

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **3 009 684**

51 Int. Cl.:

**C12Q 1/68**

(2008.01)

12

## TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **25.10.2019** **PCT/CN2019/113387**

87 Fecha y número de publicación internacional: **29.04.2021** **WO21077415**

96 Fecha de presentación y número de la solicitud europea: **25.10.2019** **E 19949899 (9)**

97 Fecha y número de publicación de la concesión europea: **11.12.2024** **EP 4048808**

54 Título: **Detección y análisis de metilación del ADN de mamíferos**

45 Fecha de publicación y mención en BOPI de la  
traducción de la patente:  
**31.03.2025**

73 Titular/es:

**CHANGPING NATIONAL LABORATORY**  
(100.00%)  
**No. 7 Shengmingkexueyuan Road, Changping**  
**District**  
**Beijing 102206, CN**

72 Inventor/es:

**CAO, YUNLONG y**  
**XIE, XIAOLIANG**

74 Agente/Representante:

**DEL VALLE VALIENTE, Sonia**

### Observaciones:

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

ES 3 009 684 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín Europeo de Patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre Concesión de Patentes Europeas).

## DESCRIPCIÓN

Detección y análisis de metilación del ADN de mamíferos

## 5 Antecedentes

### Campo de la invención

10 Las realizaciones de la presente invención se refieren generalmente a métodos y composiciones para la secuenciación de ADN en cantidades bajas, tales como ADN de unas pocas células a 100 células o una única célula.

### Descripción de la técnica relacionada

15 La capacidad de realizar estudios de metilación del ADN de mamíferos de alta cobertura y alta precisión en ADN a nivel de una única célula es esencial en los estudios donde la variación de célula a célula y la heterogeneidad de la población desempeñan un papel clave, como el crecimiento tumoral, la reprogramación de las células madre, la formación de la memoria, el desarrollo embrionario, etc. Esto también es importante cuando las muestras de células sujetas a análisis son preciadas o raras o en cantidades mínimas, como cuando la muestra es una única célula o el genoma, en su totalidad o en parte, de una única célula o ADN libre de células.

20 Para analizar la metilación del ADN con una resolución de una única base, el ADN debe pasar por una conversión química o enzimática para distinguir eficazmente la citosina metilada y la citosina no metilada. Los métodos de conversión comunes incluyen, pero no se limitan a, el bisulfito de sodio, la secuenciación de piridina-borano asistida por TET (TAPS, por sus siglas en inglés), Liu y col., *Nature Biotechnology* (2019). "Bisulfite-free direct detection of 5-methylcytosine and 5-hydroxymethylcytosine at base resolution", Conversión de secuencia enzimática-metilica (EM-seq, por sus siglas en inglés), Williams y col., *New England Biolabs, Inc.* (2019). "Enzymatic Methyl-seq: The Next Generation of Methylome Analysis" [world wide website international.neb.com/tools-and-resources/feature-articles/enzymatic-methyl-seq-the-next-generation-of-methylome-analysis].

30 El principal desafío en toda conversión química o enzimática es la pérdida o el daño del ADN que tiene lugar simultáneamente con la conversión. Las condiciones necesarias para completar las conversiones, como los largos tiempos de incubación, la temperatura elevada, la alta concentración de bisulfito y un entorno altamente oxidativo o reductor, pueden provocar la degradación, el daño de las bases del ADN y la fragmentación de hasta el 90 % del ADN incubado. Las purificaciones del ADN entre las etapas de la conversión o después de estas también pueden provocar una pérdida de ADN de hasta un 90 %. La pérdida y el daño extensos del ADN son problemáticos y más aún cuando se trata de una cantidad limitada o baja de ADN inicial o incluso de ADN a nivel de una única célula. La secuenciación de bisulfito unicelular de baja cobertura se ha logrado realizando directamente la conversión de bisulfito en una única célula, seguida de la amplificación del ADN. Guo, H., y col. (2013). "Single-cell methylome landscapes of mouse embryonic stem cells and early embryos analyzed using reduced representation bisulfite sequencing". *Genoma Res* 23(12): 2126-2135; Smallwood, S. A. y col. (2014). "Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity." *Nat Methods* 11(8): 817-820.

45 El daño del ADN durante la conversión química o enzimática, así como la pérdida de ADN durante las purificaciones del ADN que se producen simultáneamente con las conversiones, podrían reducirse significativamente mediante la adición de ADN portador, como el ADN lambda sonificado de tamaño 100 pb a 4 kb. Sin embargo, la mezcla del ADN portador con el ADN dirigido da como resultado una mezcla de muestras que no se pudieron distinguir más adelante, lo que provocó dificultades o incluso fallas en la detección y el análisis de la metilación posteriores.

