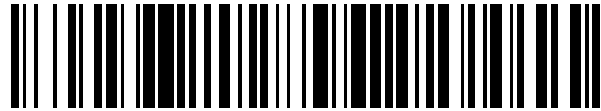


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 609 845**

51 Int. Cl.:

C12Q 1/68

(2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **27.07.2012 PCT/GB2012/051819**

87 Fecha y número de publicación internacional: **07.02.2013 WO13017853**

96 Fecha de presentación y número de la solicitud europea: **27.07.2012 E 12753553 (2)**

97 Fecha y número de publicación de la concesión europea: **12.10.2016 EP 2737085**

54 Título: **Métodos para la detección de modificación de nucleótidos**

30 Prioridad:

29.07.2011 US 201161513356 P
01.03.2012 US 201261605702 P
12.04.2012 US 201261623461 P
01.05.2012 US 201261641134 P

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
24.04.2017

73 Titular/es:

CAMBRIDGE EPIGENETIX LIMITED (100.0%)
1010 Cambourne Business Park
Cambourne, Cambridgeshire CB23 6DP, GB

72 Inventor/es:

BOOTH, MICHAEL, JOHN y
BALASUBRAMANIAN, SHANKAR

74 Agente/Representante:

SÁEZ MAESO, Ana

ES 2 609 845 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

Métodos para la detección de modificación de nucleótidos

5 Esta invención se refiere a la detección de los residuos de citosina modificada y, en particular a la secuenciación de ácidos nucleicos que contienen residuos de citosina modificada.

10 La 5-metilcitosina (5mC) es una marca epigenética bien estudiada que tiene un papel importante en el silenciamiento génico y en la estabilidad del genoma, y se localiza en zonas con predominio de los dinucleótidos CpG (1). En metazoarios, 5mC puede oxidarse a 5-hidroximetilcitosina (5hmC) mediante las enzimas de la familia de enzimas Translocación Diez Once (TET, por sus siglas en inglés) (2,3). Los niveles globales de 5hmC son aproximadamente 10 veces más bajos que aquellos de 5mC y varían entre los tejidos (4). Cantidades relativamente altas de 5hmC (~0.4 % de todas las citosinas) se presentan en las células madres embrionarias (ES), donde 5hmC, se sugiere, tiene un papel en el establecimiento y/o mantenimiento de la pluripotencialidad (2,3, 5-9). Se propone que 5hmC es un intermediario en la desmetilación activa del ADN, por ejemplo mediante la desaminación o mediante una oxidación adicional de 5hmC a 5-formilcitosina (5fC) y 5-carboxicitosina (5cC) por las enzimas TET, seguido de la reparación por escisión de base que involucra a la timina ADN glicosilasa (TDG), o un fallo para mantener la marca durante la replicación (10). Sin embargo, 5hmC puede, además, constituir una marca epigenética *por sí misma*.

20 Es posible detectar y cuantificar el nivel de 5hmC presente en todo el ADN genómico mediante métodos analíticos que incluyen la cromatografía de capa fina y la cromatografía líquida seguida de espectrometría de masas (2, 11, 12). Hasta el momento, el mapeo de las localizaciones genómicas de 5hmC se logró mediante métodos de enriquecimiento que usan química o anticuerpos para la precipitación específica de los fragmentos de ADN ricos en 5hmC que se secuencian posteriormente (6-8, 13-15). Estos procedimientos de inmunoprecipitación tienen una resolución relativamente pobre (10s a 100s de nucleótidos) y solamente proporcionan una cantidad relativa de información que es probable se asocie a algún sesgo de distribución durante el enriquecimiento. La secuenciación de un solo nucleótido cuantificable de 5mC se realiza mediante el uso del método de secuenciación por bisulfito (BS-Seq), que produce la desaminación de la citosina a uracilo, por lo que la transformación correspondiente de 5mC es mucho más lenta (16). Sin embargo, se reconoce que tanto 5mC como 5hmC se desaminan muy lentamente en la reacción del bisulfito y por tanto, esas dos bases no pueden discriminarse (17, 18). Dos métodos de molécula únicas, relativamente nuevos y elegantes, demuestran ser promisorios en la detección de 5mC y 5hmC con una resolución de un solo nucleótido. La secuenciación de una sola molécula en tiempo real (SMRT) demostró detectar derivados de 5hmC en ADN genómico (19). Sin embargo, se requiere el enriquecimiento de fragmentos de ADN que contienen 5hmC, lo cual conduce a la pérdida de información cuantitativa (19). La 5mC puede detectarse, aunque con una exactitud más baja, mediante SMRT (19). Además, SMRT tiene una relativamente alta tasa de errores de secuenciación (20), el máximo de las llamadas modificaciones es impreciso (19) y la plataforma aún no secuenció un genoma completo. Los nanoporos de estado sólido y de proteínas pueden diferenciar 5mC de 5hmC y tienen el potencial para secuenciar moléculas de ADN no amplificadas con un desarrollo ulterior (21, 22).

40 FlusBerg y otros (en Direct detection of ADN methylation during single-molecule real-time sequencing; Nature Methods; 7; 6; 461-465; 1 de julio de 2010) describen un método para la identificación de citosina modificada (es decir 5mC y 5hmC) mediante el uso de la secuenciación de una sola molécula en tiempo real. El método no describe el uso de bisulfito o agentes oxidantes.

45 Nomura y otros (en Discrimination between 5-hydroxymethylcytosine and 5-methylcytosine by a chemically designed peptide; Chemical Communications; 47; 29; 8277-8279; 7 de agosto de 2011 describen que 5hmC se discrimina selectivamente de 5mC mediante el uso de un péptido dedo de zinc. El método no describe el uso de bisulfito o agentes oxidantes.

50 Gupta y otros (en Advances in genome-wide ADN methylation analysis; Biotechniques; 49; 4; iii-xi; Oct 2010 revisan los métodos para el análisis de metilación del ADN en el genoma completo, pero no describen el uso de bisulfito o agentes oxidantes.

55 Wanunu y otros (en Discrimination of methylcytosine from hydroxymethylcytosine in ADN molecules; Journal of the American Chemical Society; 133; 3; 486-492 26 de enero de 2011) describen un método para discriminar entre hmC y mC mediante el uso de nanoporos de estado sólido, que se basa en las diferencias físicas entre las dos bases. El método no describe el uso de bisulfito o agentes oxidantes.

60 Los presentes inventores idearon métodos que permiten que los residuos de citosina modificada, tales como 5-metilcitosina (5mC), 5-hidroximetilcitosina (5hmC) y 5-formilcitosina (5fC) se distingan de la citosina (C) con la resolución de un solo nucleótido. Esos métodos se aplican a todas las plataformas de secuenciación y pueden utilizarse, por ejemplo en el análisis de ADN genómico y/o ARN.

65 Un aspecto de la invención proporciona un método de identificación de un residuo de citosina modificada en una secuencia de nucleótidos de la muestra, que comprende;

i) proporcionar una población de polinucleótidos los cuales comprenden la secuencia de nucleótidos de la muestra,

- ii) oxidar o reducir una primera porción de dicha población,
- iii) tratar la primera porción oxidada o reducida de dicha población y una segunda porción de dicha población con bisulfito,
- iv) secuenciar los polinucleótidos en la primera y en la segunda porciones de la población que siguieron las etapas ii) y iii) para producir una primera y segunda secuencia de nucleótidos, respectivamente, e;
- v) identificar los residuos en la primera y segunda secuencias de nucleótidos que correspondan a un residuo de citosina en la secuencia de nucleótidos de la muestra, en donde la citosina modificada es 5-metilcitosina (5mC), 5-hidroximetilcitosina (5hmC) o 5-formilcitosina (5fC).

Los residuos que se identifican en la primera y segunda secuencias de nucleótidos que se corresponden a un residuo de citosina en la secuencia de nucleótidos de la muestra indican la modificación del residuo de citosina.

Por ejemplo, los residuos de citosina pueden presentarse en una o más posiciones en la secuencia de ácidos nucleicos de la muestra. Los residuos en esa única o más posiciones en la primera y segunda secuencias de nucleótidos pueden identificarse. Una citosina modificada en una posición en la secuencia de nucleótidos de la muestra puede identificarse de la combinación de residuos que se identifican en la primera y segunda secuencias de nucleótidos respectivamente (es decir C y C, U y U, C y U, o U y C) en esa posición. Las modificaciones de citosina que se indican por combinaciones diferentes se muestran en la tabla 1.

Un residuo de citosina modificada puede contener una modificación en la posición 5. Las citosinas modificadas adecuadas incluyen las citosinas sustituidas en 5.

Los grupos que pueden sustituirse en la posición 5 de la citosina incluyen los grupos metilo (m), hidroximetilo (hm) o formilo (f).

Los métodos descritos en la presente descripción pueden utilizarse en identificar y/o distinguir citosina (C), 5-metilcitosina (5mC), 5-hidroximetilcitosina (5hmC) y 5-formilcitosina (5fC) en una secuencia de nucleótidos de la muestra. Por ejemplo, los métodos descritos en la presente descripción pueden utilizarse en distinguir un residuo del grupo que consiste de citosina (C), 5-metilcitosina (5mC), 5-hidroximetilcitosina (5hmC) y 5-formilcitosina (5fC) de otros residuos en el grupo.

Preferentemente, los residuos de citosina modificada, tal como 5-hidroximetilcitosina, en la primera porción de dicha población no se marcan, por ejemplo con grupos sustituyentes, tal como la glucosa, antes de la oxidación o reducción de la etapa ii).

En algunas modalidades de la invención, la primera porción de los polinucleótidos de la población puede oxidarse. Por ejemplo, los residuos de 5-hidroximetilcitosina en la primera porción del polinucleótido pueden convertirse en 5-formilcitosina (5fC) mediante oxidación y después la primera porción del polinucleótido se trata con bisulfito.

Un método de identificación de un residuo de citosina modificada en una secuencia de nucleótido de la muestra puede comprender;

- i) proporcionar una población de polinucleótidos los cuales comprenden una secuencia de nucleótidos de la muestra,
- ii) oxidar una primera porción de dicha población,
- iii) tratar la primera porción oxidada de dicha población y una segunda porción de dicha población con bisulfito.
- iv) secuenciar los polinucleótidos en la primera y segunda porciones de la población que siguieron las etapas ii) y iii) para producir una primera y segunda secuencias de nucleótidos, respectivamente e;
- v) identificar los residuos en la primera y segunda secuencias de nucleótidos que se correspondan a un residuo de citosina en la secuencia de nucleótidos de la muestra, en donde la citosina modificada es 5-metilcitosina (5mC), 5 hidroximetilcitosina (5hmC) o 5-formilcitosina (5fC).

La identificación de un residuo, en una posición en una o en ambas de la primera y segunda secuencias de nucleótidos, como la citosina, en una o en ambas de la primera y la segunda secuencias de nucleótidos indica que el residuo de citosina en la secuencia de nucleótidos de la muestra es 5-metilcitosina o 5-hidroximetilcitosina.

La 5-hidroximetilcitosina (5hmC) puede identificarse en la secuencia de nucleótidos de la muestra. Un residuo de uracilo en una posición en la primera secuencia de nucleótidos que corresponde a una citosina en la secuencia de nucleótidos de la muestra y una citosina en la misma posición en la segunda secuencia de nucleótidos indican que el residuo de citosina en la secuencia de nucleótidos de la muestra es 5-hidroximetilcitosina (5hmC).

Por ejemplo, un método de identificación de un residuo de 5-hidroximetilcitosina (5hmC) en una secuencia de nucleótidos de la muestra o para distinguir 5-hidroximetilcitosina de citosina (C), 5-metilcitosina, y 5-formilcitosina (5fC) en una secuencia de nucleótidos de la muestra puede comprender;

- i) proporcionar una población de polinucleótidos los cuales comprenden la secuencia de nucleótidos de la muestra,
- ii) oxidar una primera porción de dicha población,
- iii) tratar la primera porción oxidada de dicha población y una segunda porción de dicha población con bisulfito,

- iv) secuenciar los polinucleótidos en la primera y segunda porciones de la población que siguieron las etapas ii) y iii) para producir una primera y segunda secuencias de nucleótidos, respectivamente e;
- v) identificar los residuos en la primera y segunda secuencias de nucleótidos que correspondan a un residuo de citosina en la secuencia de nucleótidos de la muestra,
 5 en donde la presencia de un residuo de uracilo en la primera secuencia de nucleótidos y una citosina en la segunda secuencia de nucleótidos indican que el residuo de citosina en la secuencia de nucleótidos de la muestra es 5-hidroximetilcitosina.

10 La 5-metilcitosina (5mC) puede identificarse en una secuencia de nucleótidos de la muestra. La citosina en una posición tanto en la primera como en la segunda secuencia de nucleótidos que corresponda a un residuo de citosina en la secuencia de nucleótidos de la muestra indica que el residuo de citosina en la secuencia de nucleótidos de la muestra es 5-metilcitosina (5mC).

15 Por ejemplo, un método de identificación de 5-metilcitosina en una secuencia de nucleótidos de la muestra o para distinguir 5-metilcitosina de citosina (C), 5-hidroximetilcitosina (5hmC), y 5-formilcitosina (5fC) en una secuencia de nucleótidos de la muestra puede comprender;

- i) proporcionar una población de polinucleótidos los cuales comprenden la secuencia de nucleótidos de la muestra,
- ii) oxidar una primera porción de dicha población,
- iii) tratar la primera porción oxidada de dicha población y una segunda porción de dicha población con bisulfito,
- 20 iv) secuenciar los polinucleótidos en la primera y segunda porciones de la población que siguieron las etapas ii) y iii) para producir una primera y segunda secuencias de nucleótidos, respectivamente e;
- v) identificar los residuos en la primera y segunda secuencias de nucleótidos que correspondan a un residuo de citosina en la secuencia de nucleótidos de la muestra,
 25 en donde la presencia de una citosina tanto en la primera como en la segunda secuencia de nucleótidos indica que el residuo de citosina en la secuencia de nucleótidos de la muestra es 5-metilcitosina (5mC).

30 Los residuos de uracilo en una posición tanto en la primera como en la segunda secuencias de nucleótidos que se correspondan a una citosina en la secuencia de nucleótidos de la muestra indican que el residuo de citosina en la secuencia de nucleótidos de la muestra no es 5-metilcitosina o 5-hidroximetilcitosina, es decir el residuo de citosina es una citosina no modificada o 5-formilcitosina.

35 Un resumen de las modificaciones de la citosina en una posición en la secuencia de nucleótidos de la muestra que se indican mediante combinaciones específicas de citosina y uracilo en la posición de la primera y segunda secuencia de nucleótidos se muestra en la Tabla 1.

La primera y segunda porciones de la población de nucleótidos pueden tratarse con bisulfito y/o secuenciarse simultáneamente o secuencialmente.

40 En algunas modalidades en las que la primera porción se oxida en la etapa ii), puede no requerirse el tratamiento de la segunda porción para identificar o distinguir un residuo de citosina modificada en la secuencia de nucleótidos de la muestra. Por ejemplo, la Tabla 1 muestra que la oxidación y el tratamiento con bisulfito de la primera porción de la población de polinucleótidos es suficiente para identificar 5-metilcitosina en la secuencia de nucleótidos de la muestra. Un método de identificación de 5-metilcitosina en una secuencia de nucleótidos de la muestra o para distinguir 5-metilcitosina de citosina (C), 5-hidroximetilcitosina (5hmC) y 5-formilcitosina (5fC) en una secuencia de nucleótidos de la muestra puede comprender;

- i) proporcionar una población de polinucleótidos los cuales comprenden la secuencia de nucleótidos de la muestra,
- ii) oxidar dicha población,
- iii) tratar la población oxidada con bisulfito
- 45 iv) secuenciar los polinucleótidos en la población que siguieron las etapas ii) y iii) para producir una secuencias de nucleótidos tratada, e;
- 50 v) identificar los residuos en la secuencia de nucleótidos tratada que se correspondan a un residuo de citosina en la secuencia de nucleótidos de la muestra,
 55 en donde la presencia de una citosina en la secuencia de nucleótidos tratada indica que el residuo de citosina en la secuencia de nucleótidos de la muestra es 5-metilcitosina (5mC).

En algunas modalidades de la invención, la primera porción de los polinucleótidos de la población puede reducirse en la etapa ii). Un método de identificación de un residuo de citosina modificada en una secuencia de nucleótidos de la muestra puede comprender;

- i) proporcionar una población de polinucleótidos los cuales comprenden la secuencia de nucleótidos de la muestra,
- 60 ii) reducir una primera porción de dicha población,
- iii) tratar la primera porción reducida de dicha población y una segunda porción de dicha población con bisulfito,
- iv) secuenciar los polinucleótidos en la primera y segunda porciones de la población que siguieron las etapas ii) y iii) para producir una primera y segunda secuencias de nucleótidos, respectivamente e;
- 65 v) identificar los residuos en la primera y segunda secuencias de nucleótidos que se correspondan con un residuo de citosina en la secuencia de nucleótidos de la muestra,
 en donde la citosina modificada es 5-metilcitosina (5mC), hidroximetilcitosina (5hmC) o 5-formilcitosina (5fC).

Esto puede usarse para identificar y/o distinguir 5-formilcitosina (5fC) en una secuencia de nucleótidos de la muestra.

5 En las modalidades en las que la primera porción se reduce, la citosina en una posición en la primera secuencia de nucleótidos que se corresponde con una citosina en la secuencia de nucleótidos de la muestra y un residuo de uracilo en esta posición en la segunda secuencia de nucleótidos indican que el residuo de citosina en la secuencia de nucleótidos de la muestra es 5-formilcitosina; el uracilo en una posición en la primera y segunda secuencia de nucleótidos que se corresponda con un residuo de citosina en la secuencia de nucleótidos de la muestra indica que el residuo de citosina en la secuencia de nucleótidos de la muestra es una citosina no modificada; y una citosina en una posición tanto en la primera como en la segunda secuencias de nucleótidos que se corresponda con un residuo de citosina en la secuencia de nucleótidos de la muestra indica que el residuo de citosina en la secuencia de nucleótidos de la muestra es 5-metilcitosina (5mC) o 5-hidroximetilcitosina (5hmC).

15 Un resumen de las modificaciones de la citosina en una posición en la secuencia de nucleótidos de la muestra que se indican mediante combinaciones específicas de citosina y uracilo en la posición de la primera y segunda secuencia de nucleótidos se muestra en la Tabla 1.

20 En algunas modalidades, los métodos de la invención pueden comprender la secuenciación de una primera porción de polinucleótidos los cuales se oxidaron y se trataron con bisulfito; una segunda porción de polinucleótidos los cuales se trataron con bisulfito; y una tercera porción de polinucleótidos de la población los cuales se redujeron y se trataron con bisulfito. Por ejemplo, un método puede comprender;

- 25 i) proporcionar una población de polinucleótidos los cuales comprenden la secuencia de nucleótidos de la muestra,
- ii) proporcionar la primera, la segunda y la tercera porciones de la población,
- iii) oxidar la primera porción de dicha población,
- iv) reducir la tercera porción de dicha población,
- v) tratar la primera, la segunda y la tercera porciones de dicha población con bisulfito,
- vi) secuenciar los polinucleótidos en la primera, la segunda y la tercera porciones de la población que siguieron las etapas iii), iv) y v) para producir una primera, segunda y tercera secuencias de nucleótidos, respectivamente e:
- 30 vii) identificar los residuos en la primera, segunda y tercera secuencias de nucleótidos que se correspondan a un residuo de citosina en la secuencia de nucleótidos de la muestra, en donde la citosina modificada es 5-metilcitosina (5mC), 5-hidroximetilcitosina (5hmC) o 5-formilcitosina (5fC).

35 Esto puede usarse, por ejemplo, para identificar y/o distinguir 5-formilcitosina (5fC) en una secuencia de nucleótidos de la muestra de la citosina y/u otras citosinas modificadas.

El uracilo en una posición en la primera secuencia de nucleótidos que corresponde a una citosina en la secuencia de nucleótidos de la muestra y una citosina en esta posición en la tercera secuencia de nucleótidos indican que el residuo de citosina en la secuencia de nucleótidos de la muestra es 5-formilcitosina.

40 La citosina en una posición en la primera, segunda y tercera secuencia de nucleótidos que se corresponda a un residuo de citosina en la secuencia de nucleótidos de la muestra indica que el residuo de citosina en la secuencia de nucleótidos de la muestra es 5-metilcitosina (5mC).

45 La citosina en una posición en la segunda y tercera secuencia de nucleótidos que se corresponda a una citosina en la secuencia de nucleótidos de la muestra y un residuo de uracilo en esta posición en la primera secuencia de nucleótidos indican que el residuo de citosina en la secuencia de nucleótidos de la muestra es 5-hidroximetilcitosina.

