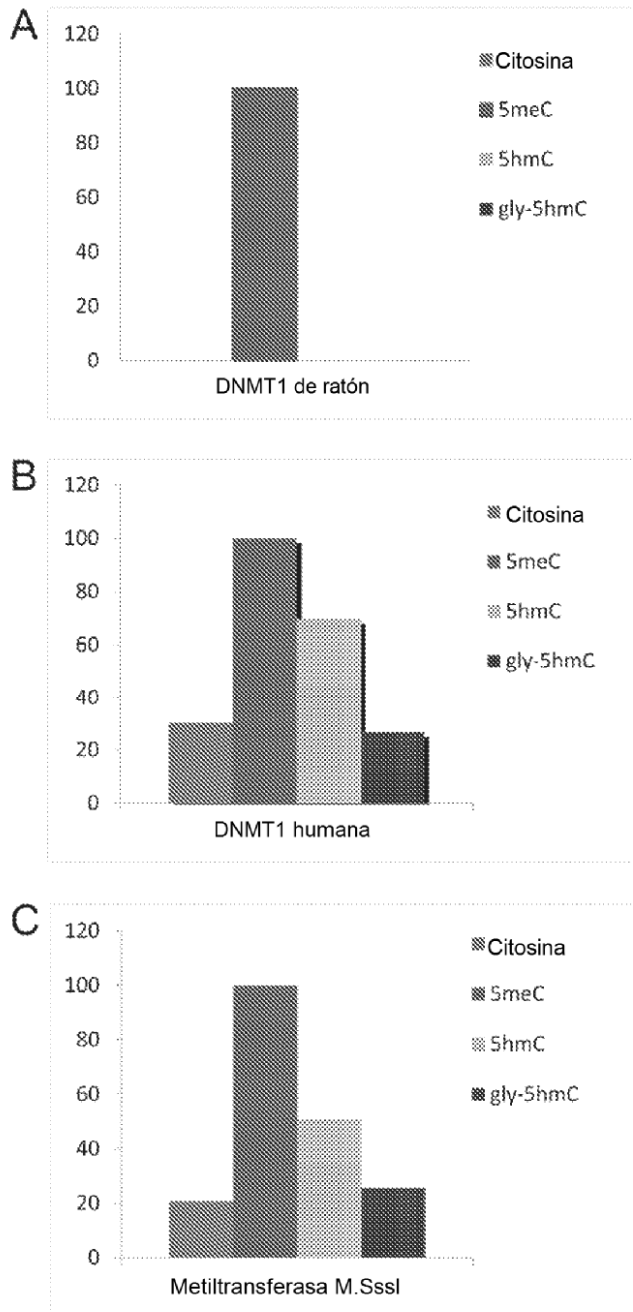


Figura 3