50 La transposición in vitro se ha utilizado en ciertas aplicaciones de amplificación de ADN. En tales métodos, el ADN diana se fragmenta y se etiqueta simultáneamente produciendo fragmentos etiquetados con las secuencias de ADN deseadas para su procesamiento posterior. Como método de preparación de genotecas, la transposición in vitro se ha utilizado en la tecnología Nextera de Illumina, Inc., para fragmentar simultáneamente el ADN y etiquetar cada fragmento con las secuencias apropiadas para la secuenciación de próxima generación (US20110287435). Como herramienta para estudiar genomas y epigenomas unicelulares, Buenrostro y col. han utilizado la transposición in vitro para perfilar la accesibilidad a la cromatina (Buenrostro, J. D., Wu, B., Litzenburger, U. M., Ruff, D., Gonzales, M. L., Snyder, M. P., ... y Greenleaf, W.J. (2015). Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature*, 523(7561), 486-490), de Ramani y col. para estudiar la conformación cromosómica tridimensional (Ramani, V., Deng, X., Qiu, R., Gunderson, K. L., Steemers, F. J., Distech, C. M., ... y Shendure, J. (2017). Massively multiplex single-cell Hi-C. *Nature Methods*, 14(3), 263-266), y de Zahn y col. para amplificar genomas unicelulares directamente en la genoteca de secuenciación (Zahn, H., Steif, A., Laks, E., Eirew, P., VanInsberghe, M., Shah, S. P., ... y Hansen, C. L. (2017). Scalable whole-genome single-cell library preparation without preamplification. *Nature Methods*, 2017). Sin embargo, todos estos métodos sufren una pérdida de aproximadamente el 50 % del ácido nucleico diana original. Esto sucede porque se utilizan dos secuencias de transposones para el etiquetado, denominadas en lo sucesivo A y B: Después de etiquetar los transposones A y B en el ADN diana, se pueden generar cuatro tipos diferentes de fragmentos de ADN, que son fragmentos marcados con A-A, B-B, A-B o B-A en los dos extremos de cada fragmento. Solo los fragmentos etiquetados con A-B o B-A, que representan el 50 % del total de los productos de transposición,

son adecuados para la amplificación por PCR o la secuenciación de extremos emparejados. El otro 50 % de los fragmentos, que están etiquetados con A-A o B-B, se perderán. Tal tasa de pérdida es ciertamente indeseable, y potencialmente inaceptable, para muestras con una cantidad limitada de ADN, incluidas muestras unicelulares raras, únicas o valiosas, como una única célula que se utilizará para el cribado genético preimplantacional. En la patente WO2016/073690 se describe un método de transposición adicional, sin embargo, tal método no reduce la pérdida del 50 % resultante del sesgo de transposición.

En consecuencia, existe la necesidad de métodos adicionales para analizar el estado de metilación de cantidades bajas de ADN de mamíferos sin incurrir en la pérdida de ADN asociada con los métodos de la técnica anterior. La patente WO 2018/165366 A1 se refiere a un método para fabricar un metiloma amplificado extendiendo fragmentos y tratando los fragmentos extendidos con una metiltransferasa y una fuente de grupos metilo para transformar el ADN bicatenario hemimetilado en ADN bicatenario completamente metilado. La patente WO 2018/217912 se refiere a un método para el ensamblaje de ADN genómico mediante la amplificación por etiquetado terminal múltiple de fragmentos genómicos.

## Resumen

La presente descripción proporciona un método para el análisis del estado de metilación de una muestra de ADN diana presente en una cantidad baja, tal como el ADN de una pluralidad de células o una única célula, incluida la fragmentación del ADN genómico utilizando una pluralidad de transposomas donde cada miembro de la pluralidad de transposomas incluye dos secuencias de ácido nucleico de transposones que tienen secuencias de sitios de cebado. Las secuencias de ácido nucleico de transposón que tienen secuencias en el sitio de cebado que se unen a la transposasa pueden estar metiladas o no metiladas para toda la citosina, dependiendo de si la química de conversión convierte la citosina metilada o la citosina no metilada. El fragmento de ADN producido por los transposomas puede tener una citosina y/o una metilcitosina presente en el mismo.