50 El uracilo en la primera, segunda y tercera secuencia de nucleótidos que se corresponda a un residuo de citosina en la secuencia de nucleótidos de la muestra indica que el residuo de citosina en la secuencia de nucleótidos de la muestra es una citosina no modificada.

55 Un resumen de las modificaciones de la citosina en una posición en la secuencia de nucleótidos de la muestra que se indican mediante combinaciones específicas de citosina y uracilo en la posición de la primera, segunda y tercera secuencia de nucleótidos se muestra en la Tabla 1.

60 La secuencia de nucleótidos de la muestra puede conocerse previamente o esta pudiera determinarse. La secuencia de nucleótidos de la muestra es la secuencia de polinucleótidos no tratados en la población es decir, polinucleótidos los cuales no se oxidaron, redujeron o se trataron con bisulfito. En la secuencia de nucleótidos de la muestra, las citosinas modificadas no se distinguen de la citosina. La 5-metilcitosina, 5-formilcitosina y 5-hidroximetilcitosina se indican todas o se identifican como residuos de citosina en la secuencia de nucleótidos de la muestra. Por ejemplo, cualquiera de los métodos descritos en la presente descripción pueden comprender adicionalmente;

- proporcionar una cuarta porción de la población de polinucleótidos que comprende la secuencia de nucleótidos de la muestra; y
- secuenciar los polinucleótidos en la cuarta porción para producir la secuencia de nucleótidos de la muestra.

ES 2 609 845 T3

La secuencia de los polinucleótidos en la cuarta porción puede determinarse por cualquiera de las técnicas de secuenciación apropiadas.

5 Las posiciones de uno o más residuos de citosina en la secuencia de nucleótidos de la muestra pueden determinarse. Esto puede hacerse mediante análisis estándar de secuencias. Debido a que los residuos de citosina modificada no se distinguen de la citosina, los residuos de citosina en la secuencia de nucleótidos de la muestra pueden ser citosina, 5-metilcitosina, 5-formilcitosina, o 5-hidroximetilcitosina.

10 La primera y la segunda secuencias de nucleótidos y, opcionalmente la tercera secuencia de nucleótidos, pueden compararse con la secuencia de nucleótidos de la muestra. Por ejemplo, los residuos en posiciones en la primera y segunda secuencias y, opcionalmente la tercera secuencia de nucleótidos, que se correspondan a uno o más residuos de citosina en la secuencia de nucleótidos de la muestra pueden identificarse.

15 La modificación de un residuo de citosina en la secuencia de nucleótidos de la muestra puede determinarse a partir de la identidad de los nucleótidos en las posiciones correspondientes en la primera y segunda secuencias de nucleótidos y, opcionalmente la tercera secuencia de nucleótidos.

20 Todos los polinucleótidos en la población contienen la misma secuencia de nucleótidos de la muestra, es decir la secuencia de nucleótidos de la muestra es idéntica en todos los polinucleótidos en la población.

El efecto de los diferentes tratamientos en los residuos de citosina dentro de la secuencia de nucleótidos de la muestra puede entonces determinarse, como se describe en la presente descripción.

25 La secuencia de nucleótidos de la muestra puede ser una secuencia genómica. Por ejemplo, la secuencia puede comprender toda o una parte de la secuencia de un gen, que incluye exones, intrones, o elementos regulatorios cadena arriba o abajo, o la secuencia puede comprender secuencias genómicas que no se asocien con un gen. En algunas modalidades, la secuencia de nucleótidos de la muestra puede comprender una o más islas de CpG.

30 Los polinucleótidos adecuados incluyen ADN, preferentemente ADN genómico, y/o ARN, tal como ARN genómico (por ejemplo ARN genómico de mamíferos, plantas o virus), ARNm, ARNt, ARNr y ARN no codificante.

Los polinucleótidos que comprende la secuencia de nucleótidos de la muestra pueden obtenerse o aislarse de una muestra de células, por ejemplo, células de mamíferos, preferentemente, células humanas.

35 Las muestras adecuadas incluyen células aisladas y muestras de tejidos tal como biopsias.

Los residuos de citosina modificada que incluyen 5hmC y 5fC se detectan en un rango de tipos celulares que incluyen las células madres embrionarias (ESCS) y las células neurales (2, 3, 11, 37, 38).

40 Las células adecuadas incluyen células somáticas y de la línea germinal.

Las células adecuadas pueden encontrarse en cualquier estado del desarrollo, que incluye células total o parcialmente diferenciadas o no diferenciadas o células pluripotentes, que incluye las células madres, tales como células madres somáticas o adultas, células madres fetales, o células madres embrionarias.

45 Las células adecuadas incluyen, además, células madres pluripotentes inducidas (iPSCs), las que pueden derivarse de cualquier tipo de células somáticas de acuerdo con técnicas estándar.

50 Por ejemplo, los polinucleótidos que comprende la secuencia de nucleótidos de la muestra pueden obtenerse o aislarse de células neurales, que incluyen neuronas y células de la glía, células de músculo esquelético, células de músculo liso, células hepáticas, células que sintetizan hormonas, células sebáceas, células de los islotes pancreáticos, células de la corteza adrenal, fibroblastos, queratinocitos, células endoteliales y uroteliales, osteocitos y condrocitos.

55 Las células adecuadas incluyen células asociadas a enfermedades, por ejemplo células cancerosas, tales como carcinomas, sarcomas, linfomas, blastomas o células tumorales de la línea germinal.

Las células adecuadas incluyen células con el genotipo de un trastorno genético como la enfermedad de Huntington, la fibrosis quística, la anemia drepanocítica, la fenilcetonuria, el Síndrome de Down o el Síndrome de Marfan.

60 Los métodos para la extracción y el aislamiento del ADN genómico y el ARN de las muestras de células se conocen bien en la técnica. Por ejemplo, el ADN genómico o ARN pueden aislarse mediante el uso de cualquier técnica de aislamiento conveniente, tales como la extracción por fenol/cloroformo y precipitación alcohólica, la centrifugación por gradiente de densidad de cloruro de cesio, la cromatografía de intercambio aniónico en fase sólida, y las técnicas basadas en gel de sílice.

65

En algunas modalidades, el ADN genómico completo y/o ARN que se aíslan de las células pueden usarse directamente como una población de polinucleótidos como se describe en la presente descripción, después del aislamiento. En otras modalidades, el ADN genómico y/o el ARN que se aíslan pueden someterse a etapas de preparación adicionales.

El ADN genómico y/o ARN pueden fragmentarse, por ejemplo mediante sonicación, cizallamiento o digestión por enzimas endonucleasas, para producir fragmentos de ADN. Una fracción del ADN genómico y/o ARN pueden usarse como se describe en la presente descripción. Las fracciones adecuadas de ADN genómico y/o ARN pueden basarse en la talla u otros criterios. En algunas modalidades, una fracción de ADN genómico y/o fragmentos de ARN los cuales son ricos en islas de CpG (CGI) pueden usarse como se describe en la presente descripción.

El ADN genómico y/o ARN pueden desnaturalizarse, por ejemplo mediante calentamiento o por tratamiento con un agente desnaturalizante. Los métodos adecuados para la desnaturalización del ADN genómico y el ARN se conocen bien en la técnica.

En algunas modalidades, el ADN genómico y/o ARN pueden adaptarse para la secuenciación antes de la oxidación o reducción y el tratamiento con bisulfito, o el tratamiento con bisulfito solo. La naturaleza de las adaptaciones depende del método de secuenciación que se va a usar. Por ejemplo, para algunos métodos de secuenciación, los cebadores pueden ligarse a los extremos libres del ADN genómico y/o fragmentos de ARN después de la fragmentación. Los cebadores adecuados pueden contener 5mC para evitar que las secuencias del cebador se modifiquen durante la oxidación o reducción y el tratamiento con bisulfito, o el tratamiento con bisulfito solo, como se describe en la presente descripción. En otras modalidades, el ADN genómico y/o ARN pueden adaptarse para la secuenciación después de la oxidación, reducción y/o tratamiento con bisulfito, como se describe en la presente descripción.

Después del fraccionamiento, desnaturalización, adaptación y/o otras etapas de la preparación, el ADN genómico y/o ARN pueden purificarse por cualquier técnica conveniente.

Después de la preparación, la población de polinucleótidos puede proporcionarse en una forma adecuada para tratamientos posteriores como se describe en la presente descripción. Por ejemplo, la población de polinucleótidos puede estar en una solución acuosa en ausencia de tampones antes del tratamiento como se describe en la presente descripción.

Los polinucleótidos para usar como se describe en la presente descripción pueden ser de simple o doble cadena.

La población de polinucleótidos puede dividirse en dos, tres, cuatro o más porciones separadas, cada una de las cuales contiene los polinucleótidos que comprende la secuencia de nucleótidos de la muestra. Estas porciones pueden tratarse y secuenciarse independientemente como se describe en la presente descripción.

Preferentemente, las porciones de polinucleótidos no se tratan para adicionar marcas o grupos sustituyentes, tal como glucosa, a los residuos de 5-hidroximetilcitosina en la secuencia de nucleótidos de la muestra, antes de la oxidación y/o reducción.

La primera porción de la población de polinucleótidos que comprende la secuencia de nucleótidos de la muestra puede oxidarse. La oxidación convierte cualquier 5-hidroximetilcitosina en la secuencia de nucleótidos de la muestra en 5-formilcitosina. La oxidación puede ser oxidación no mediada por enzimas, por ejemplo mediante el uso de un agente químico oxidante orgánico o inorgánico, preferentemente bajo condiciones desnaturalizantes.

La primera porción puede oxidarse mediante el tratamiento con un agente oxidante. El agente oxidante es cualquier agente adecuado para generar un aldehído a partir de un alcohol. El agente oxidante o las condiciones que se usan en la etapa de oxidación pueden seleccionarse de modo que cualquier 5-hidroximetilcitosina se oxide selectivamente. De esta manera, sustancialmente ninguna otra función en el polinucleótido se oxida en la etapa de oxidación. La etapa de oxidación, por consiguiente, no resulta en la reacción de cualquier residuo de timina o de 5-metilcitosina, donde aquellas estén presentes. Los agentes o condiciones se seleccionan para minimizar o evitar cualquier degradación del polinucleótido.

El uso de un agente oxidante puede resultar en la formación de algún producto correspondiente a 5-carboxicitosina. La formación de este producto no influye negativamente en los métodos de identificación descritos en la presente descripción. Bajo las condiciones de la reacción de bisulfito que se usan para convertir 5-formilcitosina en uracilo, se percibe además, que 5-carboxicitosina se convierte igualmente en uracilo. Se comprende que una referencia a 5-formilcitosina que se obtiene por oxidación de 5-hidroximetilcitosina puede ser una referencia a un producto que, además, comprende 5-carboxicitosina que se obtiene, además, mediante esa oxidación.

El agente oxidante puede ser un agente oxidante no enzimático, por ejemplo, un compuesto químico orgánico o inorgánico.

Los agentes oxidantes adecuados se conocen bien en la técnica e incluyen óxidos de metal, tales como KRuO_4 , MnO_2 y

KMnO₄. Particularmente los agentes oxidantes útiles son aquellos que pueden usarse en condiciones acuosas, como tales son los más convenientes para la manipulación de los polinucleótidos. Sin embargo, los agentes oxidantes que son adecuados para usar en solventes orgánicos pueden usarse además donde sea factible.

5 En algunas modalidades, el agente oxidante puede comprender un anión perrutenato (RuO₄⁻). Los agentes oxidantes perrutenatos adecuados incluyen sales de perrutenato orgánicas e inorgánicas, tales como el perrutenato de potasio (KRuO₄) y otros perrutenatos metálicos, perrutenato de tetraalquilamonio, tales como el perrutenato de tetrapropilamonio (TPAP) y el perrutenato de tetrabutilamonio (TBAP); perrutenato en soportes de polímero (PSP) y el rutenato de tetrafenilfosfonio.

10 Favorablemente, el agente oxidante o las condiciones oxidantes pueden, además, preservar el polinucleótido en un estado desnaturalizado.

15 Después del tratamiento con el agente oxidante, los polinucleótidos en la primera porción pueden purificarse.

La purificación puede realizarse mediante el uso de cualquier técnica conveniente para la purificación de ácidos nucleicos. Las técnicas adecuadas para la purificación de ácidos nucleicos incluyen la cromatografía en columnas de centrifugación.

20 Los polinucleótidos pueden someterse a otras etapas de oxidación repetidas. Tales etapas se acometen para maximizar la conversión de 5-hidroximetilcitosina a 5-formilcitosina. Esto puede necesitarse donde un polinucleótido tiene suficiente estructura secundaria que es capaz de rehibridarse ("re-anneling"). Cualquier porción híbrida del polinucleótido puede limitar o impedir el acceso del agente oxidante a esa porción de la estructura, lo cual tiene el efecto de proteger la 5-hidroxicitosina de la oxidación.

25 En algunas modalidades, la primera porción de la población de polinucleótidos puede, por ejemplo, someterse a múltiples ciclos de tratamiento con el agente oxidante seguido de la purificación. Por ejemplo, pueden realizarse uno, dos, tres, o más de tres ciclos.

30 En algunas modalidades, la primera porción de la población de polinucleótidos que comprende la secuencia de nucleótidos de la muestra puede reducirse. En otras modalidades, la tercera porción de la población de polinucleótidos que comprende la secuencia de nucleótidos de la muestra puede reducirse. La reducción de la primera o tercera porciones de los polinucleótidos convierte los residuos de 5-formilcitosina en la secuencia de nucleótidos de la muestra en 5-hidroximetilcitosina

35 La primera o tercera porciones de los polinucleótidos puede reducirse mediante tratamiento con un agente reductor. El agente reductor es un agente adecuado para la generación de un alcohol de un aldehído. El agente o condiciones reductoras que se usan en la etapa de reducción pueden seleccionarse de modo que cualquier 5-formilcitosina se reduce selectivamente (es decir, el agente reductor o las condiciones reductoras son selectivas para 5-formilcitosina). De esta manera, sustancialmente ninguna otra funcionalidad en el polinucleótido se reduce en la etapa de reducción. Los agentes o condiciones reductoras se seleccionan para minimizar o evitar cualquier degradación del polinucleótido.

40 Los agentes reductores adecuados se conocen bien en la técnica e incluyen NaBH₄, NaCNBH₄ y LiBH₄. Los agentes reductores particularmente útiles son aquellos que pueden usarse en condiciones acuosas, como tales son más convenientes para la manipulación de los polinucleótidos. Sin embargo, los agentes reductores que son adecuados para usar en solventes orgánicos pueden usarse además donde sea factible.

45 Después de la oxidación y la reducción respectivamente, la primera porción de la población, y opcionalmente la tercera porción, se tratan con bisulfito. Una segunda porción de la población que no se oxida o reduce se trata, además, con bisulfito.

50 El tratamiento con bisulfito convierte en uracilo ambos residuos de citosina y 5-formilcitosina en un polinucleótido. Como se estableció anteriormente, donde esté presente cualquier 5-carboxicitosina (como un producto de la etapa de oxidación), esta 5-carboxicitosina se convierte en uracilo en el tratamiento con bisulfito. Sin desear estar atado a la teoría, se cree que la reacción de 5-formilcitosina ocurre a través de la pérdida del grupo formilo hasta producir citosina, seguido de una desaminación subsecuente para dar uracilo. Se cree que 5-carboxicitosina produce uracilo a través de una secuencia de etapas de descarboxilación y desaminación. El tratamiento con bisulfito puede realizarse bajo condiciones que conviertan tanto a la citosina como a los residuos de 5-formilcitosina o 5-carboxicitosina del polinucleótido, en uracilo, como se describe en la presente descripción.

60 Una porción de la población puede tratarse con bisulfito mediante la incubación con iones de bisulfito (HSO₃²⁻).

65 El uso de los iones de bisulfito (HSO₃²⁻), para convertir las citosinas no metiladas del ácido nucleico en uracilo, es común en la técnica y los reactivos y condiciones adecuadas se conocen bien por un experto en la técnica (39-42). Numerosos protocolos y reactivos están, además, disponibles comercialmente (por ejemplo, EpiTect™, Qiagen NL; EZ ADN Methylation™ Zymo Research Corp CA; CpGenome Turbo Bisulfite Modification Kit; Millipore).

- Una característica de los métodos descritos en la presente descripción es la conversión de citosina no metilada (la cual puede generarse in situ a partir de 5-formilcitosina o 5-carboxicitosina) en uracilo. Esta reacción se logra típicamente a través del uso de bisulfito. Sin embargo, en los aspectos generales de la invención, cualquier reactivo o condiciones de reacción pueden usarse para efectuar la conversión de citosina en uracilo. Tales reactivos o condiciones se seleccionan de manera que poca o ninguna 5-metilcitosina reaccione, y más específicamente, de manera tal que poca o ninguna 5-metilcitosina reaccione para formar uracilo. El reactivo, u opcionalmente un reactivo adicional, pueden, además, efectuar la conversión de 5-formilcitosina o 5-carboxicitosina en citosina o uracilo.
- Después de la incubación, las porciones de los polinucleótidos pueden inmovilizarse, lavarse, desulfonarse, eluirse y/o tratarse de otro modo según se requiera.
- En algunas modalidades, la primera, segunda, y tercera porciones de polinucleótidos de la población pueden amplificarse después del tratamiento como se describe anteriormente. Esto puede facilitar la manipulación y/o la secuenciación posterior. Las alteraciones de las secuencias en la primera, segunda y tercera porciones de polinucleótidos se conservan después de la amplificación. Las técnicas de amplificación de polinucleótidos adecuadas se conocen bien en la técnica e incluyen el PCR. La presencia de un residuo de uracilo (U) en una posición en la primera, segunda y/o tercera porciones del polinucleótido puede indicarse o identificarse por la presencia de un residuo de timina (T) en esa posición en el polinucleótido amplificado correspondiente.
- Como se describió anteriormente, los polinucleótidos pueden adaptarse después de la oxidación, reducción, o tratamiento con bisulfito para ser compatibles con una técnica o plataforma de secuenciación. La naturaleza de la adaptación dependerá de la técnica o plataforma de secuenciación. Por ejemplo, para la secuenciación Solexa-Illumina, los polinucleótidos tratados pueden fragmentarse, por ejemplo mediante sonicación, o tratamiento restrictivo con endonucleasas, los extremos libres de los polinucleótidos repararse según se requiera, y los cebadores ligarse en los extremos.
- Los polinucleótidos pueden secuenciarse mediante cualquier técnica o plataforma de secuenciación de alto o bajo rendimiento conveniente, que incluyen la secuenciación de Sanger (43), la secuenciación Solexa-Illumina (44), la secuenciación basada en ligación (SOLiD™) (45), la pirosecuenciación (46); secuenciación estroboscópica (SMRT™) (47, 48); y la secuenciación en chip semiconductores (Ion Torrent™) (49).
- Los protocolos, reactivos y equipamientos adecuados para la secuenciación de polinucleótidos se conocen bien en la técnica y están disponibles comercialmente.
- Los residuos en posición en la primera, segunda, y/o tercera secuencia de nucleótidos que se correspondan a citosina en la secuencia de nucleótidos de la muestra pueden identificarse.
- La modificación de un residuo de citosina en una posición en la secuencia de nucleótidos de la muestra puede determinarse a partir de la identidad de los residuos en las posiciones correspondientes en la primera, segunda, y opcionalmente, tercera secuencias de nucleótidos, como se describe anteriormente.
- La extensión o la cantidad de la modificación de citosina en la secuencia de nucleótidos de la muestra pueden determinarse. Por ejemplo, puede determinarse la proporción o cantidad de 5-hidroximetilcitosina y/o 5-metilcitosina en la secuencia de nucleótidos de la muestra en comparación con las citosinas no modificadas.
- Los polinucleótidos como se describen en la presente descripción, por ejemplo la población de polinucleótidos o 1, 2, 3, o los 4 de la primera, segunda, tercera y cuarta porciones de la población, pueden inmovilizarse sobre un soporte sólido.
- Un soporte sólido es un cuerpo insoluble, no gelatinoso que presenta una superficie sobre la cual los polinucleótidos pueden inmovilizarse.
- Los ejemplos de soportes adecuados incluyen láminas de cristal, pocillos, membranas o microperlas. Los soportes pueden ser en forma de partículas o sólidos, que incluyen por ejemplo una placa, un tubo de ensayo, perlas, bolilla, filtro, tela, polímero o membrana. Los polinucleótidos pueden, por ejemplo, fijarse a un polímero inerte, una placa de 96 pocillos, otro dispositivo, aparato o material que se use en la secuenciación de ácidos nucleicos u otros contextos de investigación. La inmovilización de un polinucleótido a la superficie de un soporte sólido se conoce bien en la técnica. En algunas modalidades, el soporte sólido mismo puede inmovilizarse. Por ejemplo, las microperlas pueden inmovilizarse sobre una segunda superficie sólida.
- En algunas modalidades, la primera, segunda, tercera y/o cuarta porciones de la población de polinucleótidos pueden amplificarse antes de la secuenciación. Preferentemente, las porciones de polinucleótidos se amplifican después del tratamiento con bisulfito.
- Los métodos adecuados para la amplificación de polinucleótidos se conocen bien en la técnica.