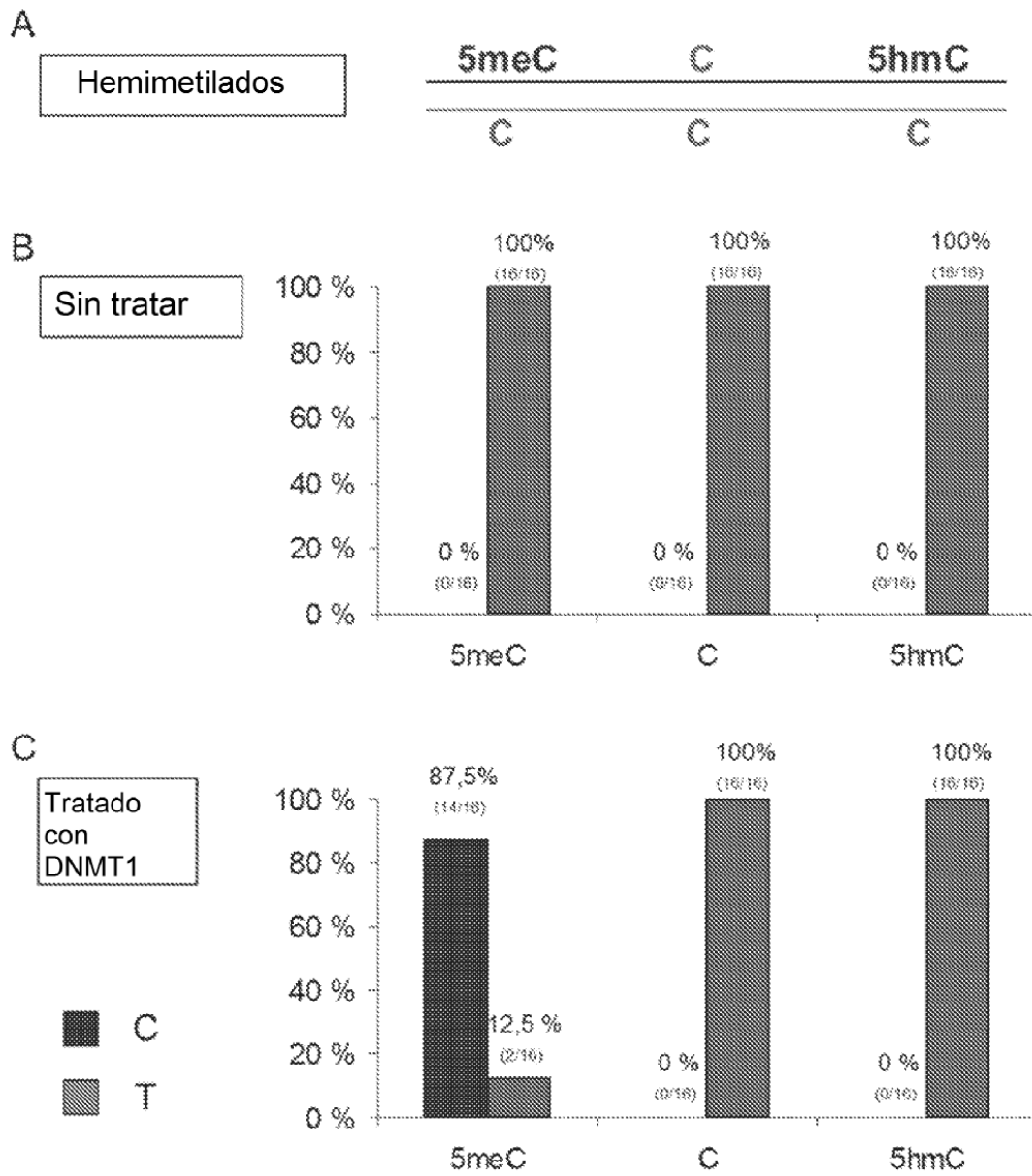


Figura 4