Los métodos para fragmentar ADN de bajo volumen utilizando transposomas, rellenar espacios, purificar, amplificar y secuenciar se proporcionan en la patente PCT/US2018/034162, incorporado aquí como referencia en su totalidad, y generalmente se denominan amplificación por etiquetado de múltiples extremos o "META". Según un aspecto, la secuencia del sitio de cebado de cada secuencia de ácido nucleico del transposón del transposoma es la misma. Según un aspecto, la secuencia del sitio de cebado de cada secuencia de ácido nucleico del transposón del transposoma es diferente. Según un aspecto, cada miembro de la pluralidad de transposomas puede incluir una secuencia de sitio de cebado única y/o diferente. Según un aspecto, cada miembro de la pluralidad de transposomas puede incluir dos secuencias de sitios de cebado únicas y/o diferentes, una para cada transposón en el transposoma. De esta manera, se proporciona un conjunto de transposomas que tienen una secuencia de sitio de unión al cebador única (o dos secuencias de sitio de cebado únicas y/o diferentes) asociada a la misma y que pueden utilizarse para distinguir los transposomas. Dicho de otro modo, las secuencias del sitio de unión al cebador de los transposones dentro del transposoma pueden ser las mismas o pueden ser diferentes o no idénticas. Las secuencias del sitio de unión al cebador de los transposomas en dos transposomas adyacentes unidos a una secuencia de ácido nucleico diana y utilizadas para fabricar un fragmento no son idénticas, por ejemplo con una alta probabilidad. Los transposones pueden denominarse transposones múltiples en la medida en que cada transposón dentro de un transposoma tenga una secuencia de sitios de cebado diferente. Los sitios de cebado dentro de una genoteca de transposomas pueden denominarse sitios de cebado múltiple en la medida en que cada transposoma tenga un sitio de cebado que sea diferente o no idéntico o único de otros sitios de cebado dentro de otros transposomas dentro del conjunto de transposomas. Según un aspecto, el método proporciona la etapa de unir transposomas de una genoteca o pluralidad de transposomas a lo largo de una secuencia de ácido nucleico diana de tal modo que los transposomas adyacentes tengan diferentes secuencias de sitios de unión al cebador. De esta manera, los extremos del sitio de fragmentación se etiquetarán con diferentes secuencias de sitios de unión al cebador. Esto se puede lograr si un transposoma tiene la misma secuencia de sitios de unión al cebador para cada uno de sus dos ADN de transposón o si un transposoma tiene una secuencia de sitios de unión al cebador diferente para cada uno de sus dos ADN de transposón. De esta manera, el método de amplificación por etiquetado múltiple de extremos descrito en la presente memoria utiliza múltiples secuencias de cebado para crear fragmentos de ADN diana marcados por diferentes secuencias en los dos extremos. El método de amplificación por etiquetado múltiple de extremos puede llevarse a cabo independientemente de que las dos secuencias de transposones dentro de un transposoma sean iguales o diferentes, siempre que dos transposomas adyacentes, es decir, directamente adyacentes para formar una secuencia de fragmentos, porten diferentes secuencias de sitios de unión a cebadores de transposones donde el fragmento tenga diferentes secuencias de sitios de unión a cebadores en cada extremo.

Según un aspecto, el uso de secuencias de sitios de cebado múltiples dentro de un conjunto de transposomas reduce la tasa de pérdida cuando se utiliza un método de transposición para fragmentar y etiquetar una secuencia de ácido nucleico genómico, tal como una secuencia de ácido nucleico genómico de una única célula. Según las enseñanzas de la presente memoria, cuando hay N secuencias de transposones diferentes en la mezcla de reacción, es decir, cuando el número de secuencias únicas del sitio de cebado es N, la probabilidad de que un fragmento de ADN esté etiquetado por la misma secuencia de transposones, es decir, la tasa de pérdida, es 1/N. Por lo tanto, la presente descripción proporciona un método para alterar el número de secuencias únicas del sitio de cebado, es decir, el

número N, para controlar la tasa de pérdida. Por ejemplo, cuando hay 20 secuencias de transposones diferentes, para su uso con el ADN obtenido de una única célula humana, la tasa de pérdida es de 1/20 o 5 %.

El método descrito en la presente memoria para crear una pluralidad de fragmentos utiliza un conjunto de transposomas en donde cada miembro del conjunto de transposomas tiene una o dos secuencias de sitios de unión a cebadores diferentes y en donde cada miembro del conjunto de transposomas tiene uno o dos sitios de unión de cebado únicos o diferentes en comparación con cada otro miembro del conjunto de transposomas, tal como con una alta probabilidad. De esta manera, los extremos contiguos de los fragmentos reciben códigos de barras con secuencias de códigos de barras finales diferentes y/o únicas durante el proceso de fragmentación para crear fragmentos que tienen secuencias de códigos de barras únicas (secuencias de sitios de cebado) en cada extremo. De esta manera, los extremos opuestos de los fragmentos reciben códigos de barras con secuencias de códigos de barras finales diferentes y/o únicas con una alta probabilidad durante el proceso de fragmentación de crear fragmentos que tengan diferentes secuencias de códigos de barras (secuencias de sitios de cebado) en cada extremo. De esta manera, los dos extremos opuestos de los fragmentos reciben códigos de barras con secuencias de códigos de barras finales diferentes y/o únicas durante el proceso de fragmentación para crear fragmentos que tienen secuencias de códigos de barras únicas (secuencias de sitios de cebado) en cada extremo.