Después de la amplificación, las porciones amplificadas de la población de polinucleótidos pueden secuenciarse.

5 Las secuencias de nucleótidos pueden compararse y los residuos en posiciones en la primera, segunda, tercera y/o cuarta secuencias de nucleótidos que se correspondan a citosina en la secuencia de nucleótidos de la muestra pueden identificarse, mediante el uso de un análisis de secuencia basado en computación.

10 Pueden identificarse las secuencias de nucleótidos, tal como islas de CpG, con modificación de citosina mayor que un valor umbral. Por ejemplo, pueden identificarse una o más secuencias de nucleótidos en las cuales más del 1 %, más del 2 %, más del 3 %, más del 4 % o más del 5 % de las citosinas están hidroximetiladas.

15 El análisis de secuencia basado en computación puede realizarse mediante el uso de cualquier sistema de computación y programas convenientes. Un sistema de computación típico comprende una unidad de procesamiento central (CPU), recursos para entrada, salida y almacenamiento de datos (tal como RAM). Se proporciona preferentemente un monitor u otro dispositivo para mostrar imagen. El sistema de computación puede operarse vinculado a un secuenciador de ADN y/o ARN.

20 Por ejemplo, un sistema de ordenador puede comprender un procesador adaptado para identificar las citosinas modificadas en una secuencia de nucleótidos de la muestra mediante comparación con la primera, segunda y/o tercera secuencia de nucleótidos como se describe en la presente descripción. Por ejemplo el procesador puede adaptarse;

(a) identificar las posiciones de los residuos de citosina en la secuencia de nucleótidos de la muestra,

(b) identificar los residuos en la primera, segunda, y/o tercera secuencias de nucleótidos en las posiciones de los residuos de citosina en la secuencia de nucleótidos de la muestra.

25 (c) determinar a partir de las identidades de dichos residuos la presencia o ausencia de modificaciones de los residuos de citosina en las posiciones en la secuencia de nucleótidos de la muestra.

30 La secuencia de nucleótidos en la muestra y la primera, segunda y tercera secuencias de nucleótidos pueden introducirse automáticamente desde el secuenciador de ADN y/o ARN. Las secuencias pueden mostrarse, por ejemplo en un monitor.

35 El sistema de ordenador puede comprender adicionalmente un dispositivo de memoria para el almacenamiento de los datos. Las secuencias de nucleótidos tales como las secuencias genómicas y las posiciones de 5fC, 5hmC, y otros residuos de citosina modificada pueden almacenarse en el mismo u otros dispositivos de memoria, y/o pueden enviarse a un dispositivo de salida o mostrarse en un monitor. Esto puede facilitar el mapeo de las citosinas modificadas tales como 5hmC y 5fC, en el ADN genómico.

40 La identificación y el mapeo de las modificaciones de citosina, tales como 5fC y 5 hmC, en el genoma pueden usarse en el estudio del desarrollo y función neurológica, y en la diferenciación, división y proliferación celular, así como en el pronóstico y diagnóstico de enfermedades tal como el cáncer.

45 La identificación y/o el mapeo de las citosinas modificadas, tales como 5fC y 5 hmC, mediante el uso de métodos descritos en la presente descripción pueden por tanto utilizarse en enfermedades

"y/o" donde se usa en la presente descripción debe tomarse como descripción específica de cada uno de los dos componentes o características que se especifican con o sin el otro. Por ejemplo "A y/o B" debe tomarse como una descripción específica de cada uno de (i) A, (ii) B, y (iii) A y B, justo como si cada uno se estableciera individualmente en la presente descripción.

50 Ciertos aspectos y modalidades de la invención se ilustrarán ahora mediante la forma de ejemplos y con referencias a las figuras descritas debajo.

55 La figura 1 muestra el método para la secuenciación con resolución de una sola base de 5hmC. La Fig. 1 A muestra la reacción de 2'-deoxi-5-formilcitosina (d5fC) con NaHSO₃ (bisulfito) que se detiene por NaOH a diferentes puntos de tiempo, después se analiza mediante la cromatografía líquida de alto rendimiento (HPLC). Las barras de error representan la desviación estándar de 3 réplicas. La Fig. 1 B muestra el esquema de la reacción oxidativa del bisulfito: la oxidación de 5hmC a 5fC seguido por el tratamiento con bisulfito y NaOH para convertir 5fC en U. El grupo R es ADN. La Fig. 1 C muestra un diagrama y una tabla que bosquejan las técnicas BS-Seq y oxBS. La técnica BS-seq consiste en el tratamiento con bisulfito del ADN inicial y después la amplificación seguida de la secuenciación. La técnica oxBS-Seq consiste de la oxidación del ADN inicial, seguido por el tratamiento con bisulfito y la amplificación y después, la secuenciación. Mediante la comparación de la muestra inicial y la final procedentes de las técnicas BS-Seq y oxBS, C, 5mC y 5hmC pueden discriminarse, mapearse y cuantificarse.

60 La figura 2A muestra el perfil de bisulfito de 2'-deoxi-5-formilcitosina que indica una descarboxilación global más una desaminación hasta formar uracilo. La figura 2B muestra el perfil de bisulfito de 2'-deoxi-5-carboxicitosina que indica una descarboxilación a citosina y después una desaminación hasta formar uracilo.

La figura 3 muestra la cuantificación por espectrometría de masa de la oxidación (Fig. 3A, 3B, 3D) y el tratamiento oxidativo de bisulfito mediante secuenciación Illumina (Fig. 3C). La Fig. 3A muestra los niveles de 5hmC y 5fC (área de los picos normalizadas a T) en un oligonucleótido de ADNsc de 15 mer, antes y después de la oxidación por KRuO₄. La Fig. 3B muestra los niveles de 5hmC y 5fC (concentración normalizada a 5mC en la secuencia del cebador) en un fragmento de ADNdc de 135 mer, antes y después de dos oxidaciones secuenciales con KRuO₄. La Fig. 3C muestra los niveles de conversión de C en T, como determina la secuenciación Illumina, de dos fragmentos de ADNdc que contienen tanto un único 5hmCpG (122 mer) o múltiples 5hmCpG (135 mer) que siguieron el tratamiento con bisulfito oxidativo (se obtuvieron al menos 950, 000 lecturas por base). La 5mC además, estaba presente en esas cadenas para la comparación de las tasas de conversión. La Fig. 3D muestra los niveles de 5hmC y 5fC (concentración normalizada a 5mC en la secuencia del cebador) en el ADN de las células ES (J1) que se midieron antes y después de la oxidación con KRuO₄. Todas las barras de error representan las desviaciones estándar.

La figura 4 muestra la extensión de la degradación de la citosina después de la oxidación como se determinó mediante las mediciones de los cambios en las relaciones de nucleósidos después de la oxidación con KRuO₄ en un ADN de simple cadena (sc) sintético de 15 mer que contiene C (tres réplicas) (Fig. 4A), en un ADNsc sintético de 15 mer que contiene 5mC (tres réplicas) (Fig. 4B) y en un ADN genómico de células ES J1(2 réplicas del sonificado y 2 del no sonificado) (Fig. 4C). El porcentaje de cambio que se mide mediante análisis por HPLC del área del pico de nucleósido después de la oxidación dividido por el área del pico antes de la oxidación. Las barras de error representan las desviaciones estándar.

La figura 5 muestra un cromatograma de HPLC de los nucleósidos que se obtienen mediante la digestión de una molécula de ADN de 140 bp que contiene 5hmC, antes (Fig. 5A) y después (Fig. 5B) de la oxidación.

La figura 6 muestra un cromatograma de HPLC de los nucleósidos que se obtienen mediante la digestión de una cadena de ADN de 15 bp que contiene 5fC, antes (Fig. 6A) y después (Fig. 6B) del tratamiento con bisulfito.

La figura 7 muestra un cromatograma de HPLC de los nucleósidos que se obtienen mediante la digestión de una molécula de ADN de 140 bp que contiene 5fC, antes (Fig. 7A) y después (Fig. 7B) de la reducción.

La figura 8 muestra la secuencia de Sanger de la cadena de ADN de 122 mer con un sitio ClaI (ATCGAT) que contiene ya sea C, 5mC o 5hmC después del tratamiento oxBS. El cromatograma muestra la secuencia opuesta a la cadena molde. En el C ADN la C en la cadena opuesta se convierte totalmente en U, en el cromatograma se muestra una A en lugar de una G. El 5mC ADN no se convierte, y aparece una G en el cromatograma. El 5hmC ADN se convierte en su mayoría, en el cromatograma se muestra una A, con pequeñas trazas de G no convertida en esta corrida.

La figura 9 muestra la cuantificación de los niveles de 5mC y 5hmC en las CGI por oxRRBS. La figura 9A muestra una comparación de la fracción de citosinas no convertidas por CGI entre el conjunto de datos de RRBS y oxRRBS; las CGI con fracciones menores estadísticamente significativas en el conjunto de datos de oxRRBS (rojo) son CGI hidroximetiladas; una tasa de descubrimientos falsos de 3.7 % se estimó mediante el uso del número de CGI con el patrón opuesto (negro). La figura 9B muestra la distribución de los niveles de 5mC y 5hmC dentro de las CGI con niveles significativos de la respectiva modificación. La figura 9C muestra ejemplos de los perfiles de RRBS genómico y oxRRBS superpuestos con los perfiles de (h)MeDIP-Seq (6). Las CGI se indican por las barras verdes; para los propósitos de claridad los datos fuera de las CGI se ocultaron (áreas grises). Cada barra en el carril de oxRRBS representa un único CpG (en cualquier cadena de ADN). Las áreas amplificadas en la parte inferior del panel resaltan la resolución de un solo nucleótido del método. La figura 9D muestra que los niveles de 5mC y 5hmC en CGI seleccionadas se validaron mediante el uso de glucMS-qPCR. Los valores de oxRRBS en sitios MspI individuales se muestran, con barras de error que representan el 95 % de los intervalos de confianza. El glucMS-qPCR se realizó en duplicado, con la barra que representa el valor de la media y los puntos negros las réplicas individuales. Las dos técnicas muestran una buena correlación.

La figura 10 muestra un cromatograma de HPLC de una cadena de ARN (Sec. con núm. ident.: 7) que se digirió hasta nucleósidos, antes (Fig. 10A) y después (Fig. 10B) de la oxidación con KRuO₄. Se usaron las mismas condiciones que para la oxidación de ADN. Los tiempos de retención de los nucleósidos fueron como sigue: C - 1.2 min, U - 1.7 min, G - 3.5 min, A - 6.5 min.

La figura 11 muestra el gráfico de una secuencia de Sanger para una cadena de ADN sintético de 100 mer que contiene una 5-formilcitosina (5fC) (se muestra la secuencia parcial - ACGGA5fCGTA) después de la reducción con NaBH₄ y el tratamiento con bisulfito (redBS-Seq). El cromatograma muestra el complemento reverso de la secuencia parcial (TACG**T**CCAT - donde las posiciones resaltadas en negrita provienen de 5fC o C). Las posiciones de 5fC y C (en paréntesis) se muestran en la cadena molde en la figura 11. Tanto 5fC como C se desaminan bajo las condiciones del bisulfito. Sin embargo, 5fC se convierte en 5hmC mediante la etapa de reducción y no se desamina, mientras que la desaminación de C no se afecta. Esto permite la discriminación de 5fC y C con la resolución de una única base.

La tabla 1 muestra los resultados de la secuenciación para la citosina y citosinas modificadas sometidas a diversos tratamientos.

La tabla 2 muestra las estructuras de la citosina (1a), 5-metilcitosina (5mC; 1b), 5-hidroximetilcitosina (5hmC; 1c) y 5-formilcitosina (5fC; 1d)

5 La tabla 3 muestra un resumen de las eficiencias de oxidación de 5hmC en el ADN para algunos ejemplos de oxidantes solubles en agua.

Las tablas 4 y 5 muestran los tiempos de retención por los picos en los cromatogramas de HPLC de ADN (Figuras 4, 5, y 6) y ARN (Figura 10) respectivamente.

10

Experimentos

1. Métodos

15

1.1 Oxidación de d5hmCTP a d5fCTP y d5cCTP con MnO₂

2.5 µL d5hmCTP (100 mM, Bioline) en 497.5 µL H₂O con 51.6 mg MnO₂ (para d5fCTP) o 500 mg MnO₂ (para d5cCTP) (Alpha Aeser) se agitaron a 50 grados durante 2 h y 30 min. Después el MnO₂ se eliminó por filtración mediante el uso de columnas Amicon Ultra 0.5 mL 10 kDa (Millipore) y la muestra se liofilizó. El nucleótido trifosfato se resuspendió (5mM) y se desfosforiló con fosfatasa alcalina (New England Biolabs) toda la noche a 37°C.

20

1.2 Tiempo de exposición del bisulfito con los nucleósidos d5fC y d5cC

9 µL d5fC o d5cC (5 mM), 0.5 µL dA (0.1 M, Roche) y 2.5 µL H₂O se mezclaron y después se adicionaron 33 µL de NaHSO₃ 4 M (MP Biochemicals). Esto se dividió en tres reacciones de 15 µl y se dejaron a 50°C en la oscuridad. Se tomaron fracciones de 0.5 ul en varios puntos de tiempo y se procesaron en 2.5 µL H₂O y 2 µL NaOH (1 M). Después de permanecer por al menos 30 min a temperatura ambiente, ellos se inyectaron en el HPLC. Las áreas de los picos se midieron, correlacionados con una curva de calibración de d5fC, d5cC, dC o dU, y se estandarizaron a los niveles de dA en el cromatograma.

25

30

1.3 Digestión del ADN para análisis por HPLC

El ADN se digirió según un protocolo de la literatura (30), se purificó con columnas Amicon Ultra 0.5 mL 10 kDa y se analizó por HPLC mediante el uso de Agilent 1100 HPLC con un flujo de 1 mL/min sobre una columna Eclipse XDB-C18 3.5 µm, 3.0x150mm. La temperatura de la columna se mantuvo a 45 grados. Los tampones de elución son tampón A (Acetato de Amonio 500 mM (Fisher) pH 5), tampón B (Acetonitrilo) y tampón C (H₂O). El tampón A se mantuvo al 1 % a través de toda la corrida y el gradiente para el resto de los tampones fue 0 min - 0.5% B, 2 min - 1% B, 8 min - 4% B, 10 min - 95% B.

35

40

Los tiempos de retención de 2'-deoxinucleósidos son como sigue: 2'-deoxi-5-carboxicitidina (1.0 min), 2'-deoxicidina (1.8 min), 2'-deoxi-5-hidroximetilcitidina (2.1 min), 2'-deoxiuridina (2.7 min), 2'-deoxi-5-metilcitidina (4.0 min), 2'-deoxiguanosina (4.5 min), 2'-deoxi-5-formilcitidina (5.4 min), 2'-deoxitimidina (5.7 min), 2'-deoxiadeosina (7.4 min).

El mismo protocolo se usó para digerir el ARN para análisis de HPL C.

45

1.4 Secuencias de ADN de simple y doble cadena

Los oligos de 15 mer que contienen bien sea citosina, 5-metilcitosina, o 5-hidroximetilcitosina se adquirieron de IBA. Los cebadores de 122 mer y el molde de ADNdc de 135 mer se adquirieron de Biomers. Todas las C en los cebadores son 5-metilcitosina. Mediante PCR la 5-hidroximetilcitosina se adicionó a la cadena en todas las otras posiciones de citosina, mediante el uso de d5hmCTP y Fermentas DreamTaq Polymerase.

50

1.5 Reducción general

El ADN (aproximadamente 1 µg - 10 µL) se incubó en hielo durante 5 minutos con 40 µL de NaBH₄ (10,000 equivalentes por µL). Después esta reacción se agitó a 25 grados con una tapa abierta en la oscuridad, durante 1 hora. La reacción se purificó con columnas de oligo de centrifugación rápida (Roche).

55

1.6 Oxidaciones

60

Oxidación general

El ADN se completó hasta 24 µl con NaOH (concentración final 0,05 M) sobre hielo, después se añadió 1 µl de una solución de KRuO₄ (Alpha Aeser) (15 mM en NaOH 0,05 M) y la reacción se mantuvo en hielo durante 1 hora, con agitación ocasional. La reacción se purificó con una mini columna de oligonucleótidos de centrifugación rápida (Roche) (después de cuatro lavados con 600 µL de H₂O).

65

Estas condiciones se usaron, además, para la oxidación de ARN.

Oxidación de ADN de simple cadena

5 1 µg de ADNsc sintético de 15 mer se oxidó de acuerdo a la oxidación general.

Oxidación de ADN sintético de doble cadena

10 El ADNdc se precipitó con etanol y después se filtró a través de una mini columna de centrifugación rápida (después de cuatro lavados con 600µl de H₂O). Para el ADNdc sintético se requirió una doble oxidación ya que la desnaturalización con NaOH no es 100 % efectiva con una solución de un solo fragmento de ADN homólogo (a diferencia del ADN genómico).

15 1 µg de ADN se desnaturalizó en NaOH 0.05 M (volumen total 19 µL) durante 30 min a 37°C. Después la reacción se enfrió rápidamente en hielo y se dejó durante 5 min. Posteriormente la reacción se oxidó de acuerdo a la oxidación general pero en un volumen total de 20 µL. Este ADN se redensnaturalizó en NaOH 0.05 M (volumen total 24 µL) durante 30 min a 37°C. Nuevamente la reacción se enfrió rápidamente en hielo, se dejó durante 5 min y se oxidó de acuerdo a la oxidación general.

20 Reacción general de oxidación para ADN genómico.

25 El ADN (1 µg o menos) se precipitó con etanol antes de la oxidación, después se filtró a través de una mini columna de oligo de centrifugación rápida (después de cuatro lavados con 600 µL de H₂O). El ADN se desnaturalizó en NaOH 0.05 M (volumen total 24 µL o 40 µL) durante 30 min a 37°C. Después este se enfrió rápidamente en hielo, se dejó durante 5 min y se oxidó de acuerdo a la oxidación general.

1.7 Secuenciación de Sanger e Illumina de ADNdc tratado con bisulfito oxidativo

30 Para la secuenciación de Sanger, 1 µg de ADN de 122 mer que contiene C, 5mC y 5hmC se oxidó de acuerdo con la doble oxidación de ADNdc y se trató con bisulfito mediante el uso del juego de Qiagen Epitect, de conformidad con las instrucciones del fabricante para muestras FFPE, excepto que el ciclo térmico se realizó dos veces más. Después esas muestras se sometieron a la secuenciación de Sanger (fuente BioScience).

35 Para la secuenciación Illumina, 1 µg de ADN de 122 mer y de 135 mer que contienen 5hmC se digirieron durante la noche con Dral (2 µl, New England Biolabs) y Sspl (1 µl, New England Biolabs). Las bandas digeridas se purificaron con el juego de extracción de gel Fermentas GeneJET y se ligaron los adaptadores metilados (Illumina) mediante el uso de la mezcla maestra de preparación de muestra NEBNext ADN. Después de la oxidación y del tratamiento con bisulfito como anteriormente, los fragmentos ligados se amplificaron (18 ciclos) mediante el uso de Pfu Turbo Cx (Agilent) y los cebadores específicos del adaptador (Illumina), seguido de purificación mediante perlas AMPure XP (Agencourt).

40 1.8 Espectrometría de masa

45 Los nucleósidos se derivaron del ADN mediante la digestión con la ADN Degradase Plus (Zymo Research) de acuerdo con las instrucciones del fabricante y se analizaron mediante LC-MS/MS en un espectrómetro de masas LTQ Orbitrap Velos (Thermo Scientific) equipado con una fuente de iones nanoelectrospray (Proxeon). Se adquirieron los datos del espectro de masas para 5hmC, 5fC y, donde fuera aplicable, 5mC y T, en modo de barrido de alta resolución (R > 40.000 para los iones pseudomoleculares protonados y > 50.000 para los fragmentos de iones de base protonados acompañantes), y además, en el modo de monitoreo de reacciones seleccionadas (SRM), se monitorearon las transiciones 258 -> 142.0611 (5hmC), 256 -> 140.0455 (5fC), 242 -> 126.0662 (5mC) y 243 -> 127.0502 (T). Los iones parentales se seleccionaron para SRM con una ventana de aislamiento de 4 unidades de masa y se fragmentaron mediante HCD con una energía de colisión relativa del 20 %, con R> 14.000 para los fragmentos de iones.