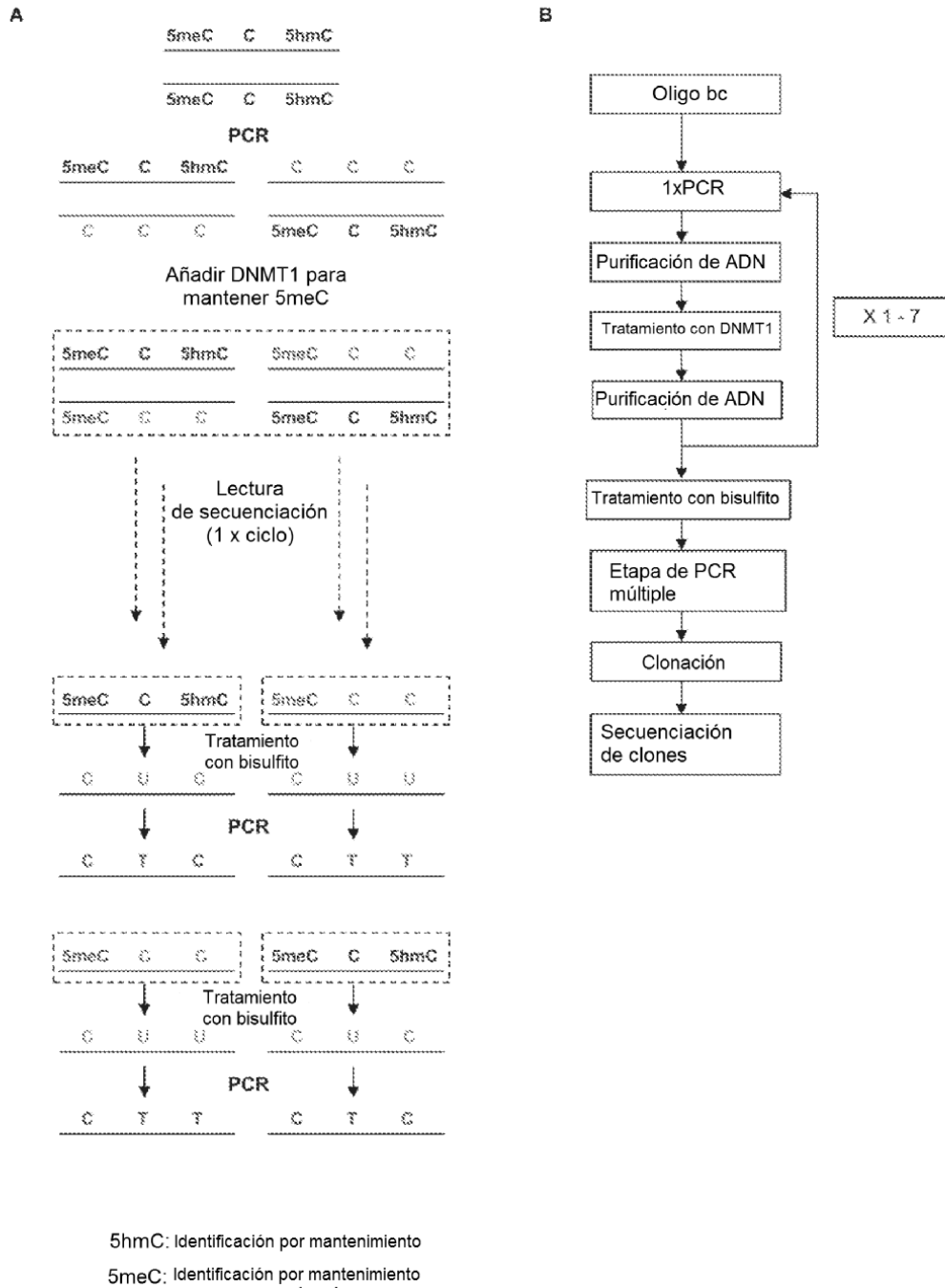


Figura 5

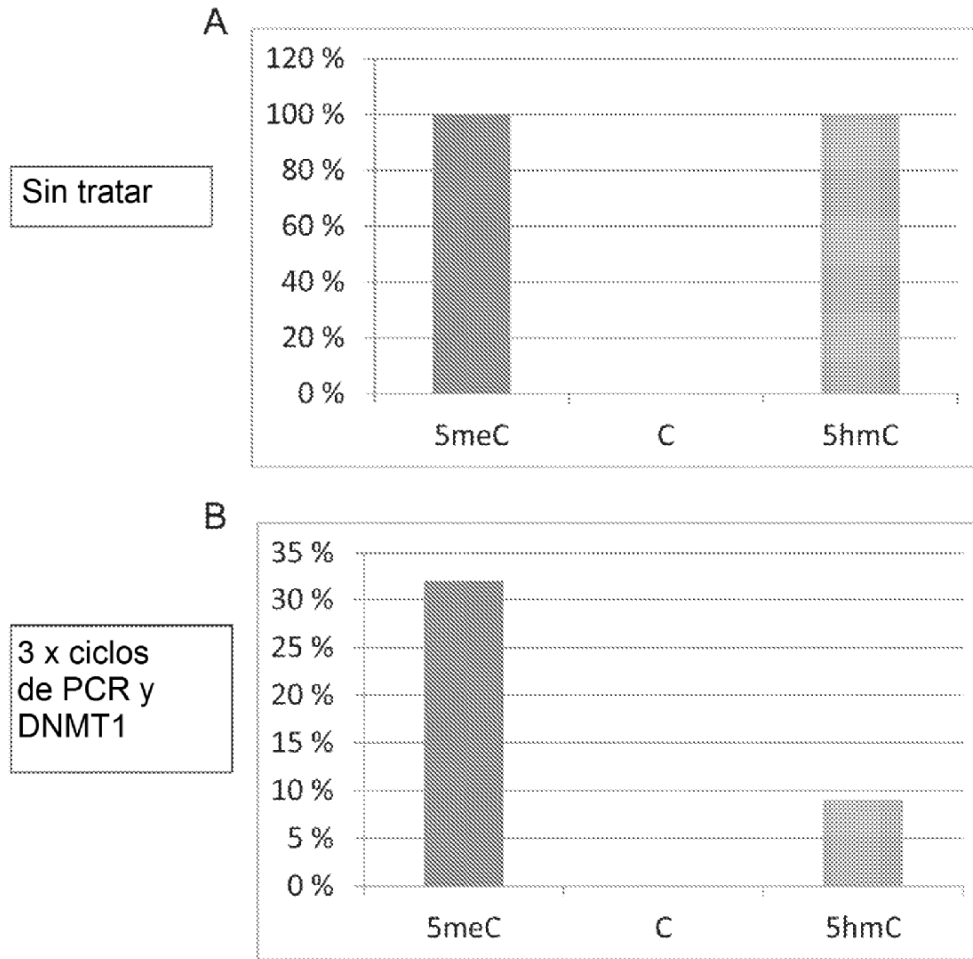