Según un aspecto, una genoteca de transposomas se utiliza para fabricar fragmentos de ADN genómico en medios acuosos donde se inserta o se une una secuencia de código de barras única a cada extremo del ADN genómico en un sitio que ha sido cortado por la transposasa del transposoma. Dado que cada transposoma tiene una o dos secuencias de sitios de cebado diferentes y/o únicas en comparación con otros miembros del transposoma del conjunto o pluralidad o genoteca, cada fragmento tendrá secuencias de sitios de cebado únicas (secuencias de códigos de barras) en cada extremo. Según un aspecto, cada fragmento incluirá una o más citosina o una o más metilcitosina. Los métodos descritos en la presente memoria utilizan ADN portador durante una etapa de conversión química que convierte la citosina en uracilo para proteger el ADN de la muestra de que se pierda o se dañe durante la etapa de conversión química. Los expertos en la técnica conocen los reactivos o enzimas utilizados para convertir la citosina en uracilo. Los reactivos enzimáticos para convertir la citosina en uracilo, es decir, las citosinas desaminasas, incluyen los de la familia ABOPEC, como APOBEC-seq o APOBEC3A. Los miembros de la familia APOBEC son las citidinas desaminasas que convierten la citosina en uracilo mientras mantienen la 5-hidroximetilcitosina, es decir, sin alterar la 5-hidroximetilcitosina. La 5-metilcitosina se puede convertir en 5-hidroximetilcitosina mediante una reacción de oxidación con la enzima TET. Tales enzimas se describen en la patente US 2013/0244237 y pueden estar disponibles en New England Biolabs. Otros reactivos enzimáticos resultarán evidentes para los expertos en la técnica basándose en la presente descripción. Según un aspecto, el reactivo no es bisulfito o excluye el bisulfito o el reactivo convierte la citosina en uracilo con la condición de que el reactivo no sea bisulfito. El ADN portador puede entonces retirarse o los fragmentos convertidos pueden amplificarse sin amplificar también el ADN portador, lo que da como resultado un ADN fragmentado amplificado. Más específicamente, se proporcionan métodos que utilizan cebadores o adaptadores insertados en el ADN diana fragmentado utilizando un método de transposoma y la adición de ADN portador para reducir la pérdida de ADN y el daño del ADN durante la conversión química o enzimática del ADN, que es necesaria para el análisis de metilación con resolución de bases. El uso de transposomas para crear fragmentos con cebadores o adaptadores en cada extremo, tal como se describe en la presente memoria (y antes de la adición del ADN portador) añade adaptadores de código de barras y PCR al ADN dirigido, de tal modo que el ADN dirigido pueda distinguirse suficientemente del ADN portador. Según los métodos descritos en la presente memoria, el ADN ligado al adaptador se amplifica mientras que el ADN portador no se amplifica. El ADN portador se convierte en ADN monocatenario, es decir, ADNmc, y se elimina de la mezcla, lo que da como resultado un ADN dirigido amplificado puro.

La presente descripción contempla fragmentar el ADN genómico en una pluralidad de fragmentos, tales como 5 o más fragmentos, 10 o más fragmentos, 100 o más fragmentos, 1000 o más fragmentos, 10.000 o más fragmentos, 100.000 o más fragmentos, 1.000.000 o más fragmentos, o 10.000.000 o más fragmentos utilizando una genoteca de transposomas como se describe en la presente memoria. Según un aspecto que depende del número de secuencias de sitios de unión a cebadores únicas y/o diferentes, una genoteca de transposomas incluye de 5 a 10 tipos o clases de miembros del transposoma, de 10 a 100 tipos o clases de miembros del transposoma, 100 o más tipos o clases de miembros del transposoma, 1000 o más tipos o clases de miembros del transposoma, 10.000 o más tipos o clases de miembros del transposoma, 100.000 o más tipos o clases de miembros del transposoma, 1.000.000 o más tipos o clases de miembros del transposoma, o 10.000.000 o más tipos o clases de miembros del transposoma o entre 5 y 50 tipos o clases de miembros del transposoma.