50 Las áreas de los pico de los cromatogramas iónicos que se extrajeron de los iones correspondientes a 5hmC y 5fC se normalizaron con aquellas de las de 5mC (donde fuera aplicable) o T, y se cuantificaron mediante calibración externa con respecto a los patrones que se obtienen por digestión de nucleótidos trifosfatos u oligonucleótidos.

55 1.9 Cultivo de células ES y extracción de ADN

60 Las células J1 ES (129S4/SvJae) se adquirieron de la ATCC (Cat. SCRC-1010) y se cultivaron sobre una capa alimentadora pMEF irradiada con rayos γ, a 37°C y 5% CO₂ en medio ES completo (DMEM glucosa 4500 mg/l, L-glutamina 4 mM y piruvato de sodio 110 mg/l, suero fetal bovino 15%, penicilina 100 U / estreptomicina 100 µg en 100 ml de medio, amino ácidos no esenciales 0.1mM, β-mercaptoetanol 50 µM, 103U LIF ESGRO®). El ADN genómico se preparó de las células ES en el pase 14 o 20 mediante el uso del mini juego Qiagen Allprep ADN/ARN.

65 1.10 oxRRBS

Las librerías RRBS de ADN oxidado y no oxidado se prepararon en base a un protocolo publicado previamente (31). Brevemente, 2 µg de ADN genómico se digirieron con MspI (Fermentas) seguido de reparación final y adición de colas de A con Klenow (Fermentas) y ligación de los adaptadores metilados (Illumina) con la ADN T4-ligasa (NEB). El ADN digerido con MspI ligado al adaptador se corrió en un gel de agarosa al 3% y se seleccionó el tamaño de las bandas (110-380 bp), seguido de la purificación con el juego de purificación rápida en gel Qiagen QIAquick y precipitación con etanol.

Antes de la oxidación, el ADN del tamaño seleccionado se filtró a través de una mini columna de oligo de centrifugación rápida (después de cuatro lavados con 600 µl de H₂O) para eliminar cualquier remanente de tampones/sales y se ajustó a un volumen final de 25 µl. Se mantuvieron 5 µl de esta solución para la generación de la librería no oxidada. El resto se oxidó de acuerdo con la oxidación general para el ADN genómico.

Tanto las muestras de ADN oxidadas como las no oxidadas se trataron con bisulfito mediante el juego Qiagen Epitect, de acuerdo con las instrucciones del fabricante para las muestras FFPE, excepto que el ciclo térmico se realizó dos veces más. La amplificación de la librería final (18 ciclos) se realizó mediante el uso de cebadores Pfu Turbo Cx (Agilent) y adaptadores específicos (Illumina), después de lo cual las librerías se purificaron mediante el uso de perlas AMPure XP (Agencourt).

1.11 Secuenciación y alineación de lectura

La secuenciación (lecturas de 40 bp, de un solo extremo) se realizó en la plataforma Illumina GAllx. Las bases se nombraron por reprocesamiento de imágenes no procesadas mediante el uso de la versión 1.8 de OLB después de aplicar el método de procesamiento bareback a los tres primeros pares de bases (32). La lectura de los alineamientos de bisulfito en el genoma del ratón (versión del genoma NCBIM37) se realizaron mediante el uso de Bismark v0.6.4 (33), mediante el uso de las opciones -n 1 -l 40 - phred64-quals -- vanilla. Los alineamientos de Bismark a las secuencias individuales de monómeros 5' LINE1 se realizaron ligeramente más estrictamente (-n 0); se usaron secuencias consenso publicadas para los alineamientos de lecturas a subtipos de monómeros L1A (34), L1Tf y L1Gf (35).

Las tasas de conversión de bisulfito se estimaron a partir del número de citosinas no convertidas en los sitios MspI 3' rellenos con Klenow en las lecturas de secuencias que fueron suficientemente cortas para leer a través de esos sitios. El parámetro phred de calidad de lectura permaneció alto en los extremos 3'. Los porcentajes estimados de conversión de bisulfito variaron entre 99.8 % y 99.9 %.

1.12 Procesamiento de datos oxRRBS

Los números de citosinas convertidas y no convertidas dentro de CGI (25) se extrajeron de cada conjunto de datos BS y oxBS. Para cada posición CpG, se tomó la cantidad de 5mC como el porcentaje de citosinas no convertidas en cada conjunto de datos oxBS y se tomó la cantidad de 5hmC restando este valor del porcentaje de citosinas no convertidas en el conjunto de datos correspondiente a BS. Se calculó un valor global por CGI mediante la unificación de los datos de todos los CpG cubiertos dentro de cada CGI. Se excluyeron los CpG con menos de 10 lecturas, al igual que los CpG para los cuales la estimación de 5mC se desvió del valor global de 5mC del CGI en más del 20 %, o la estimación de 5hmC se desvió del valor global en más del 10 %. Después de la exclusión de los atípicos, sólo se analizaron las CGI con 5 o más CpG representativos.

Para evaluar las CGI que contenían niveles de 5mC significativamente por encima del error de conversión de bisulfito en el conjunto de datos oxBS, se aplicó una prueba binomial mediante el uso del valor de corte corregido de Benjamini-Hochberg de p 0.01. Del mismo modo, se utilizó una prueba binomial para seleccionar las CGI con cantidades significativas de citosinas no convertidas en el conjunto de datos BS; dentro de éstas, las diferencias entre los conjuntos de datos BS y oxBS se evaluaron mediante la aplicación de la prueba de Fisher y mediante el uso de un valor de corte corregido de p 0.05. Las CGI con una fracción significativamente menor de citosinas no convertidas en el conjunto de datos oxBS se tomaron como CGI hidroximetiladas. Las CGI con el patrón opuesto se asumieron como artefactos y se utilizaron para estimar una tasa de descubrimientos falsos.

1.13 GlucMS-qPCR

La cuantificación de los niveles de 5mC y 5hmC en sitios MspI por glucMS-qPCR se realizó como se describe anteriormente (6).

2. Resultados

Nosotros hemos seguido una estrategia que discriminaría 5mC de 5hmC en el ADN mediante el aprovechamiento de la reactividad química que es selectiva para 5hmC, en particular, mediante la eliminación química del grupo hidroximetilo, y que por tanto transforma 5hmC en C, la cual podría ser fácilmente transformada en U por la desaminación mediada por bisulfito. Durante nuestros estudios de reactividad química en 5-formilcitosina (5fC), observamos la descarbonilación y la desaminación de 5fC a uracilo (U) bajo las condiciones de bisulfito que dejarían la 5mC en la misma forma (Fig. 1 A).

Esta transformación, que no se describió previamente, indicó que la secuenciación de 5hmC podría realizarse mediante la oxidación selectiva de 5hmC en 5fC y después convertir 5fC en U en un procedimiento de dos etapas (Fig. 1B). Mientras que la BS-Seq convencional indica que ambas 5mC y 5hmC se detectan como C, este procedimiento de "secuenciación oxidativa con bisulfito" (oxBS-Seq) solamente produce C en los sitios de 5mC y por tanto nos permite determinar la cantidad de 5hmC en una posición particular del nucleótido mediante la comparación de las lecturas de BS-Seq y oxBS-Seq (Fig. 1C).

Se determinó el perfil de bisulfito para 2'deoxi-5-formilcitosina y 2'deoxi-5-carboxicitosina (Fig. 2A y 2B). La 2' deoxi-5-formilcitosina y 2' deoxi-5-carboxicitosina se incubaron con NaHSO₄ 2.9 M. Se tomaron pequeñas muestras de la reacción en diferentes puntos de tiempo y se procesaron en NaOH 0.3 M. Estas se inyectaron directamente en un HPLC para análisis. Los perfiles de HPLC son consistentes con descarboxilación o descarbonilación global, respectivamente, a citosina seguido por una desaminación rápida a uracilo.

Por tanto, se requirió la oxidación específica de 5hmC a 5fC mediante el uso de un oxidante que fuera moderado, compatible con el medio acuoso y selectivo sobre otras bases y el esqueleto del ADN. Se evaluaron una serie de oxidantes solubles en agua potencialmente adecuados (Tabla 3) y encontramos que el perrutenato de potasio (KRuO₄) posee las propiedades y la eficiencia de conversión que nosotros buscamos. El KRuO₄ puede, en principio, oxidar ambos, alcoholes y dobles enlaces carbono carbono (23). Sin embargo, en nuestros estudios de reactividad sobre un ADN sintético de simple cadena de 15 mer (ADNsc) que contiene 5hmC, establecimos condiciones bajo las cuales la reactividad del KRuO₄ fue altamente específica para el alcohol primario de 5hmC (conversión cuantitativa de 5hmC en 5fC por espectrometría de masa, Fig. 3A). El ADNsc de 15 mer que contiene C o 5mC, más que 5hmC, no mostró ninguna reacción base específica con KRuO₄ (fig. 4A, B). Fuimos conscientes, además, de que las oxidaciones con KRuO₄ podrían proceder hacia el ácido carboxílico (23), sin embargo en el contexto de 5hmC en el ADN, solamente observamos el aldehído (5fC), aun con un exceso moderado del oxidante. La oxidación con KRuO₄ es además, capaz de oxidar 5hmC en muestras que se presentan como ADN de doble cadena (ADNdc), con una etapa inicial de desnaturalización antes de la adición del oxidante; esto resulta en un rendimiento cuantitativo de 5hmC en 5fC, como se distingue mediante la espectrometría de masa (Fig. 3B).

Se preparó una molécula de ADN de 140 bp (Sec. con núm. ident.: 1) la cual contiene 45 nucleósidos 5hmC que se incorporaron a través del PCR mediante el uso de cebadores de 5-metilcitosina y hmCTP. El ADN se oxidó mediante el uso de KRuO₄. Antes y después de la oxidación, el ADN se digirió hasta nucleósidos con Benzonasa, Fosfodiesterasa I y Fosfatasa Alcalina. Después esta mezcla se inyectó en el HPLC, para proporcionar los cromatogramas que se muestran en las Figuras 5A (antes de la oxidación) y 5B (después de la oxidación). Se observó una conversión casi completa de 5hmC en 5fC, sin actividad sobre otros nucleósidos.

Una molécula de ADN de 15 bp de cadena simple (Sec. con núm. ident.: 2) que contiene 3 residuos de 5fC se trató con bisulfito como se describe anteriormente. Antes y después del tratamiento con bisulfito, el ADN se digirió hasta nucleósidos con Benzonasa, Fosfodiesterasa I y Fosfatasa Alcalina. Después, esta mezcla se inyectó en el HPLC, para proporcionar los cromatogramas que se muestran en las Figuras 6A y 6B.

Después del tratamiento con bisulfito, sólo quedó un pico muy pequeño para 5fC, y la citosina se presenta en cantidades insignificantes. El pico de uracilo en la figura 6B se derivó de 5fC, así como de la desaminación de C no modificada.

Se preparó una molécula de ADN de 140 bp (Sec. con núm. ident.: 1) que contiene 45 nucleósidos 5fC que se incorporaron mediante PCR. El ADN se reduce mediante el uso de NaBH₄ como se describe anteriormente. Antes y después de la reducción, las muestras de ADN se digirieron hasta nucleósidos con Benzonasa, Fosfodiesterasa I y Fosfatasa Alcalina. Después, esta mezcla de nucleósidos se inyectó en el HPLC, para proporcionar los cromatogramas que se muestran en las Figuras 7A (antes de la reducción) y 7B (después de la reducción). Se observó la conversión completa de 5fC en 5hmC.

Se investigó la conversión de bisulfito oxidado de un sitio ClaI (ATCGAT) en un ADN de doble cadena de 122 pares de bases (Sec. con núm. Ident: 3) para evaluar la eficiencia y la selectividad del método de bisulfito oxidativo. Se amplificó por PCR un fragmento de ADN de doble cadena de 122 pares de bases con un solo CpG en el centro (en el contexto de un sitio de restricción ClaI ATCGAT; Sec. con núm. ident: 3) mediante el uso de cebadores 5-metilcitosina y también CTP, 5mCTP o 5hmCTP. El producto amplificado contiene 5-metilcitosina en las regiones del cebador y CpG, 5mCpG o 5hmCpG en el centro CpG.

Como se describió anteriormente, los tres ADNdc sintéticos de 122 mer que contienen bien sea C, 5mC o 5hmC se oxidaron cada uno con KRuO₄ y después se sometieron a un protocolo convencional de conversión de bisulfito. Se realizó la secuenciación de Sanger en cada una de las tres cadenas (Fig.8).

La cadena que contiene C se convirtió completamente en U (Figura 8, panel de la izquierda), la cadena que contiene 5mC no se convirtió (Figura 8, panel central) y la cadena que contiene 5hmC casi se convirtió cuantitativamente en U, con una pequeña cantidad de C no convertida (Figura 8 panel de la derecha). Esto aparece como un pico mayor de adenina en el material convertido, y un pico de guanina residual que proviene de una minoría de material no convertido.

Para obtener una medida exacta de la eficiencia de conversión de 5hmC en U, después del tratamiento con bisulfito oxidativo se realizó la secuenciación de Illumina en la cadena sintética que contiene 5hmC. Se observó un nivel de conversión global de 5hmC en U de 94.5 % (Fig. 3C). El protocolo de bisulfito oxidativo se aplicó además a una segunda cadena que contiene múltiples residuos 5hmC en un rango de diferentes contextos y también mostró una eficiencia de conversión similarmente alta (94.7 %) de 5hmC en U (Figura 3C). Por último, se realizó la oxidación con KRuO_4 en el ADN genómico y se mostró por espectrometría de masas un rendimiento cuantitativo de la conversión de 5hmC en 5fC (Fig. 3D) sin degradación significativa de C (figura 4C). Estos experimentos de prueba de concepto demuestran que el protocolo de bisulfito oxidativo convertirá específicamente 5hmC en U en el ADN, dejando C y 5mC sin cambios, lo que permite la secuenciación cuantitativa con resolución de un solo nucleótido en plataformas ampliamente disponibles (oxBS-Seq).

Después, se usó el principio de bisulfito oxidativo para el mapeo cuantitativo de 5hmC a alta resolución en el ADN genómico de células ES de ratón. Optamos por combinar el bisulfito oxidativo con la secuenciación de bisulfito de representación reducida (RRBS) (24), lo que permite la secuenciación selectiva de una parte del genoma altamente enriquecida en islas CpG (CGI), lo cual garantiza una adecuada profundidad de secuenciación para detectar estas marcas menos abundantes. Por tanto, generamos conjuntos de datos de RRBS y oxRRBS, que lograron una media de profundidad de lecturas de secuenciación de ~120 lecturas por CpG, los cuales, cuando se agruparon rindieron un promedio de ~ 3,300 metilaciones por CGI. Después de aplicar los límites de profundidad y amplitud (ver Materiales y Métodos), el 55% (12.660) de todos las CGI (25) se cubrieron en nuestros conjuntos de datos. Nuestros datos de RRBS (es decir, no oxidados) se correlacionan bien con los conjuntos de datos publicados de RRBS y BS-Seq (24, 26).

Para identificar las CGI que contienen 5hmC, evaluamos las diferencias entre los conjuntos de datos de RRBS y oxRRBS mediante el uso de criterios estrictos (ver Materiales y Métodos). Se esperaba que la mayoría de las diferencias significativas se derivaran de las CGI que tenían una menor proporción de citosinas no convertidas en el conjunto de oxRRBS en comparación con el conjunto de RRBS. Las CGI que tenían la tendencia inversa se utilizaron para estimar una tasa de descubrimientos falsos, que fue de 3.7 % (Figura 9A). Nosotros identificamos 800 CGI que contienen 5hmC, las cuales tenían un promedio de 3.3 % (intervalo 0.2-18.5 %) de CpG con hidroximetilación (Figuras 9A y B). Además, identificamos 4.577 CGI que contienen 5mC, que promedian el 8.1% de metilación de CpG (Figura 9B). Realizamos la secuenciación de una muestra biológica independiente, duplicada de la misma línea de células ES, pero con un número de pases diferente, la cual, por espectrometría de masas, tenía niveles reducidos de 5hmC (0.10 contra 0.16 de todas las C) y consistentemente encontramos menos CGI que contienen 5hmC. Es importante destacar, que las CGI que contienen 5hmC presentes en ambas muestras mostraron una buena reproducibilidad cuantitativa.

Para validar nuestro método, seleccionamos 21 CGI que contienen sitios de restricción *MspI* y cuantificamos los niveles de 5hmC y 5mC en estos CpG mediante glucMS-qPCR (28) (Figura 9D). Encontramos una buena correlación entre la cuantificación por oxRRBS y glucMS-qPCR ($r = 0.86$, $p = 5E-7$ y $r = 0.52$, $p = 0.01$ para 5mC y 5hmC, respectivamente),

Se investigó la conversión reducida de bisulfito (reBS-Seq) de la cadena de ADN que contiene una 5-formilcitosina (5fC).

Una cadena de ADN sintético de 100 mer (Sec. con núm. ident: 8) que contiene la secuencia ACGGA5fCGTA se expuso a una reducción con NaBH_4 , y después se sometió a un protocolo convencional de conversión de bisulfito. Después se realizó la secuenciación de Sanger sobre la cadena (Figura 11).

La Figura 11 muestra las curvas de secuenciación que representan parte de la secuencia complementaria inversa (TACGTCAT). Las posiciones de 5fC y C se resaltan en negrita y se muestran entre paréntesis en la cadena molde de la figura 11. Como se mostró previamente, tanto 5fC como C se desaminan bajo las condiciones de bisulfito para formar U, el cual se muestra como A en la secuencia complementaria inversa de la Figura 11. Sin embargo, la reducción con NaBH_4 convierte 5fC en 5hmC, la cual no se desamina en U, y se muestra como G en la secuencia complementaria inversa de la Figura 11. La secuenciación reducida de bisulfito (redBS-Seq) permite, por tanto, la discriminación de 5fC y C con la resolución de una sola base.

En resumen, demostramos que el método oxBS-Seq mapea y cuantifica confiablemente tanto 5mC como 5hmC a nivel de un solo nucleótido. El bisulfito oxidativo es, además, compatible con procedimientos que no dependen de la secuenciación tal como Sequenom, como se demuestra aquí. Por tanto, mediante la comparación de la secuencias de ADN genómico tratado con bisulfito y oxidado con la secuencia del ADN genómico tratado con bisulfito, es posible determinar la presencia de 5-metilcitosina y 5-hidroximetilcitosina, junto con citosina no modificada.

Por ejemplo, los residuos de uracilo en la misma posición en ambas secuencias, tanto la de ADN tratado con bisulfito y oxidado, como la de ADN genómico tratado con bisulfito, indican la presencia de citosina no modificada. Los residuos de citosina en la misma posición en ambas secuencias, tanto la de ADN tratado con bisulfito y oxidado, como la de ADN genómico tratado con bisulfito, indican la presencia de 5-metilcitosina. Un residuo de citosina en la secuencia de ADN genómico oxidado y tratado con bisulfito, además, indica la presencia de 5-metilcitosina. Un residuo de citosina en la secuencia de ADN genómico tratado con bisulfito y un residuo de uracilo en la misma posición en la secuencia de ADN genómico tratado con bisulfito y oxidado, indican la presencia de 5-hidroximetilcitosina.

La 5-formilcitosina, además, puede secuenciarse con resolución de un solo nucleótido. La 5fC puede reducirse cuantitativamente a hmC en el ADN genómico mediante el uso de NaBH₄ (como se muestra por HPLC). Mediante la comparación de las secuencias de ADN genómico no tratado, tratado con bisulfito y oxidado, tratado con bisulfito y reducido, y tratado con bisulfito, pueden determinarse la presencia de las tres modificaciones de citosina conocidas en mamíferos, 5-metilcitosina, 5-hidroximetilcitosina y 5-formilcitosina, junto con la citosina no modificada. Por ejemplo, los residuos de uracilo en la misma posición en las secuencias de ADN genómico (UUU) i) tratado con bisulfito, ii) tratado con bisulfito y oxidado, y iii) tratado con bisulfito y reducido indican la presencia de citosina no modificada.

Los residuos de citosina en la misma posición en las secuencias de ADN genómico (CCC) i) tratado con bisulfito, ii) oxidado y tratado con bisulfito y, iii) reducido y tratado con bisulfito indican la presencia de 5-metilcitosina.