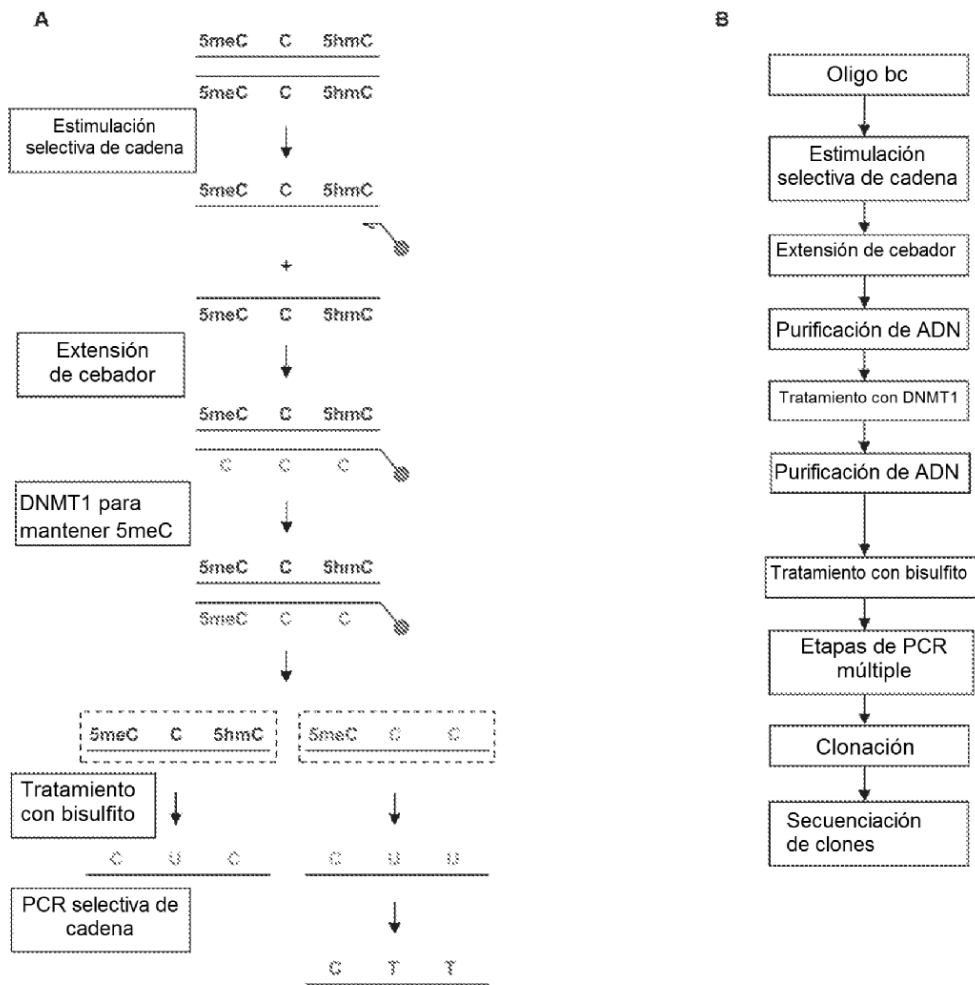