Según un aspecto, cada transposoma incluye dos transposasas y dos ADN de transposón. Cada uno de los dos ADN de transposón del transposoma incluye un sitio de unión a la transposasa y una secuencia del sitio de unión al cebador. Según un aspecto, el ADN de transposón incluye un único sitio de unión a la transposasa y una secuencia única del sitio de unión al cebador. Cada ADN de transposón es un ácido nucleico separado unido a una transposasa en el sitio de unión de la transposasa. El transposoma es un dímero de dos transposasas separadas, cada una unida a su propio ADN de transposón. El dímero puede tener las mismas secuencias del sitio de unión al cebador en cada transposón o puede tener diferentes secuencias del sitio de unión al cebador en cada transposón. Según un aspecto, el transposoma incluye dos ADN de transposón separados e individuales, cada uno unido a su propia transposasa correspondiente. Según un aspecto, el transposoma incluye solo dos transposasas y solo dos ADN de transposón. Según un aspecto, los dos ADN de transposón que forman parte del transposoma son ADN de transposón separados, individuales o no unidos, cada uno unido a su propia transposasa correspondiente.

Según un aspecto, cada miembro del transposoma de la genoteca incluye una secuencia de sitios de cebado única y diferente. La misma secuencia de sitio de cebado única y diferente puede estar presente en cada ADN de transposón del transposoma o puede estar presente una secuencia de sitio de cebado diferente, única y diferente, en cada ADN de transposón del transposoma. De esta manera, cada transposoma incluye una secuencia de sitio de cebado única y diferente que es única y diferente de las secuencias de sitios de cebado de cualquier otro transposoma en la genoteca de transposomas. Según un aspecto, la genoteca de transposomas puede incluir miembros del transposoma que tienen las mismas secuencias de sitios de cebado que otros miembros del transposoma, aunque la probabilidad es relativamente pequeña o insignificante. De esta manera, la genoteca de transposomas puede considerarse un subconjunto de la colección preparada de transposomas, donde el subconjunto incluye solo transposomas con una secuencia de sitios de cebado única y diferente, ya que el objetivo es fragmentar el ADN genómico donde cada sitio de corte de fragmento tiene diferentes secuencias de sitios de cebado. Debe entenderse que el objetivo de fragmentar el ADN genómico en donde cada sitio de corte del fragmento tiene una secuencia de sitio de cebado diferente puede lograrse cuando los transposomas adyacentes tienen cada uno una secuencia de sitio de cebado única y diferente, aunque puede ser compartida por los dos transposones del transposoma. Debe entenderse que el objetivo de fragmentar el ADN genómico donde cada sitio de corte del fragmento tiene una secuencia de sitio de cebado diferente puede lograrse cuando los transposomas adyacentes tienen cada uno dos secuencias de sitio de cebado únicas y diferentes, donde cada transposón del transposoma tiene secuencias de sitio de cebado únicas y diferentes.

Debe entenderse que un número insignificante de sitios de corte puede compartir la misma secuencia de sitios de cebado debido a la preparación de la genoteca de transposomas. Por ejemplo, para un método de preparación de una genoteca determinado, es matemáticamente posible que existan múltiples moléculas de transposomas con la misma secuencia del sitio de cebado, pero la genoteca se prepara de tal modo que el número de secuencias del sitio de cebado diferentes supere significativamente el número de moléculas del transposoma que realmente se insertarán en el genoma diana. Según un aspecto, la genoteca de transposomas puede incluir miembros del transposoma que tienen las mismas secuencias de los dos sitios de cebado, es decir, las secuencias del sitio de cebado son idénticas o iguales, aunque esta secuencia del sitio de cebado es única en comparación con cualquier otro ADN de transposón de los miembros de transposomas de la genoteca de transposomas. Para crear una genoteca de transposomas de este tipo, cada miembro del transposoma se fabrica por separado mezclando la transposasa y el ADN de transposón que contiene la secuencia única del sitio de cebado. Todos los miembros del transposoma se mezclan entonces para formar la genoteca de transposomas.

Según un aspecto, una genoteca de transposomas se prepara mezclando todas las secuencias de transposones junto con la transposasa para formar el transposoma. En este método, la mayoría de los transposomas tienen diferentes secuencias de transposones, pero la probabilidad de que un transposoma lleve las mismas secuencias de transposones es  $1/N$ . Según otro método para fabricar una genoteca de transposomas, cada tipo de secuencia de transposones se mezcla con la transposasa por separado y, a continuación, todos los transposomas se mezclan para formar la genoteca de transposomas. En este método, todos los transposomas tendrán las mismas secuencias de transposones.