Un residuo de citosina en la secuencia del ADN genómico tratado con bisulfito; un residuo de uracilo en la misma posición en la secuencia del ADN genómico oxidado y tratado con bisulfito y, opcionalmente, un residuo de citosina en la misma posición en la secuencia del ADN genómico reducido y tratado con bisulfito (CUC) indican la presencia de 5-hidroximetilcitosina.

Un residuo de uracilo en la secuencia del ADN genómico tratado con bisulfito; un residuo de citosina en la misma posición en la secuencia de ADN genómico reducido y tratado con bisulfito; y opcionalmente, un residuo de uracilo en la misma posición en la secuencia del ADN genómico oxidado y tratado con bisulfito (UCU) indican la presencia de 5-formilcitosina.

Cuando se secuencia el ADN genómico no tratado tanto las citosinas modificadas como las no modificadas se leen como citosina.

Los cromatogramas de HPLC que se muestran en la Figura 10 confirman que no se observó ninguna degradación significativa del ARN después de la oxidación de una cadena de ARN de 28 nucleótidos (Sec. con núm. ident.: 7). Este resultado significa que el procedimiento de oxidación es, además, compatible para la secuenciación de residuos de citosina modificadas, tal como 5hmC, en el ARN, como se describe en la presente descripción.

Referencias

1. A. M. Deaton y otros. *Genes Dev.* 25, 1010 (15 de mayo de 2011).
2. M. Tahiliani y otros. *Science* 324, 930 (15 de mayo de 2009).
3. S. Ito y otros. *Nature* 466, 1129 (26 de agosto de 2010).
4. A. Szwagierczak y otros. *Nucleic Acids Res.* (4 de agosto de 2010).
5. K. P. Koh y otros. *Cell Stem Cell* 8, 200 (4 de febrero de 2011).
6. G. Ficiz y otros. *Nature* 473, 398 (19 de mayo de 2011).
7. K. Williams y otros. *Nature* 473, 343 (19 de mayo de 2011).
8. W. A. Pastor y otros. *Nature* 473, 394 (19 de mayo de 2011).
9. Y. Xu y otros. *Mol. Cell* 42, 451 (20 de mayo de 2011).
10. M. R. Branco y otros. *Nat. Rev. Genet.* 13, 7 (enero, 2012).
11. S. Kriaucionis y otros. *Science* 324, 929 (15 de mayo de 2009).
12. M. Munzel y otros. *Angew. Chem. Int. Ed.* 49, 5375 (julio, 2010).
13. H. Wu y otros. *Genes Dev.* 25, 679 (1 de abril de 2011).
14. S. G. Jin y otros. *Nuc. Acids. Res.* 39, 5015 (julio, 2011).
15. C. X. Song y otros. *Nat. Biotechnol.* 29, 68 (enero, 2011).
16. M. Frommer y otros. *PNAS. U.S.A.* 89, 1827 (marzo 1992).
17. Y. Huang y otros. *PLoS One* 5, e8888 (2010).
18. C. Nestor y otros. *Biotechniques* 48, 317 (abril, 2010).
19. C. X. Song y otros. *Nat. Methods*, (20 de noviembre de 2011).
20. J. Eid y otros. *Science* 323, 133 (2 de enero de 2009).
21. E. V. Wallace y otros. *Chem. Comm.* 46, 8195 (21 de noviembre de 2010).
22. M. Wanunu y otros. *J. Am. Chem. Soc.*, (14 de diciembre de 2010).
23. G. Green, W y otros. *J Chem Soc Perk T 1*, 681 (1984).
24. A. Meissner y otros. *Nature* 454, 766 (7 de agosto de 2008).
25. R. S. Illingworth y otros. *PLoS genetics* 6, (septiembre, 2010).
26. M. B. Stadler y otros. *Nature* 480, 490 (22 de diciembre de 2011).
27. J. Borgel y otros. *Nat. Genet.* 42, 1093 (diciembre, 2010).
28. S. M. Kinney y otros. *J. Biol. Chem.* 286, 24685 (15 de julio de 2011).
29. N. Lane y otros. *Genesis* 35, 88 (febrero, 2003).
30. E. P. Quinlivan y otros 3rd, *Anal. Biochem.* 373, 383 (febrero, 2008).
31. H. Gu y otros. *Nat. Protoc.* 6, 468 (abril, 2011).
32. F. Krueger y otros. *PLoS One* 6, e16607 (2011).
33. F. Krueger y otros *Bioinformatics* 27, 1571 (1 de junio de 2011).
34. S. A. Schichman y otros *Mol. Biol. Evol.* 10, 552 (mayo, 1993).
35. J. L. Goodier y otros. *Genome research* 11, 1677 (octubre, 2001).

36. C. Qin y otros. *Mol. Carcinog.* 49, 54 (enero, 2010).
 37. Li y otros. *Nucleic Acids* (2011) Artículo ID 870726
 38. Pfaffeneder, T. y otros (2011) *Angewandte.* 50. 1-6
 39. Lister, R. y otros (2008) *Cell.* 133. 523-536
 5 40. Wang y otros (1980) *Nucleic Acids Research.* 8 (20), 4777-4790
 41. Hayatsu y otros (2004) *Nucleic Acids Symposium Series No. 48 (1), 261-262*
 42. Lister y otros (2009) *Nature.* 462. 315-22
 43. Sanger, F. y otros. *PNAS USA,* 1977, 74, 5463
 44. Bentley y otros, *Nature,* 456, 53-59 (2008)
 10 45. KJ McKernan y otros. *Genome Res.* (2009) 19: 1527-1541
 46. M Ronaghi y otros. *Science* (1998) 281 5375 363-365
 47. Eid y otros *Science* (2009) 323 5910 133-138
 48. Korlach y otros. *Methods in Enzymology* 472 (2010) 431-455)
 49. Rothberg y otros (2011) *Nature* 475 348-352).

15
 Secuencias modelos

Los nucleótidos modificados se resaltan en negrita e itálica

20 Modelo de ADN de doble cadena de 140 pares de bases (Sec. con núm. ident.:1):

CACATCCCACACTATACTCATACTCATACTACCTGCTCAGCAGCAGCTGTACACCTACGTACTCGTGCACGCTCGTC
 ACGTGATCGACCATGACTCTGACGCACTGAGGTATGGGAAGTAGTGAGTAGATTGTAGTAAGGAG

25 Modelo de ADN de simple cadena de 15 nucleótidos de longitud (Sec. con núm. ident.:2):

GAGACGACGTACAGG

Modelo de ADN de doble cadena de 122 pares de bases (Sec. con núm. ident.:3):

30 CACATCCCACACTATACTCATACTACATACCATTTAAATAAAATTAATAATATTAATATATCGATTAAATAATAAAT
 AATAAATTAATAATATTGGGAAGTAGTGAGTAGATTGTAGTAAGGAG

Modelo de ADN de doble cadena de 135 pares de bases (Sec. con núm. ident.:4):

35 CACATCCCACACTATACTCATACTACATACCATTTAACGATAAAATTACAATAACGTATCTAATCATATCGATTAAAC
 TAATCGAAATAATAATTACGCATTAATATTGGGAAGTAGTGAGTAGATTGTAGTAAGGAG

Cebador sentido ADNdc (Sec. con núm. ident.:5):

40 CACATCCCACACTATACTCATACTACATACC

Cebador antisentido ADNdc (Sec. con núm. ident.:6):

45 CTCCTTACTACAATCTACTACTACTTCCC

Secuencia del Modelo de ARN de 28 nucleótidos (Sec. con núm. ident.:7):

50 UGUGGGGAGGGCGGGGCGGGGUCUGGGG

Secuencia que contiene 100 nucleótidos 5fC (Sec. con núm. ident.:8)

[La posición de 5fC se indica en negrita, itálica]

55 GACGGACGTACGATCGAGCGAGGCTTTGGGTCAGCAGGTGGCGACTGTTAGCTCAGATGGCTAGCAAGTGGGTAT
 GTATGAGTGTATAGTGTGGGATGTG

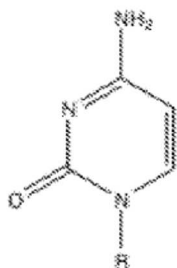
60

65

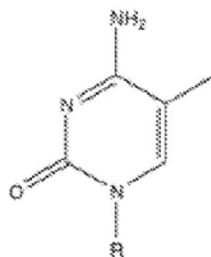
5	Base	Secuenciación Regular	Secuenciación por Bisulfito	Oxidación seguida de Secuenciación por Bisulfito	Reducción seguida de Secuenciación por Bisulfito
	C	C	U	U	U
	5mC	C	C	C	C
10	5hmC	C	C	U	C
	5fC	C	U	U	C

Tabla 1

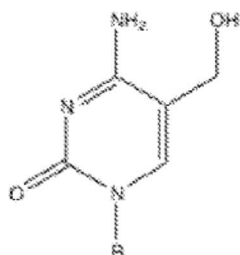
a)



b)



c)



d)

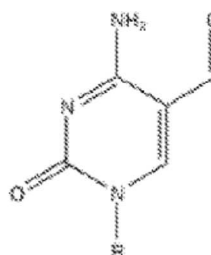


Tabla 2

Oxidante	Comentario
KRuO ₄	Conversión completa a aldehído
CrO ₃	No se observa oxidación
PDC	No se observa oxidación
PCC	No se observa oxidación
MnO ₂	Se observó una pequeña cantidad de aldehído, pero una degradación sustancial con el exceso de oxidante

Tabla 3

5

10

15

Base	Tiempo de retención/min
C	1.8
5hmC	2.1
U	2.7
G	4.5
5fC	5.3
T	5.7
A	7.3

Tabla 4

Tiempo de retención de los picos de HPLC (ADN)

20

25

30

Base	Tiempo de retención/min
C	1.3
U	1.8
G	3.7
A	6.7

Tabla 5

Tiempo de retención de los picos de HPLC (ARN)

35

Reivindicaciones

1. Un método para la identificación de residuos de citosina modificada en una secuencia de nucleótidos de la muestra que comprende;
 - i) proporcionar una población de polinucleótidos que comprende la secuencia de nucleótidos de la muestra,
 - ii) Oxidar o reducir una primera porción de dicha población,
 - iii) Tratar la primera porción, oxidada o reducida, de dicha población y una segunda porción de dicha población con bisulfito,
 - iv) secuenciar los polinucleótidos en la primera y segunda porciones de la población que siguieron las etapas ii) y iii) para producir una primera y segunda secuencias de nucleótidos respectivamente e;
 - v) identificar el residuo en la primera y segunda secuencias de nucleótidos que corresponda a un residuo de citosina en la secuencia de nucleótidos de la muestra, en donde la citosina modificada es 5-metilcitosina (5mC), 5-hidroximetilcitosina (5hmC) o 5-formilcitosina (5fC).
2. Un método de conformidad con la reivindicación 1 que comprende oxidar la primera porción de dicha población, tratar la porción oxidada con bisulfito y tratar una segunda porción de la población con bisulfito.
3. Un método de conformidad con la reivindicación 2 en donde la primera porción se oxida mediante el uso de un agente químico oxidante.
4. Un método de conformidad con la reivindicación 3 en donde la primera porción se oxida mediante el uso de un agente oxidante perrutenato.
5. Un método de conformidad con cualquiera de las reivindicaciones 2 a 4 en donde la identificación de uracilo en una posición en la primera secuencia de nucleótidos que corresponde a un residuo de citosina en la secuencia de nucleótidos de la muestra y citosina en la misma posición en la segunda secuencia de nucleótidos indican que el residuo de citosina en la secuencia de nucleótidos de la muestra es 5-hidroximetilcitosina (5hmC).
6. Un método de conformidad con cualquiera de las reivindicaciones 2 a 4 en donde la identificación de citosina en una posición tanto en la primera como en la segunda secuencia de nucleótidos que corresponde a un residuo de citosina en la secuencia de nucleótidos de la muestra, o la identificación de citosina en una posición en la primera secuencia de nucleótidos que corresponda a un residuo de citosina en la secuencia de nucleótidos de la muestra indican que el residuo de citosina en la secuencia de nucleótidos de la muestra es 5-metilcitosina (5mC).
7. Un método de conformidad con cualquiera de las reivindicaciones 2 a 6 que comprende;
 - vi) reducir la tercera porción de la población de polinucleótidos que comprende la secuencia de nucleótidos de la muestra,
 - vii) tratar la tercera porción reducida de la población con bisulfito,
 - viii) Secuenciar los polinucleótidos en la tercera porción que siguieron las etapas vi) y vii) para producir una tercera secuencia de nucleótidos, e;
 - ix) identificar el residuo en la tercera secuencia de nucleótidos que corresponda a un residuo de citosina en la secuencia de nucleótidos de la muestra.
8. Un método de conformidad con la reivindicación 7 en donde la identificación de uracilo en una posición en la primera y segunda secuencias de nucleótidos que corresponda a un residuo de citosina en la secuencia de nucleótidos de la muestra, y una citosina en esta posición en la tercera secuencia de nucleótidos indican que el residuo de citosina en la secuencia de nucleótidos de la muestra es 5-formilcitosina.
9. Un método de conformidad con la reivindicación 1 en donde la primera porción de dicha población se reduce en la etapa ii).
10. Un método de conformidad con la reivindicación 9 en donde la identificación de citosina en una posición en la primera secuencia de nucleótidos que corresponda a una citosina en la secuencia de nucleótidos de la muestra y uracilo en esa misma posición en la segunda secuencia de nucleótidos indican que el residuo de citosina en la secuencia de nucleótidos de la muestra es 5-formilcitosina; y, en donde la identificación de uracilo en una posición tanto en la primera como en la segunda secuencia de nucleótidos que corresponda a citosina en la secuencia de nucleótidos de la muestra indica que el residuo de citosina en la secuencia de nucleótidos de la muestra es una citosina no modificada.
11. Un método de conformidad con cualquiera de las reivindicaciones 1 a 10 que comprende; proporcionar una cuarta porción de la población de nucleótidos que comprende la secuencia de nucleótidos de la muestra; y, secuenciar los polinucleótidos en la cuarta porción para producir la secuencia de nucleótidos de la muestra.
12. Un método de identificación de 5-metilcitosina en una secuencia de nucleótidos de la muestra que comprende;

- 5
- 10
- i) proporcionar una población de polinucleótidos que comprende la secuencia de nucleótidos de la muestra,
 - ii) oxidar dicha población,
 - iii) Tratar dicha población oxidada con bisulfito,
 - iv) secuenciar los polinucleótidos en la población que siguieron las etapas ii) y iii) para producir una secuencia de nucleótidos tratada, e;
 - v) identificar los residuos en la secuencia de nucleótidos tratada que correspondan a residuos de citosina en la secuencia de nucleótidos de la muestra, en donde la presencia de una citosina en la secuencia de nucleótidos tratada indica que el residuo de citosina en la secuencia de nucleótidos de la muestra es 5-metilcitosina (5mC).
13. Un método de conformidad con cualquiera de las reivindicaciones 1 a 12 en donde una o más de la primera, segunda, tercera y cuarta porciones de la población se amplifican antes de la secuenciación.