Figura 6



5hmC: No transferido a la nueva cadena

5meC: Transferido a la nueva cadena

Figura 7



**Figura 8**

DNMT1 (*Mus musculus*) Recombinante

Numero de referencia: GenBank: AAH53047.1

(Sinónimos del nombre del gen: Dnmt1; Dnmt; MCMT; Met1; Cxx9; MTase; Met-1; Dnmt1o; m.Mmul; MommeD2)

MPARTAPARVPALASPAGSLPDHVRRRLKDLERDGLTEKECVREKLNLLHEFLQTEIKSQLCDLETKLHK  
 EELSEEGYLAKVKSLLNKDLSENGHTLTQKANGCPANGSRPTWRAEMADSNRSRPSRPRKPRGPRRSKS  
 DSDTLCKDTRHTAVETSFPSSVATRRTRTQTITAHFTKGPTRKPKKEESEEGNSAESAAEBERDQDKKRRV  
 VDTESGAAA VEKLEEV TAGTQLGPEEPCEQEDDNRSRRLRHTRELSLRKSKEDPDREARPETHLDEDED  
 GKDKRSSRPRSQPRDPAKRPRKEAPEQVAPETPEDRDEDEREEKRRKTRKKLESHTVPVQSRSEK  
 AAQSKSVIPKINSPKCEPGQHLDDPNLKYQQHPEDA VDEPQMLTSEKLSIYDSTSTWFDTYEDSPMHRF  
 TSFSVYCSRGLCPVDTLGIEKNVELYFSGCAKAIHDENPSMEGGINGKNLGFINQWWSGFDGGEKVLII  
 GFSTAF AEYILMEPSKEYEPIFGLMQEKIYISKIVVEFLQNNPD AVYEDLINKIETTVPSTINVNRFT  
 DSSLRHAQFVVSQVESYDEAKDDDETPIFLSPCMRALIHLAGVSLGQRRATRVRVMGATKEKDKAPT KATT  
 TKLVYQIFDTFFSEQIEKYDKEDKENAMKRRRCGVCEVCQQPECGKCKACKDMVKFGGTGRSKQA CLKRR  
 CPNLAVKEADDDDEADDDVSEMPSPKHLHQGKKKQNKDRI SWLGQPMKIEENRTYYQKVS IDEEMLEVG  
 DCVSVIPDDSSKPLYLARVTALWEDKNGQMMFHAWFCAGTDTVLGATSDPLELFLVGECEENMQLSYIHS  
 KVKVIYKAPSENWAMEGGTDPETTLPGAEDGKTYFFQLWYNQ EYARFESPPKTQPTEDNKHKFCLSCIRL  
 AELRQKEMPKVLEQIEEVDGRVYCSSITKNGVVYRLGDSVYLPPEAFTFNIKVASPVKRPKDPVNETLY  
 PEHYRKYSDYIKGSNLD APEPYRIGRIKEIHC GKKGKVNEADIKLRLYK FYRPENTHRSYNGSYHTDIN  
 MLYWSDEEAVVNFSDVQGRCTVEYGEDLLESIQDYSQGGPDRFYFLEAYNSKTKNFEDPPNHARSPGNKG  
 KGKKGKGGKGHQVSEPKPEAAIKLPKLR TLDVFSGCGLSEGFHQAGISETLWAIEMWDPAAQAFRLN  
 NPGTTVFTEDCNVLLKLV MAGEVTNSLGQRLPQKGDVEMLCGGPPCQGFSGMNRFNRSRTYSKFKNSLVVS  
 FLSYCDYRPRFFLENNRNFVSYRRSMVLKLT LRCLVRMGYQCTFGVLQAGQYGVAQTRRAI I LAAAP  
 GEKLP LFPPEPLHV FAPRACQLSVVDDKKFVSNITRLSSGPFRTITVRDTMSDLPEIQNGASNSEIPYNG  
 EPLSWFQRQLRGSHYQPILRDHIC KDMSP LVAARMRHIPLFP GSDWRDL PNIQVRLGDGVIAHKLQYTFH  
 DVKNGYSSTGALRGVCSAEGKACDPESRQFSTLI PWCLPHTGNRHHWAGLYGRLEWDGFFSTVTNPE  
 PMGKQGRVLHPEQHRVSVRECAR SQGFPDSYRFFGNILDRHRQVGNVPPPLAKAIGLEIKLCLLSSAR  
 ESASA AVKAKEEAATKD