Según un aspecto, el número de secuencias de sitios de cebado únicas y/o diferentes está entre 5 y 50, 10 y 50, 15 y 45, 20 y 40 o entre 1 y 1.000, 1 y 10.000, 1 y 100.000, 1 y 1.000.000 o 1 y 10.000.000. Según un aspecto, el número de sitios de corte en el ADN genómico se determina o ajusta mediante la concentración de transposomas, con la mayor concentración dando como resultado un mayor número de sitios de corte y una menor concentración dando como resultado un menor número de sitios de corte. Según un aspecto, el número de transposomas y las secuencias de sitios de cebado diferentes y/o únicas asociadas se selecciona de tal modo que sustancialmente todos los sitios de corte tengan dos secuencias de sitios de cebado diferentes y/o únicas. Según un aspecto, más del 90 % de los sitios de corte tienen dos secuencias de sitio de cebado diferentes y/o únicas, más del 95 % de los sitios de corte tienen dos secuencias de sitio de cebado diferentes y/o únicas, el 96 % de los sitios de corte tienen dos secuencias de sitio de cebado diferentes y/o únicas, el 97 % de los sitios de corte tienen dos secuencias de sitio de cebado diferentes y/o únicas, el 98 % de los sitios de corte tienen dos secuencias de sitio de cebado diferentes y/o únicas, el 99 % de los sitios de corte tienen dos secuencias de sitio de cebado diferentes y/o únicas, el 99,5 % de los sitios de corte tienen dos secuencias de sitio de cebado diferentes y/o únicas, o el 100 % de los sitios de corte tienen dos secuencias de sitio de cebado diferentes y/o únicas.

La genoteca de transposomas se utiliza después para cortar el ADN genómico y cada transposoma inserta o une sus secuencias del sitio de cebado en cada uno de los ADN de transposón en los extremos del sitio de corte. Cuando los transposomas adyacentes tienen secuencias de sitio de cebado únicas y diferentes en comparación entre sí, el sitio de corte tendrá una secuencia de sitio de cebado única y diferente en cada extremo del sitio, es decir, las secuencias del sitio de cebado insertadas serán diferentes. De esta manera, una pluralidad o la mayoría o sustancialmente todos los fragmentos producidos por la genoteca de transposomas tienen una secuencia de sitios de cebado diferente y/o única en cada extremo, es decir, extremos opuestos, del fragmento, en la medida en que los transposomas adyacentes tengan secuencias de sitios de cebado únicas y diferentes en comparación entre sí. La transposasa se puede eliminar entonces de cada fragmento, seguida de una etapa de llenado de espacios, por ejemplo, una etapa de extensión de la polimerasa. La secuencia de fragmentos de ácido nucleico bicatenario resultante puede amplificarse después, por ejemplo, mediante amplificación por PCR múltiple. Los fragmentos pueden secuenciarse a continuación y puede determinarse la secuencia del ADN genómico.

Según un aspecto, el ADN de transposón del transposoma puede incluir secuencias que faciliten los métodos de amplificación, tales como secuencias cebadoras específicas o secuencias promotoras de la transcripción que pueden unirse a los fragmentos de modo que los fragmentos puedan amplificarse antes de la secuenciación, tal como mediante PCR o transcripción de ARN utilizando métodos conocidos por los expertos en la técnica. Debe entenderse que la presente descripción contempla diferentes métodos de amplificación para amplificar los fragmentos y que los diferentes métodos de secuenciación para secuenciar los amplicones no se limitan a ningún método de amplificación o secuenciación en particular.

Algunas realizaciones de la presente descripción se refieren a un método de amplificación por etiquetado múltiple de ácidos nucleicos, como el ADN genómico, como una cantidad baja o pequeña de ADN genómico o una cantidad limitada de ADN, como una secuencia genómica o secuencias genómicas obtenidas de una única célula o una pluralidad de células del mismo tipo de célula o de una muestra de tejido, fluido o sangre obtenida de un individuo o un sustrato. Según ciertos aspectos de la presente descripción, los métodos descritos en la presente memoria se pueden realizar en un solo tubo con una única mezcla de reacción. Según ciertos aspectos de la presente descripción, la muestra de ácido nucleico puede estar dentro de un lisado no purificado o sin procesar de una única célula. Los ácidos nucleicos que se van a someter a los métodos descritos en la presente memoria no necesitan purificarse, tal como mediante purificación en columna, antes de ponerse en contacto con los diversos reactivos y en las diversas condiciones descritas en la presente memoria. Los métodos descritos en la presente memoria reducen la tasa de pérdida, es decir, la pérdida del ácido nucleico diana original para ayudar a proporcionar una cobertura sustancial y uniforme de todo el genoma de una única célula que produce ADN amplificado para una secuenciación de alto rendimiento.