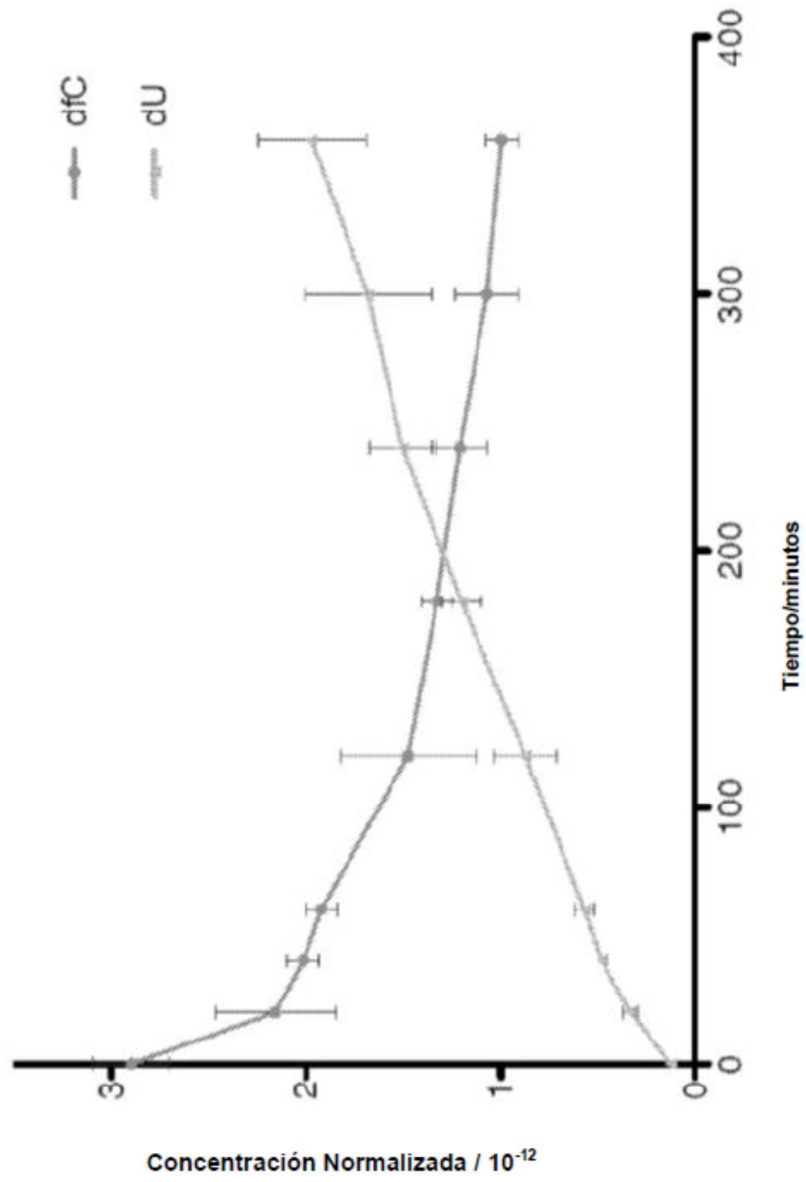


Figura 1 A

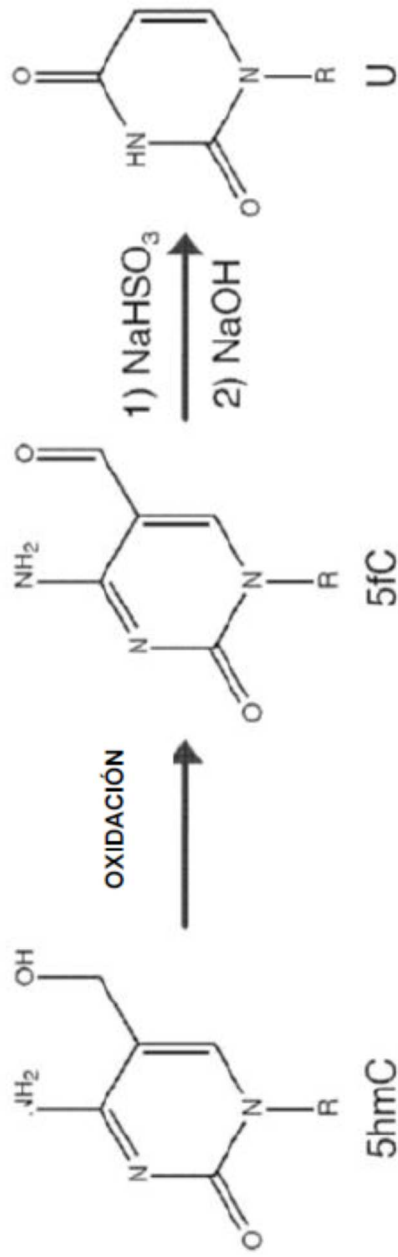


Figura 1 B

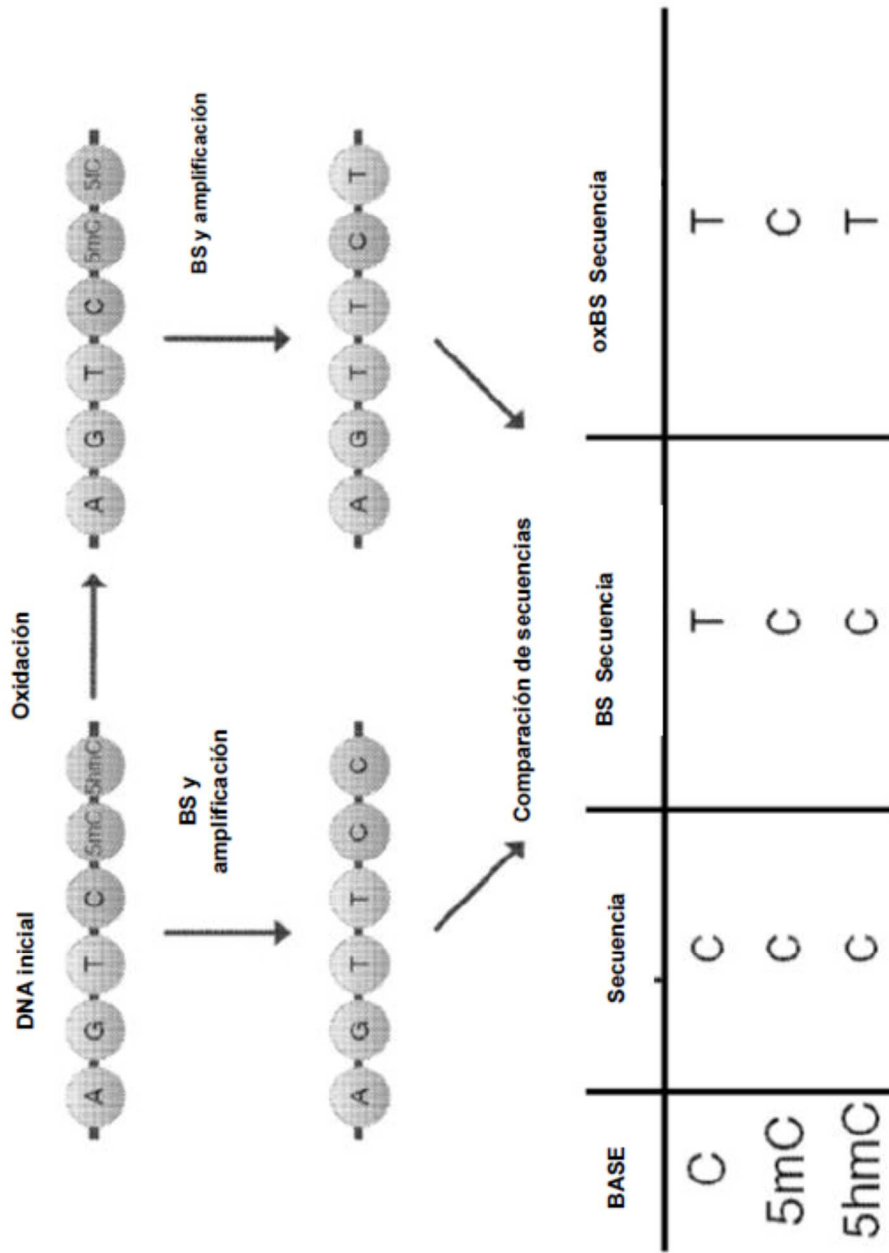


Figura 1 C

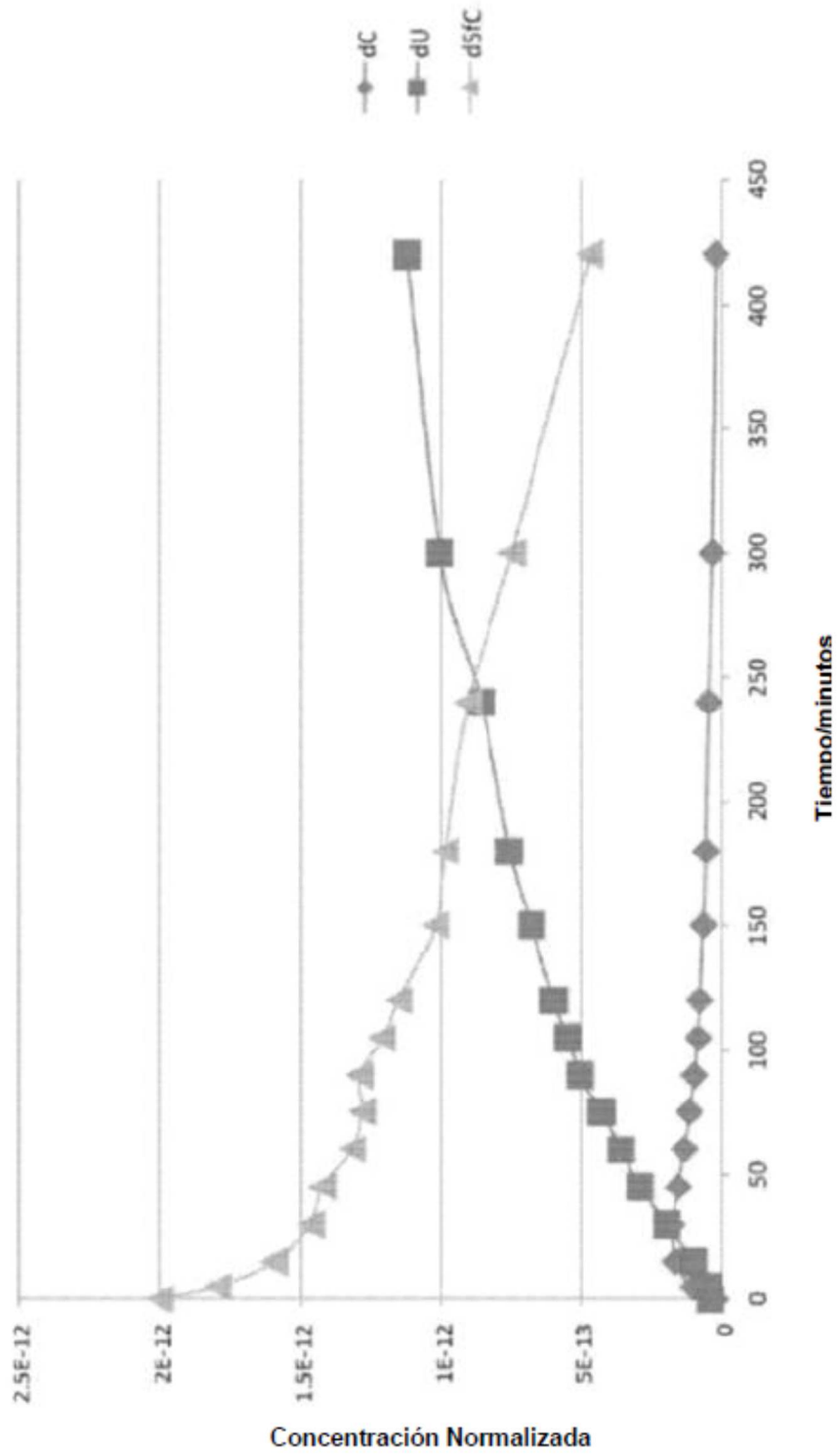


Figura 2 A

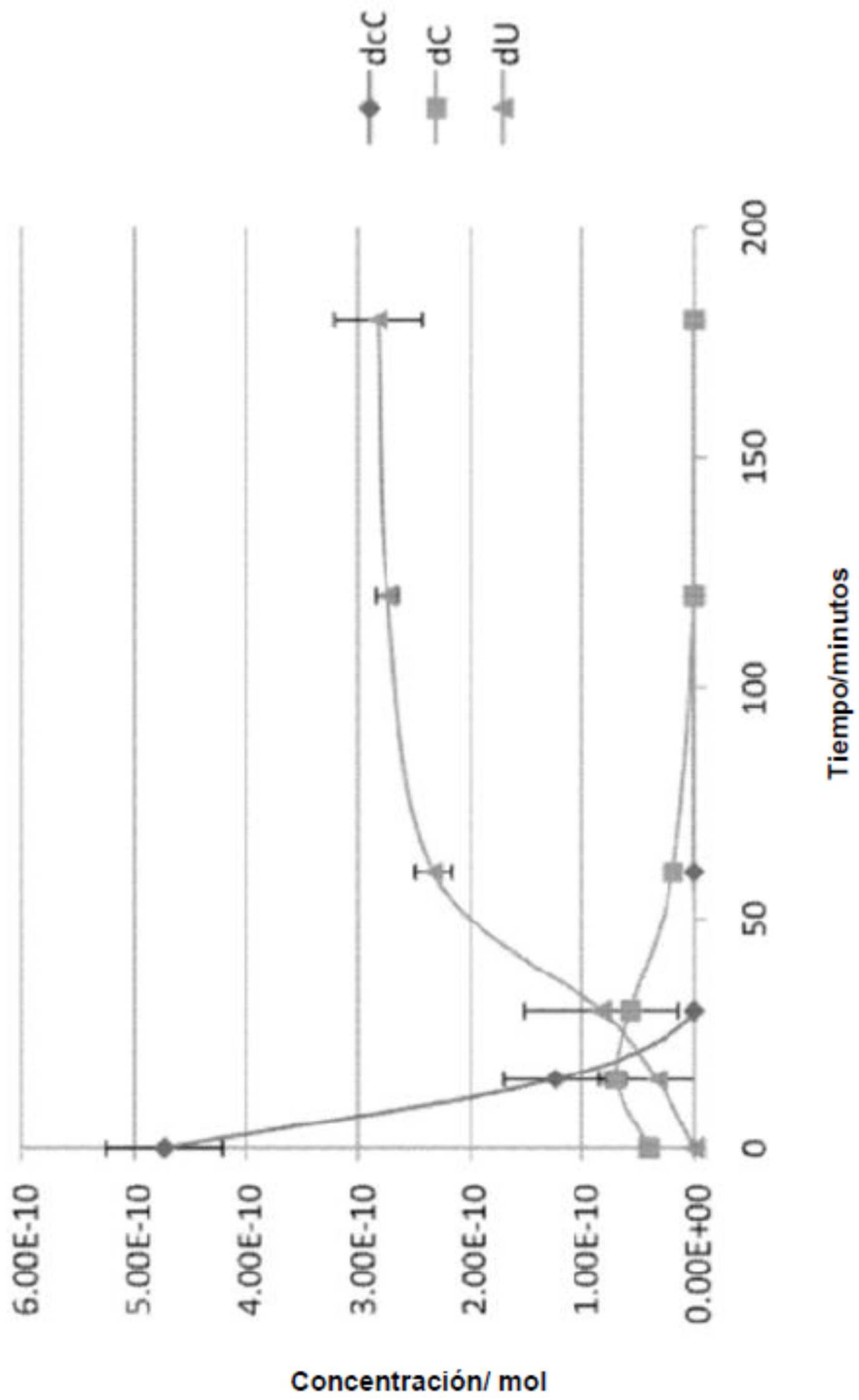


Figura 2 B

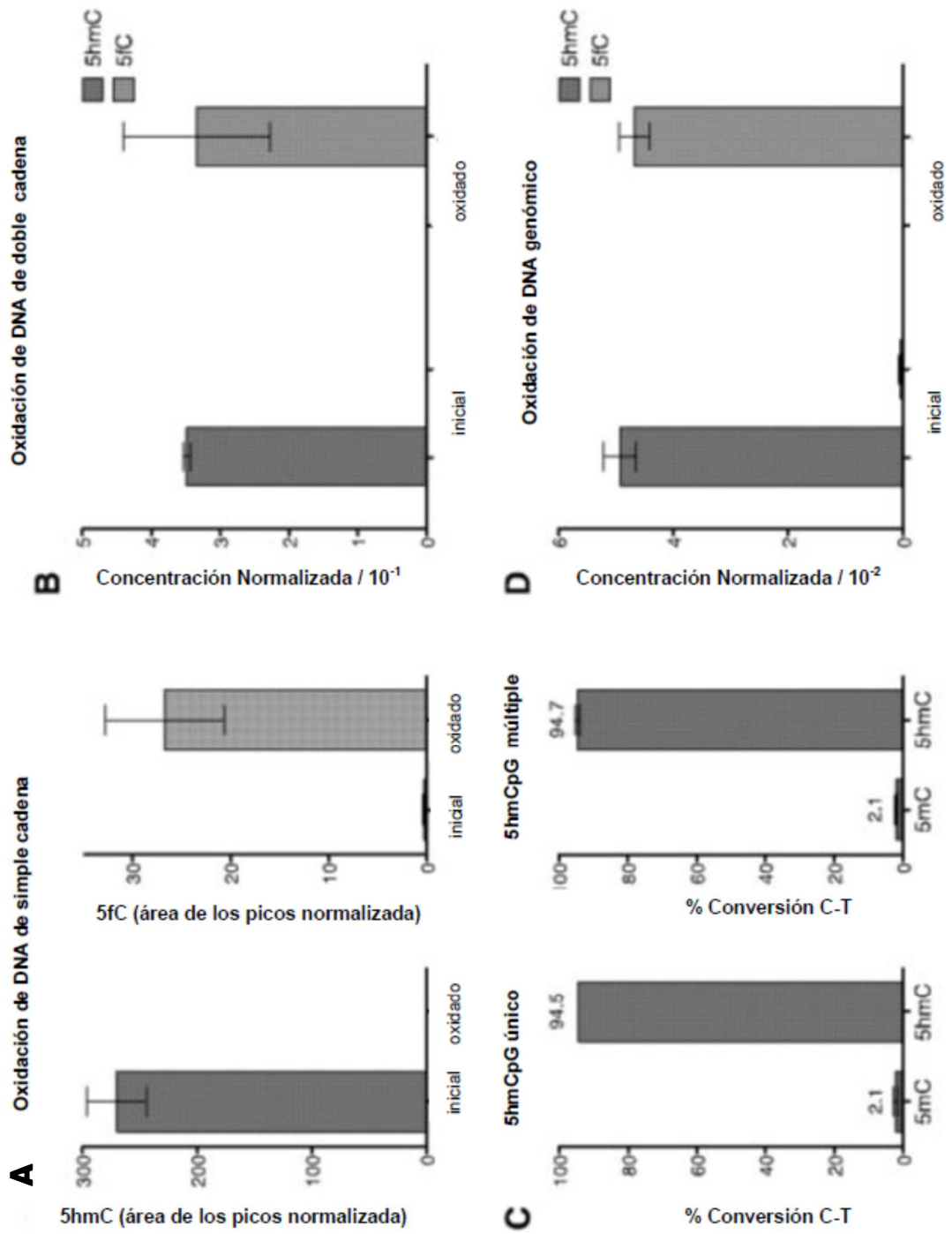


Figura 3

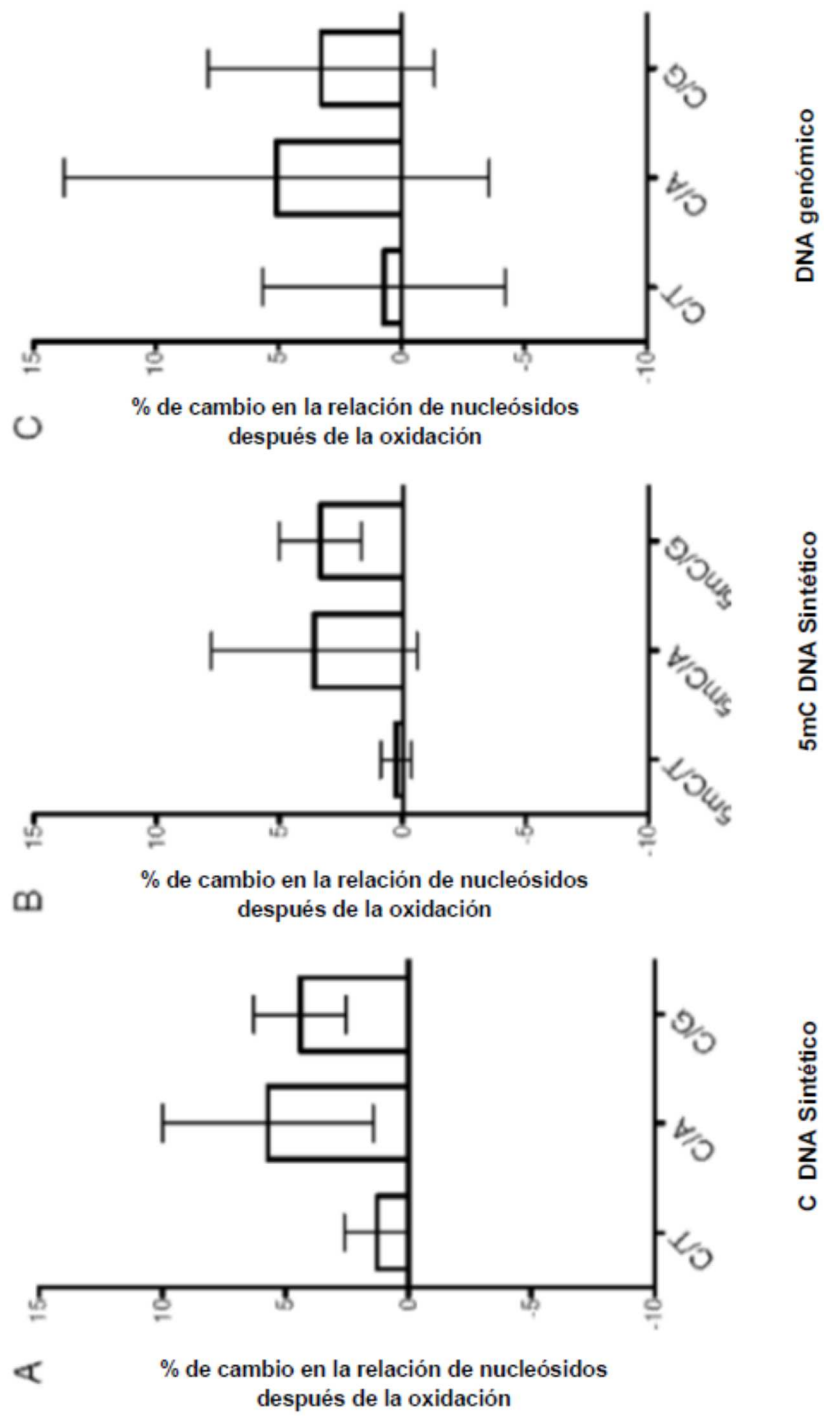


Figura 4