**Figura 9**

DNMT1 (*Homo sapiens*)

Número de referencia: GenBank: AAI44094.1

MPARTAPARVPTLAVPAISLPDDVRRRLKDLERDSLTEKECVKEKLNLLHEFLQTEIKNQLCDLETKLRK  
EELSEEGYLAKVKSLNKDLSLENGAHAYNREVNGLRLENGNQARSEARRVGMADANSPPKPLSKPRTPRR  
SKSDGEAKRSRDPASASQVTGIRAEPSPSPRI TRKSTRQTTITSHFAKGPAPKRPQEESERAKSDESİK  
EEDKDQDEKRRRVTSRERVAREPLPAEEPERAKSGTRTEKEEEERDEKEEKRLRSQTKEPTPKQKLKEEPDR  
EARAGVQADEDEDGDEKDEKHKRSQPKDLAAKRRPEEKEPEKVN PQISDEKDEDEKEEKRRKTTTPKEPTE  
KKMARAKTVMNSKTHPPKCIQCGQYLDDPDLKYQHPPDAVDEPQMLTNEKLSIFDANESGFESYEALPQ  
HKLTCSFSVYCKHGHLCPIDTGLIEKNIELFFSGSAKPIYDDPSLEGGVNGKNLGPINWWITGFDGGEK  
ALIGFSTSFAYEYILMDPSPEYAFIFGLMQEKIYISKIVVEFLQSNSDSTYEDLINKIETTVPSPGLNLNR  
FTEDSLLRHAQFVVEQVESYDEAGDSDEQPIFLTFCMRDLIKLAGVTLGQRRRAQARRQTI RHSTREKDRG  
PTKATTTKLVYQIFDTFFAEQIEKDDREDKENAFKRRRCGVCEVCQQPECGKCKACKDMVKFGGSGRSKQ  
ACOERRCPNMMKEADDDEEVDNIPEMPSPKMHOGKKKONKNRISWVGEAVKTDGKKSYYKVCIDA  
ETLEVGDVCSVIPDDSSKPLYLARVTALWEDSSNGQMFHAHWFCAGTDTVLGATSDPLELFLVDECEMDQ  
LSYIHSKVKVIYKAPSENWAMEGGMDPESLLEGDDGKTYFYQLWYDQDYARFESPPKTQPTEDNKFKFCV  
SCARLAEMRQKEIPRVLEQLEDLDSRVLYSATKNGILYRVGDGVYLPPEAFTFNIKLSSPVKRRPKEPV  
DEDLYPEHYRKYSDYIKGSNLDAPEPYRIGRIKEIFCPKKSNGRPNETDIKIRVNFYRPENTHKSTPAS  
YHADINLLYWSDEEAVVDFKAVQGRCTVEYGEDLPECVQVYSMGGPNRFYFLEAYNAKSKSFEDFPNHAR  
SPGNKKGKGGKGGKPKSQACEPSEPEIEIKLPKLRTLDVFSGCGGLSEGFBHQAGISDTLWAIEMWDPAA  
QAFRLNPNPGSTVFTEDCNILLKLV MAGETTNSRGQRLPQKGDVEMLCGGPPCQGFSGMNRFNRSRTYSKFK  
NSLVVSFLSYCDYRPRFFLLENVRNFVSFKRSMVLKLTLRCLV RMGYQCTFGVLQAGQYGV AQTRRRAI  
ILAAAPGEKLPFPEPLHVAFPRACQLSVVDDKFKVSNITRLSSGPFRTITVRDTMSDLPEVRNGASAL  
EISYNGEPQSWFQRQLRGAQYQPILRDHICKDMSALVAARMRHIPLAPGSDWRDLPIEVRLSDGTMARK  
LRYTHHDRKNGRSSSGALRGVCSCVEAGKACDPAARQFNTLIPWCLPHTGNRHNHWAGLYGRLEWDGFFS  
TTVTNPEPMGKQGRVLHPEQHRVVSVRECARSQGF PD TYRLFGNILDKHRQVGNVPPPLAKAIGLEIKL  
CMLAKARESASAKIKEEEAAKD