Las realizaciones de la presente invención se refieren en general a métodos y composiciones para fabricar fragmentos de ADN, por ejemplo, fragmentos de ADN de todo el genoma de una única célula que después pueden someterse a métodos de amplificación y secuenciación conocidos por los expertos en la técnica y tal como se describen en la presente memoria. Según ciertos aspectos, los métodos para fabricar fragmentos de ácido nucleico descritos en la presente memoria utilizan una genoteca de transposomas. Según un aspecto, una transposasa como parte de un transposoma se utiliza para crear un conjunto de fragmentos de ADN genómico bicatenario. Según ciertos aspectos, las transposasas tienen la capacidad de unirse al ADN de transposón y dimerizarse cuando se ponen en contacto entre sí, tal como cuando se colocan dentro de un recipiente de reacción o volumen de reacción, formando un dímero complejo de transposasa/ADN de transposón llamado transposoma. Cada ADN de transposón del transposoma incluye un sitio de unión a la transposasa bicatenaria y una primera secuencia de ácido nucleico que incluye una secuencia del sitio de cebado y, opcionalmente, secuencias funcionales tales como un sitio promotor de la transcripción. La primera secuencia de ácido nucleico puede estar en forma de una extensión monocatenaria. Cada transposoma de la genoteca de transposomas incluye una secuencia de sitios de cebado única y diferente que es diferente de las secuencias de sitios de cebado de cada miembro restante de la genoteca de transposomas. Según un aspecto, cada transposoma de la genoteca de transposomas incluye dos secuencias de sitios de cebado únicas y diferentes que son diferentes de las secuencias de sitios de cebado de cada miembro restante de la genoteca de transposomas.

Los transposomas tienen la capacidad de unirse aleatoriamente a ubicaciones diana a lo largo de los ácidos nucleicos bicatenarios, tales como el ADN genómico bicatenario, formando un complejo que incluye el transposoma y el ADN genómico bicatenario. Las transposasas en el transposoma escinden el ADN genómico bicatenario, con una transposasa que escinde la cadena superior y una transposasa que escinde la cadena inferior. Cada uno de los ADN de transposón del transposoma está unido al ADN genómico bicatenario en cada extremo del sitio de corte, es decir, un ADN de transposón del transposoma está unido al sitio de corte de la izquierda y el otro ADN de transposón del transposón está unido al sitio de corte de la derecha. Cuando el ADN de transposón del transposoma tiene cada uno secuencias de sitios de unión al cebador diferentes, el sitio de corte de la izquierda y el sitio de corte de la derecha se “codifican” con un código de barras diferente y único, es decir, secuencias del sitio de cebado. Cuando el ADN de transposón del transposoma tiene cada uno la misma secuencia de sitios de unión al cebador, el sitio de corte de la izquierda y el sitio de corte de la derecha tienen un “código de barras” con el mismo código de barras, es decir, la secuencia del sitio de cebado. Cuando los transposomas adyacentes utilizados para hacer un fragmento tienen cada uno una secuencia de sitios de unión al cebador diferente y única, el fragmento resultante tendrá un sitio de unión al cebador diferente y único en cada extremo del fragmento. Según ciertos aspectos, una pluralidad de dímeros del complejo de transposasa/ADN de transposón, es decir, los transposomas, se unen a una pluralidad correspondiente de ubicaciones diana a lo largo de un ADN genómico bicatenario, por ejemplo, y después escinden el ADN genómico bicatenario en una pluralidad de fragmentos bicatenarios, teniendo cada fragmento ADN de transposón con una secuencia de código de barras diferente unida en cada extremo del fragmento bicatenario.

Según un aspecto, el ADN de transposón está unido al ADN genómico bicatenario y existe un espacio monocatenario entre una cadena del ADN genómico y una cadena del ADN de transposón. Según un aspecto, la extensión del espacio se lleva a cabo para llenar el espacio y crear una conexión bicatenaria entre el ADN genómico bicatenario y el ADN de transposón bicatenario. Según un aspecto, una secuencia de ácido nucleico que incluye el sitio de unión a la transposasa y la secuencia del sitio de cebado se une a cada extremo del fragmento bicatenario. Según ciertos aspectos, la transposasa está unida al ADN de transposón que está unido en cada extremo del fragmento bicatenario. Según un aspecto, las transposasas se eliminan del ADN de transposón que está unido a cada extremo de los fragmentos de ADN genómico bicatenario.