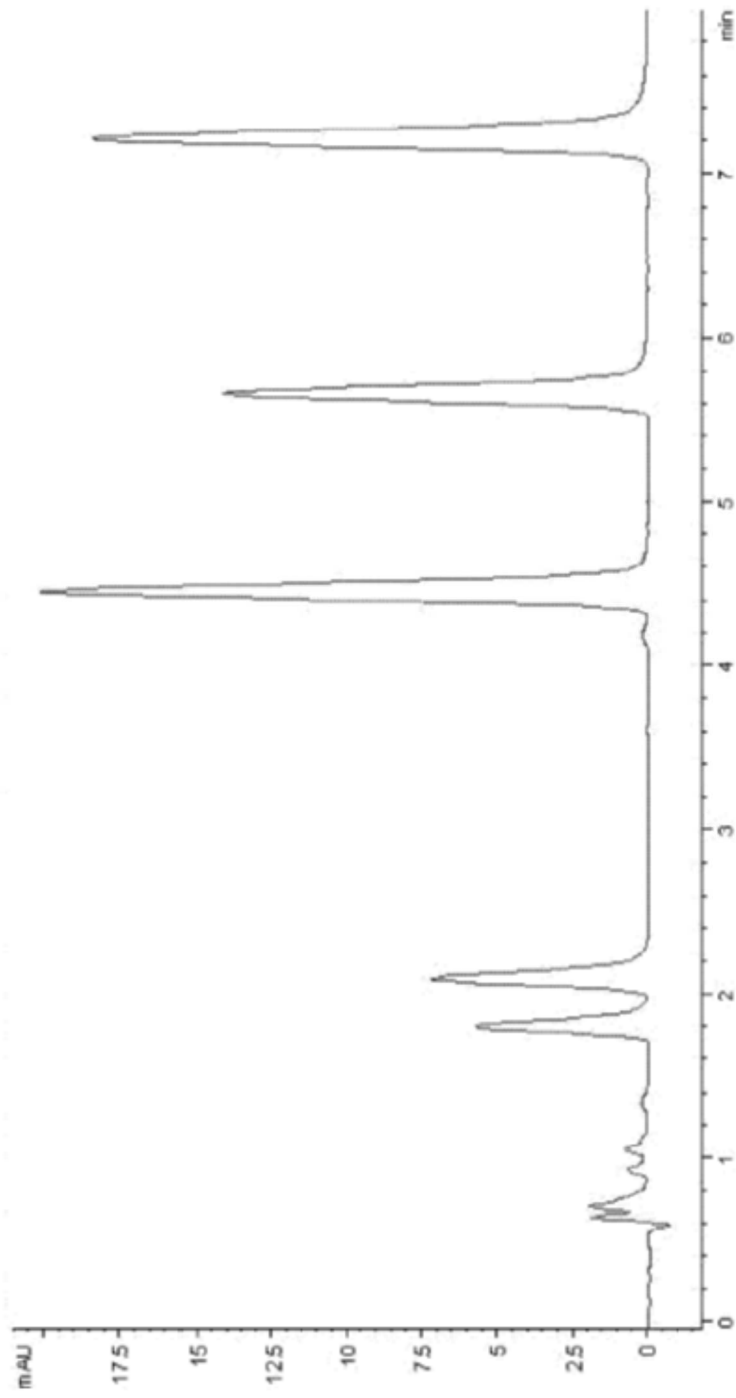


Figure 5 A

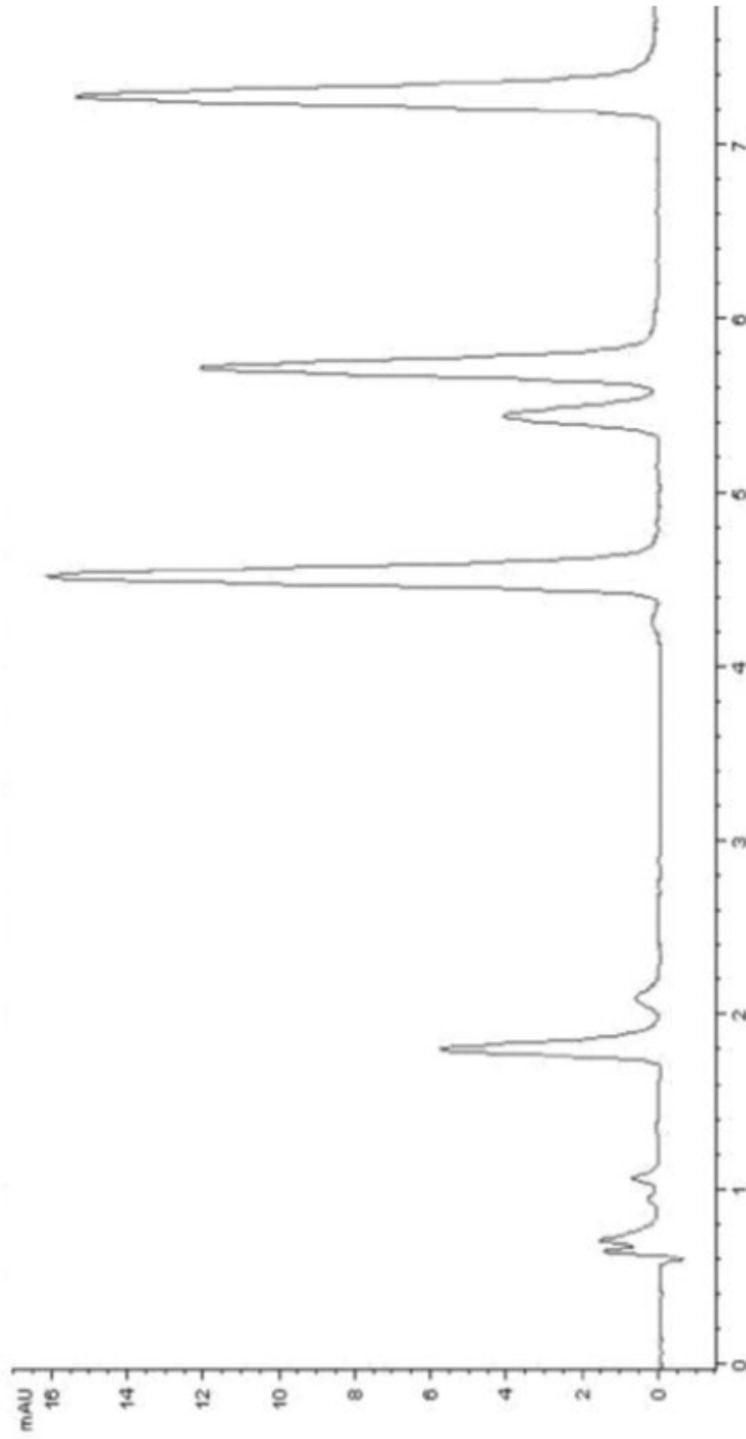


Figure 5 B

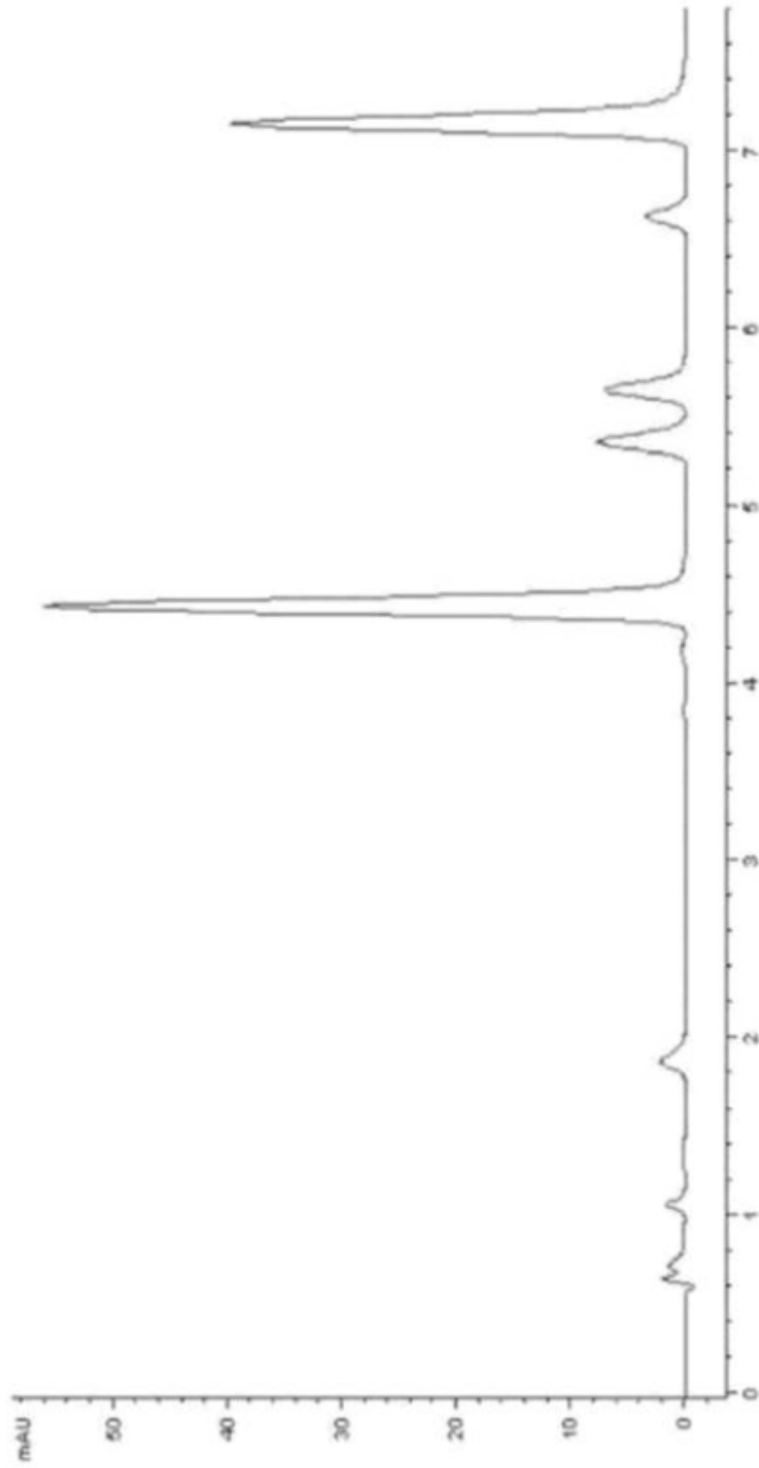


Figura 6 A

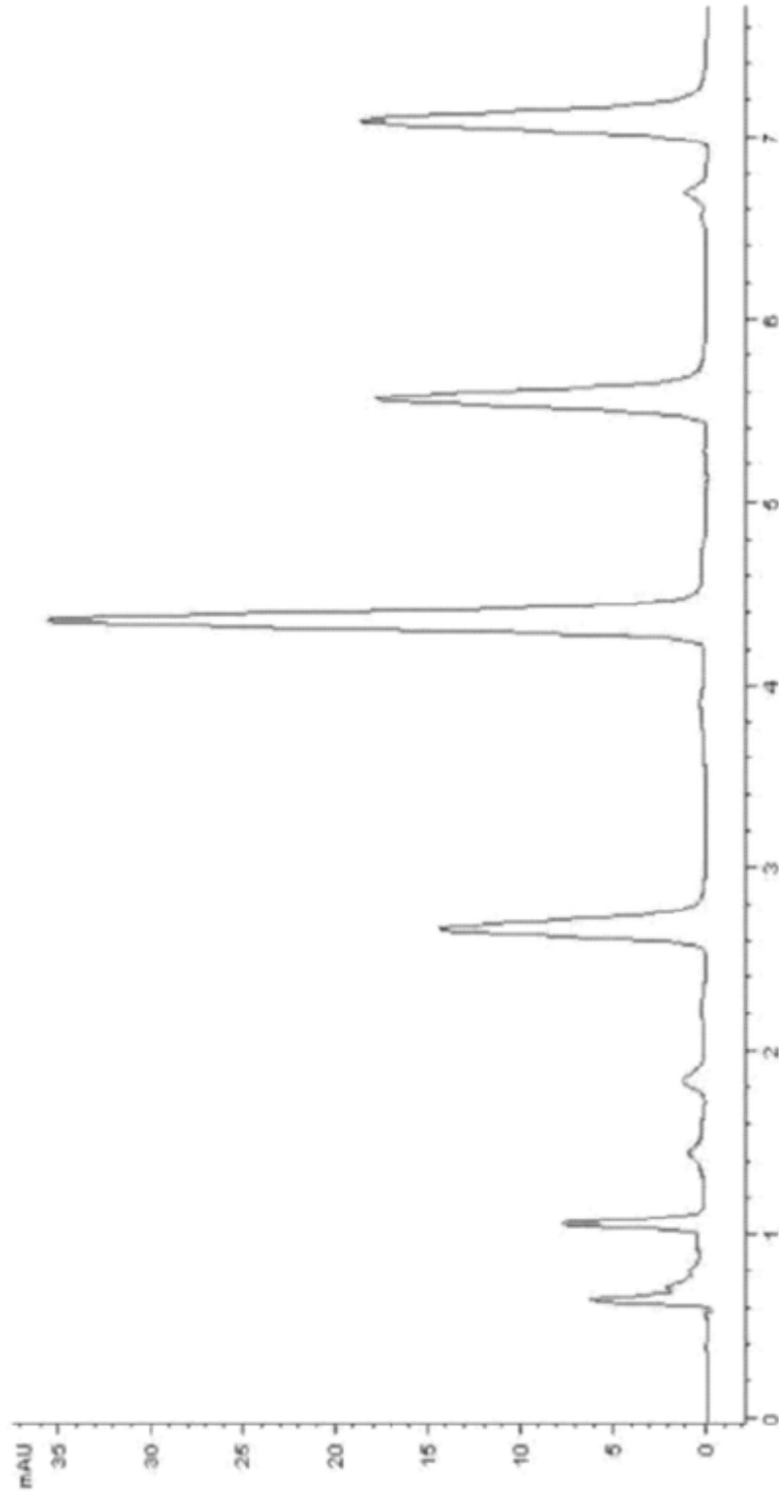


Figura 6 B

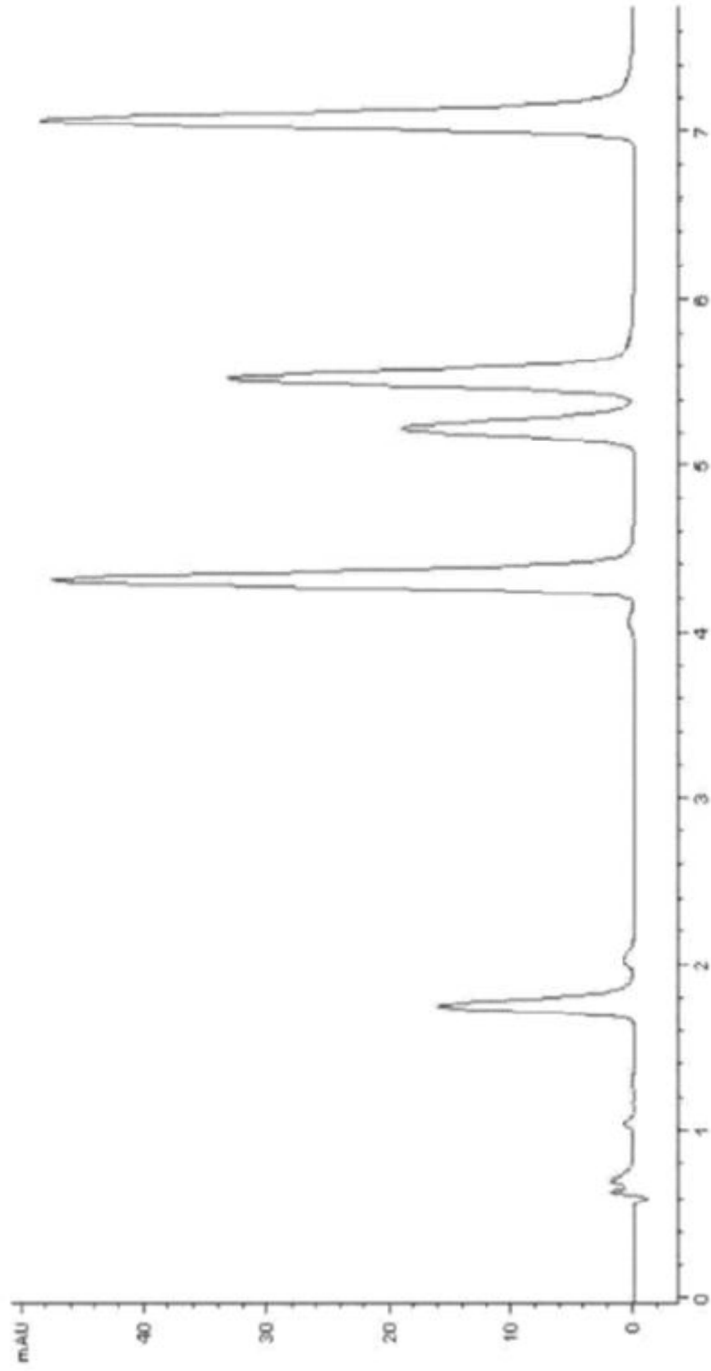


Figura 7 A

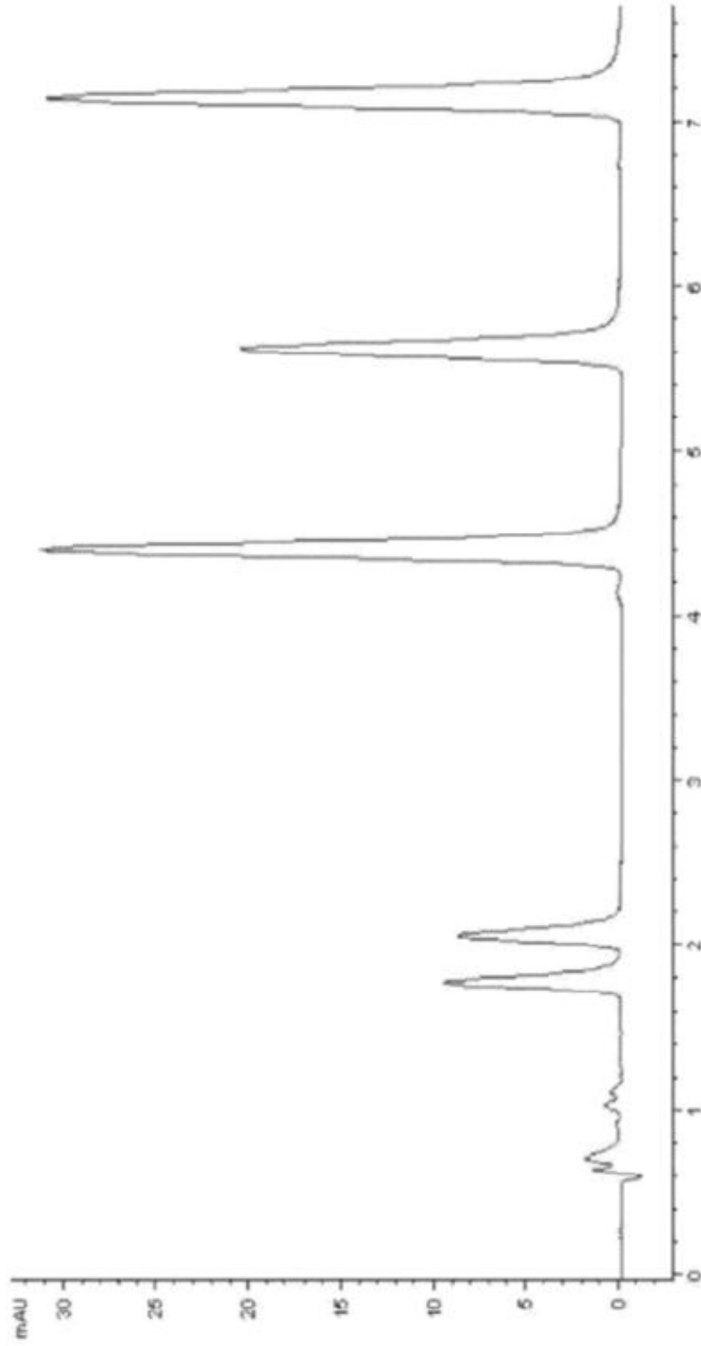


Figura 7 B

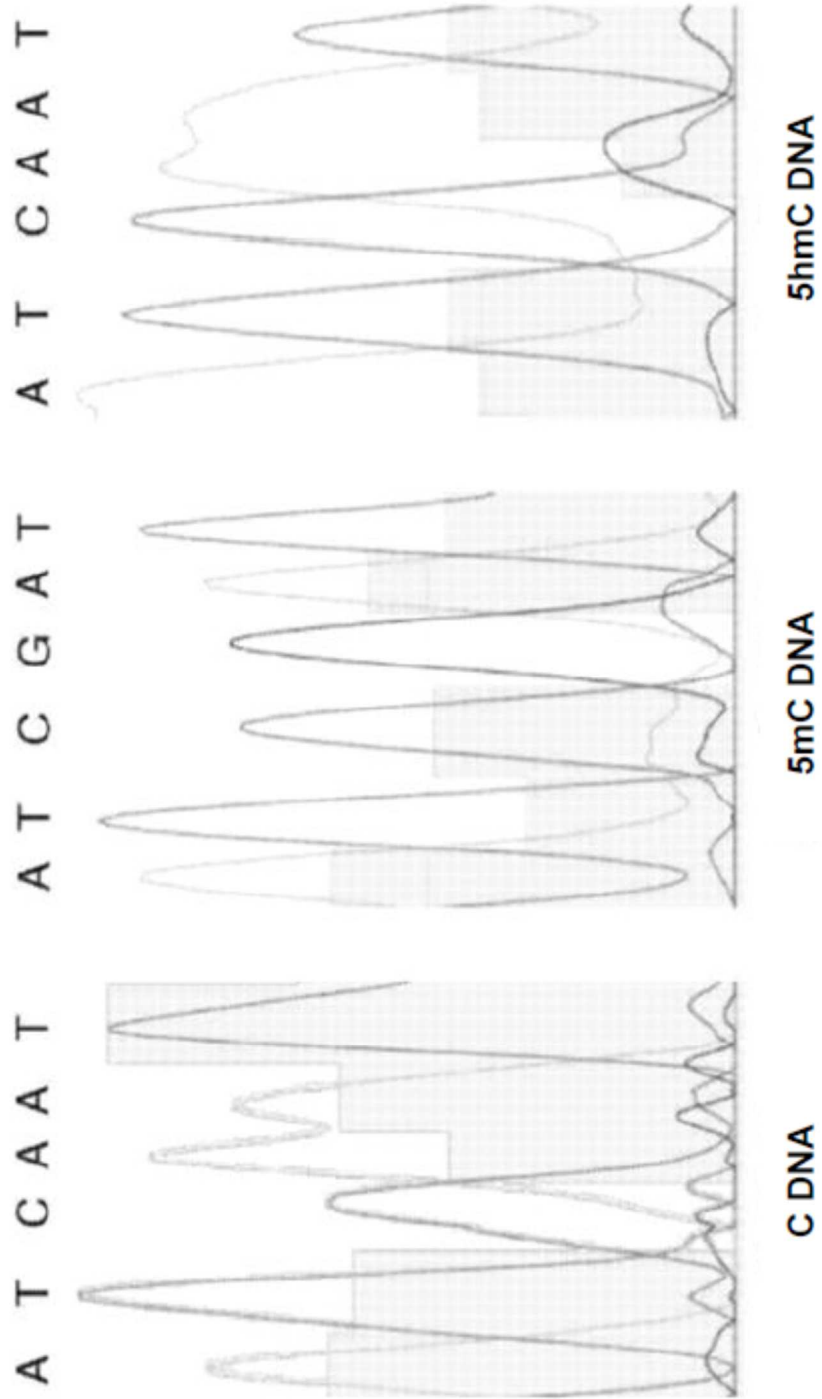


Figura 8