**Figura 10**

M.SS1 (*Spiroplasma* sp. (cepa MQ1))

ADN metiltransferasa específica de sitio (específica de citosina) (EC 2.1.1.73) Sssl- *Spiroplasma* sp. (cepa MQ1)

1 mskvenkttk lrvfeafagi gaqrkalekv rkdeyeivgl aewyvpaivm yqaihnfhf  
61 kleyksvsre emidylenkt lswnsknpsvsgywkrrkdd elkiiynaik lsekegnifd  
121 irdlykrtlk nidlltysfp cqdlsqqgiq kgmkrsgtr sgllweiera ldstekndlp  
181 kyllmenvga llhkkneeel nqwkqklesl gyqnsievl n aadfgssqar rrvfmistln  
241 cfvelpkgdk kpksikkvln kivsekdiln nllkynllef kktksninka sligyskfn  
301 egyvydpeft gptltasgan srikikdgsn irkmnsdtef lyigfdsqdg krveiefl  
361 enqkifvcgn sisvevleai idkigg

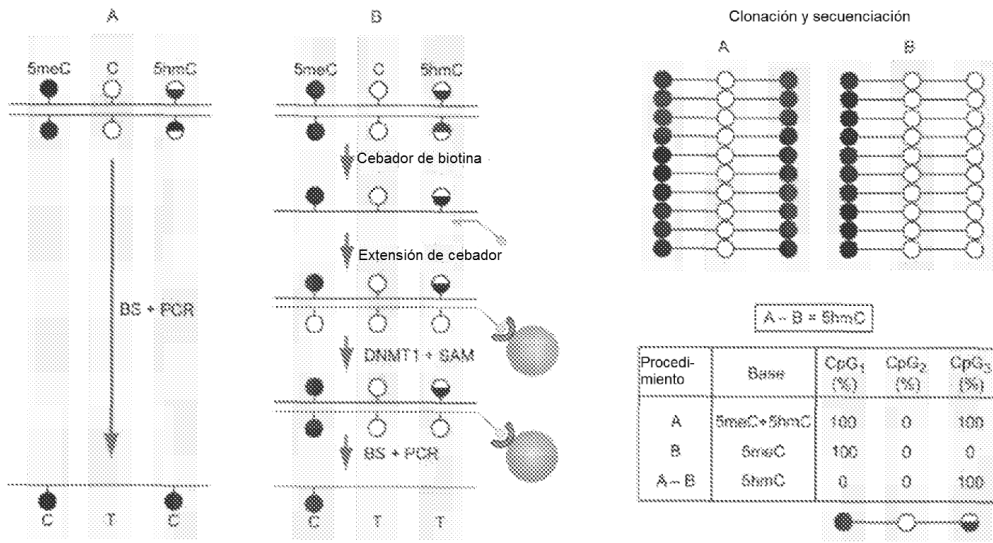


Figura 11

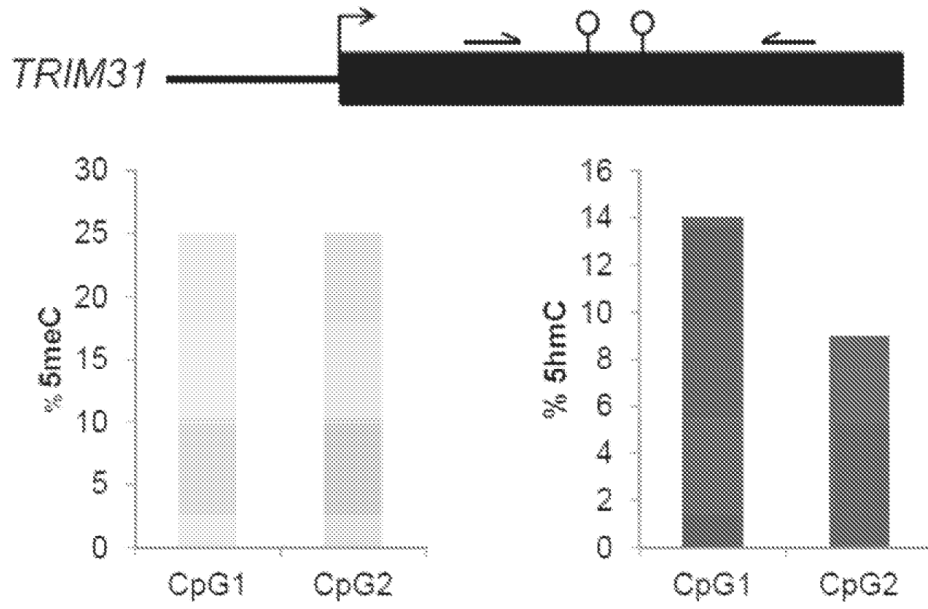


Figura 12

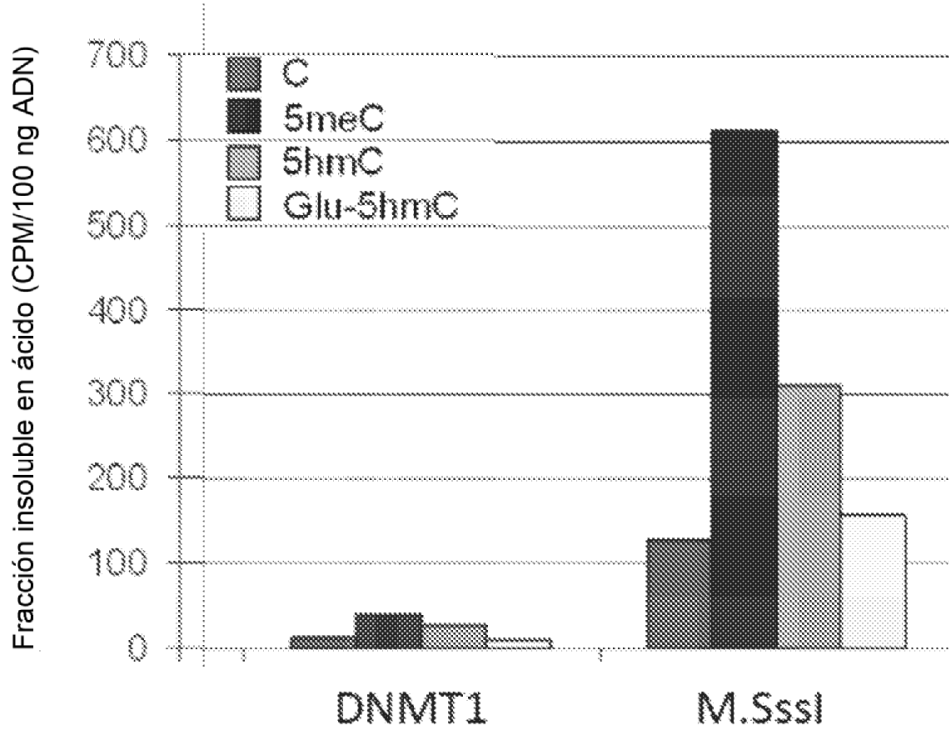


Figura 13