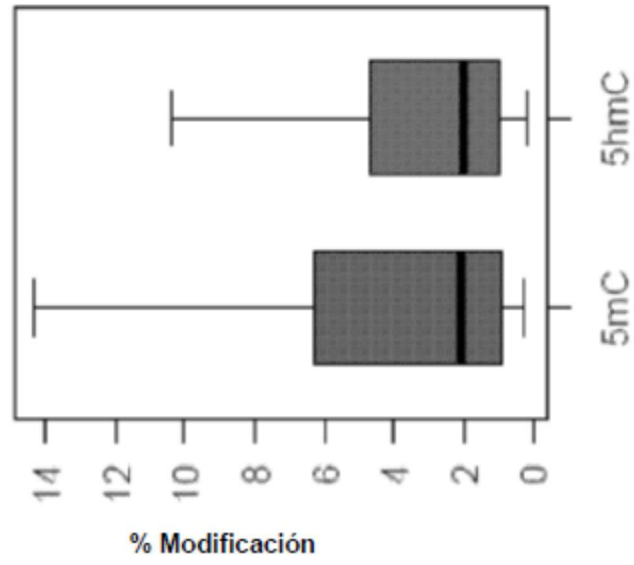


Figura 9 B

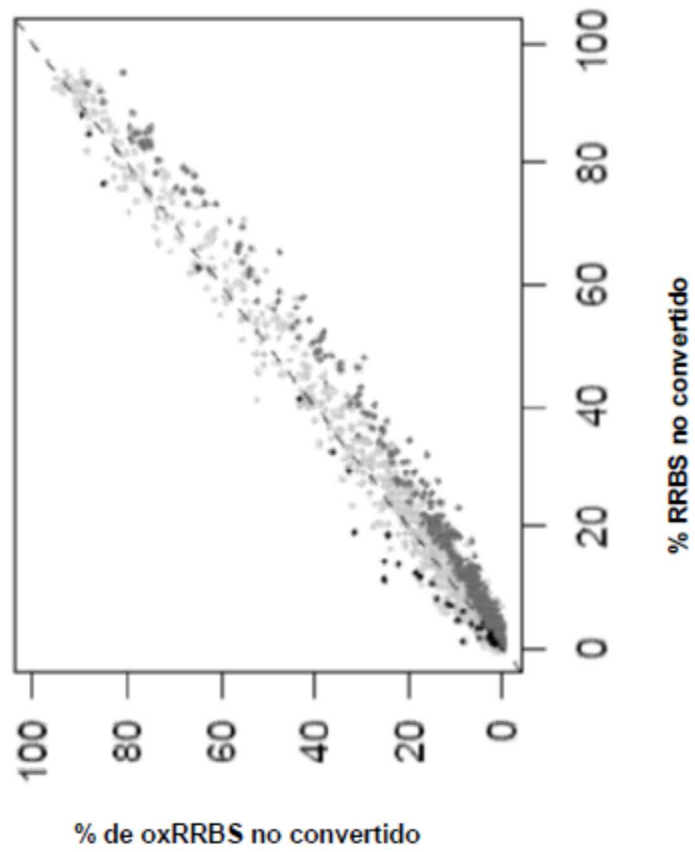


Figura 9 A

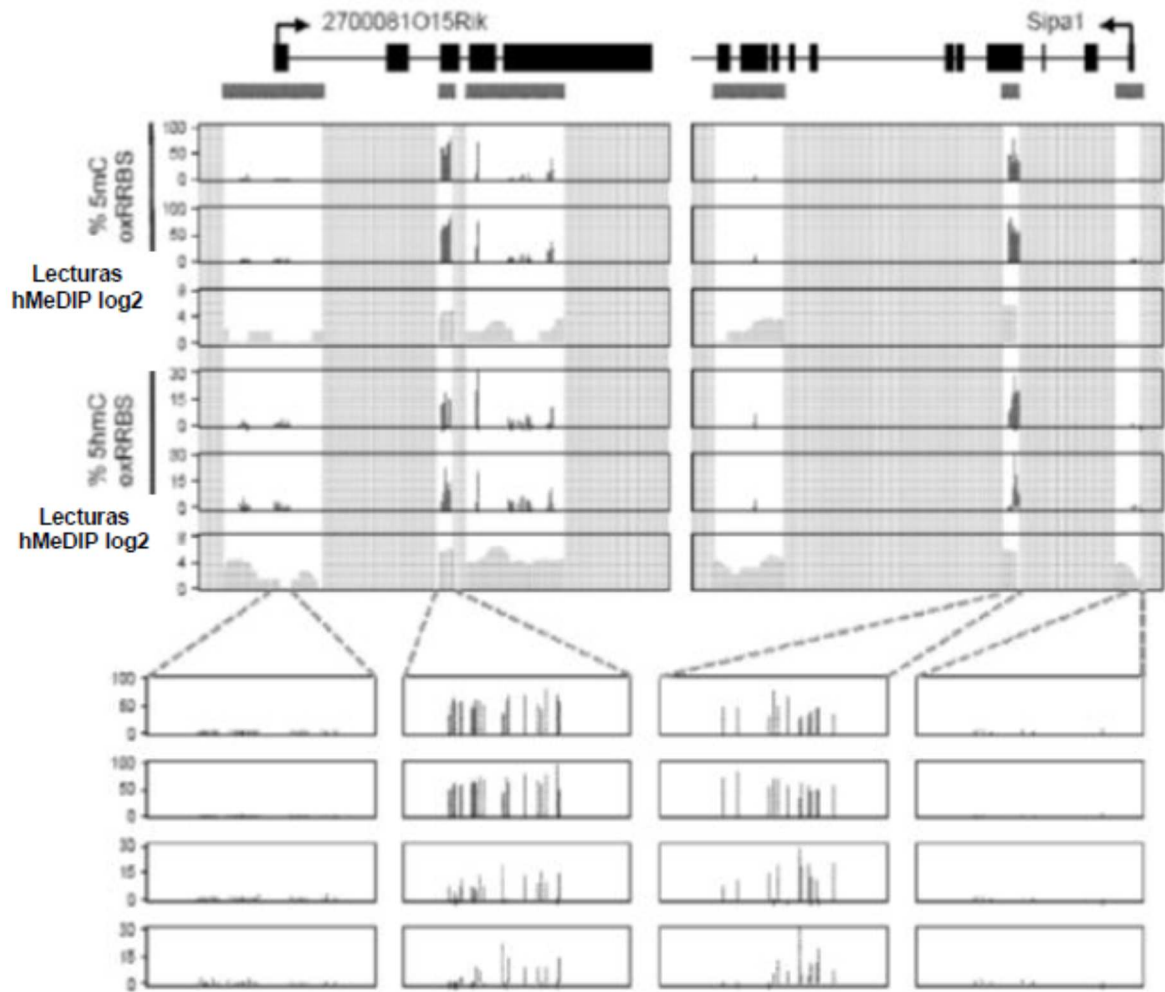


Figura 9 C

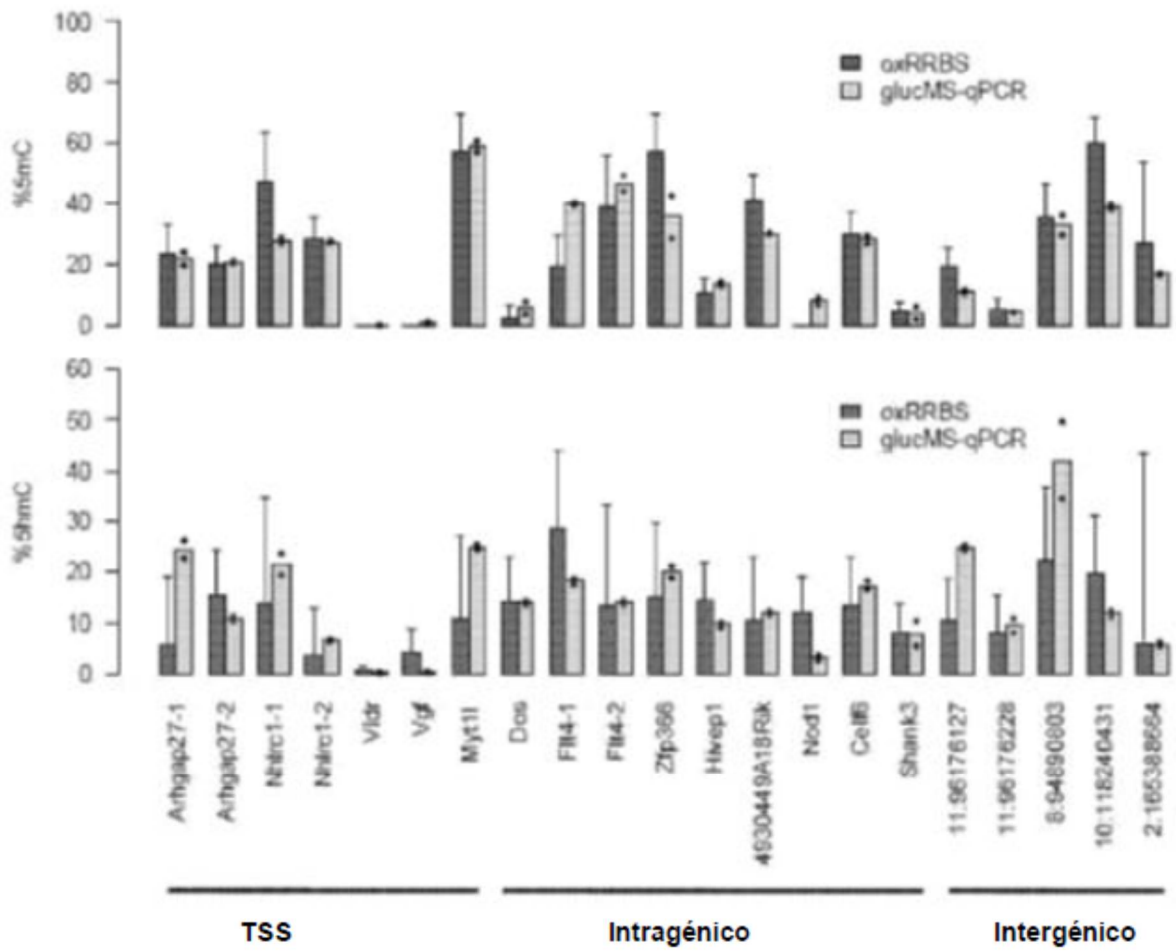


Figura 9 D

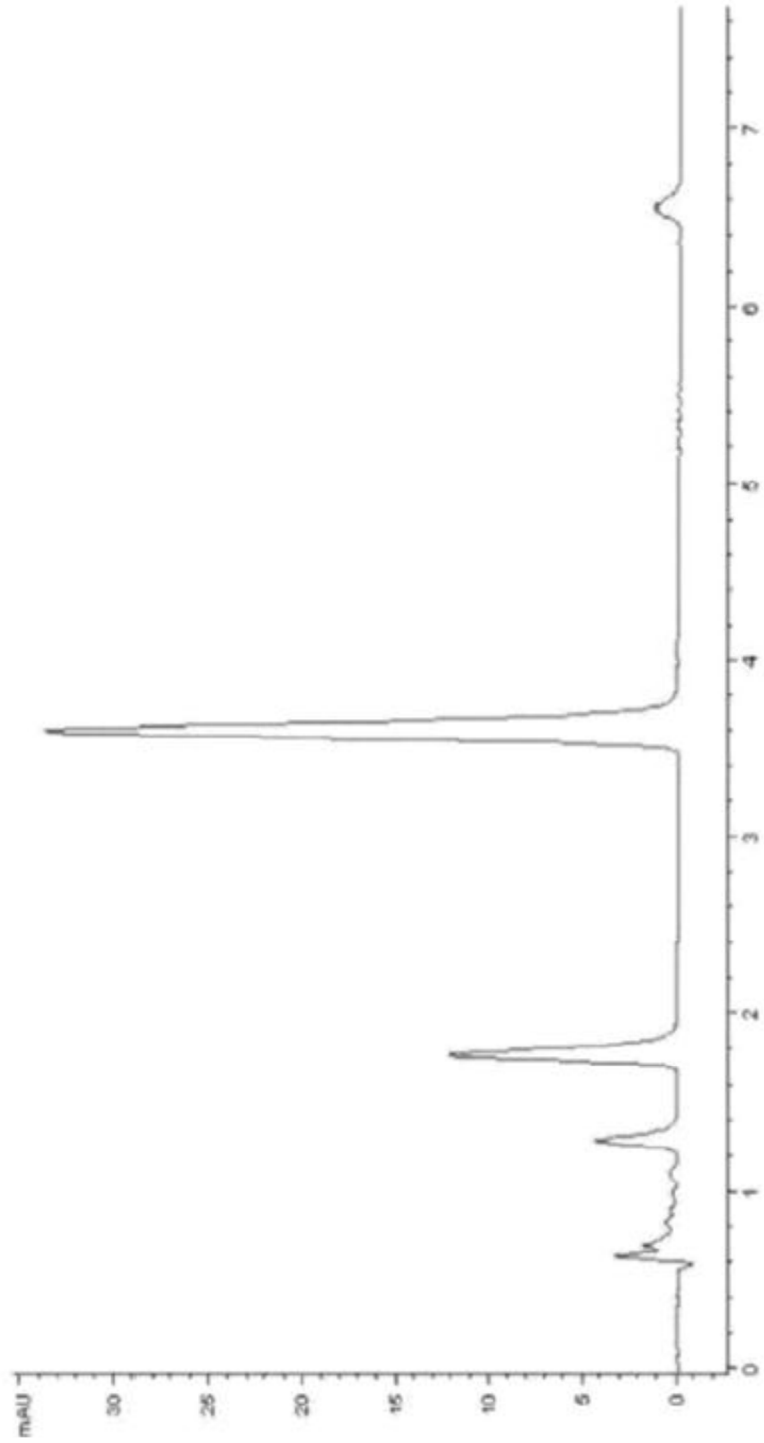


Figura 10 A

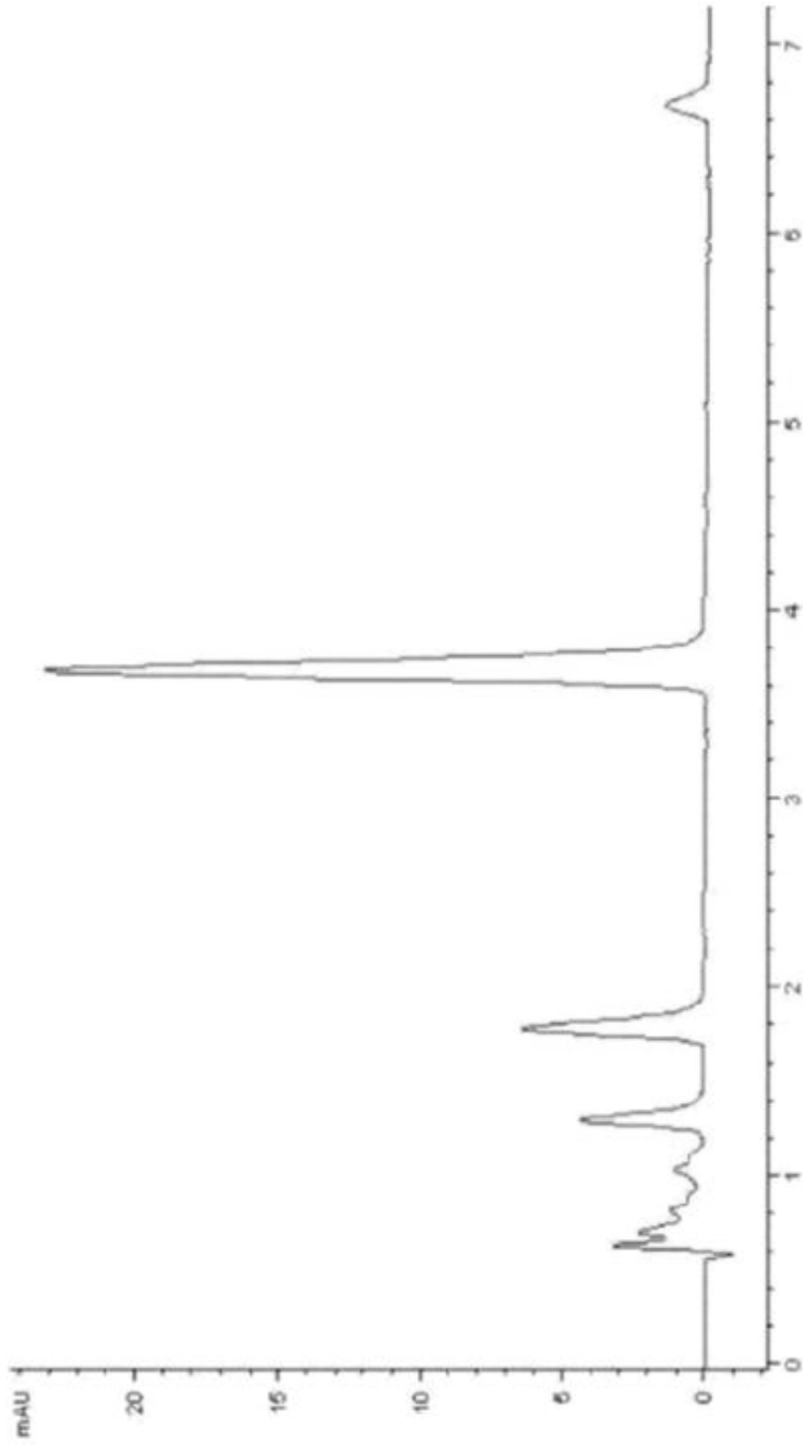


Figure 10 B

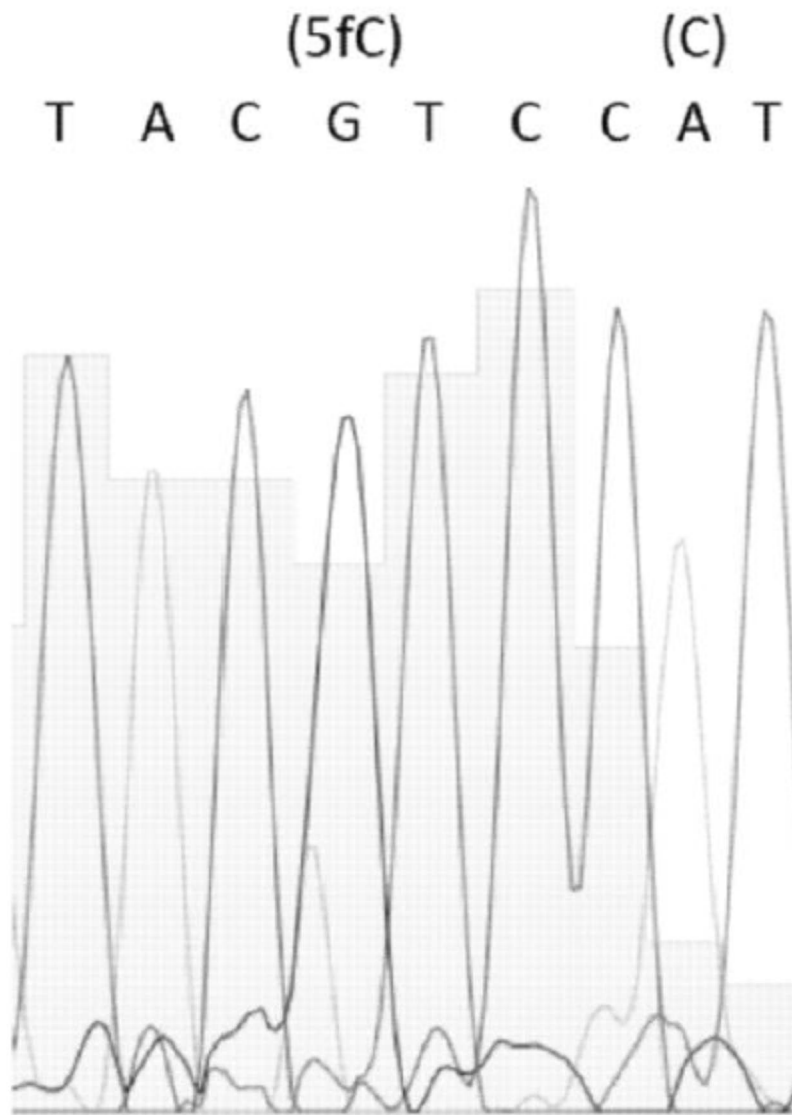


Figura 11