Según un aspecto de la presente descripción, los fragmentos de ADN genómico bicatenario que tienen el ADN de transposón con diferentes secuencias del sitio de cebado unidas en cada extremo de los fragmentos de ADN genómico bicatenario se rellenan entonces y se extienden utilizando el ADN de transposón como plantilla. En consecuencia, se produce un producto de extensión de ácido nucleico bicatenario que incluye el fragmento de ADN genómico bicatenario y un ADN de transposón bicatenario que incluye una secuencia de sitio de cebado diferente en cada extremo del ADN genómico bicatenario.

En esta fase, los productos de extensión de ácidos nucleicos bicatenarios que incluyen el fragmento de ADN genómico y las diferentes secuencias del sitio de cebado en cada extremo pueden amplificarse utilizando métodos conocidos por los expertos en la técnica para producir amplicones del fragmento de ADN genómico y el sitio de unión al cebador diferente en cada extremo. Se pueden utilizar secuencias de cebadores y reactivos de PCR para la amplificación. Los transposones tal como se describen en la presente memoria también pueden incluir un sitio de unión a la ARN polimerasa para la producción de transcritos de ARN que después pueden transcribirse de forma inversa en ADNc para una amplificación lineal. Los productos de extensión de ácido nucleico bicatenario, que incluyen el fragmento de ADN genómico y las diferentes secuencias del sitio de cebado en cada extremo, pueden combinarse con reactivos de amplificación y el fragmento de ácido nucleico genómico bicatenario puede amplificarse a continuación utilizando métodos conocidos por los expertos en la técnica para producir amplicones del fragmento de ácido nucleico genómico bicatenario.

A continuación, los amplicones pueden recogerse y/o purificarse antes de un análisis adicional. Los amplicones se pueden secuenciar utilizando métodos conocidos por los expertos en la técnica. Una vez secuenciadas, las secuencias pueden analizarse computacionalmente para identificar el ADN genómico.

Algunas realizaciones de la presente descripción se refieren a un método de amplificación del ADN mediante el etiquetado múltiple de los extremos, en donde el ADN es una cantidad baja o pequeña de ADN genómico o una cantidad limitada de ADN, tal como una secuencia genómica o secuencias genómicas obtenidas de una única célula o una pluralidad de células del mismo tipo de célula o de una muestra de tejido, fluido o sangre obtenida de un individuo o un sustrato. Según ciertos aspectos de la presente descripción, los métodos descritos en la presente memoria pueden realizarse en un solo tubo para crear los fragmentos que tienen secuencias diferentes y únicas en cada extremo que después se amplifican y secuencian utilizando plataformas de secuenciación de alto rendimiento conocidas por los expertos en la técnica.

El método de fragmentación y codificación de transposomas descrito en la presente memoria es útil para amplificar y después secuenciar cantidades bajas, pequeñas o limitadas de ADN. Los métodos descritos en la presente memoria tienen una aplicación particular en sistemas biológicos o muestras de tejido caracterizadas por poblaciones celulares altamente heterogéneas, tales como masas tumorales y neurales. Los métodos descritos en la presente memoria pueden utilizar diversas fuentes de materiales de ADN, incluidos tejidos genéticamente heterogéneos (p. ej., cánceres), muestras raras y preciadas (p. ej., células madre embrionarias) y células que no se dividen (p. ej., neuronas) y similares, así como plataformas de secuenciación y métodos de genotipado conocidos por los expertos en la técnica.

Otras características y ventajas de ciertas realizaciones de la presente descripción se harán más evidentes en la siguiente descripción de las realizaciones y dibujos de las mismas, y a partir de las reivindicaciones.

#### **Breve descripción de las figuras**

Las anteriores y otras características y ventajas de la presente invención se comprenderán más completamente a partir de la siguiente descripción detallada de las realizaciones ilustrativas tomadas junto con las figuras adjuntas, en los que:

La figura 1 representa esquemáticamente una estructura de un ADN de transposón con una extensión en 5' que es lineal, donde T es el sitio de unión de la transposasa bicatenaria y M es un sitio de cebado múltiple en un extremo de la extensión.

La figura 2 es un esquema de una realización general de la transposasa y ADN de transposón que forma espontáneamente un transposoma, que puede producirse dentro de una gotita u otro medio de formación. Antes de la formación del transposoma, cada transposón tiene una secuencia de sitios de cebado diferente y única representada por diferentes patrones. Tras la formación del transposoma, cada transposón del transposoma tiene una secuencia de sitios de cebado diferente y única representada por diferentes patrones.

La figura 3A es un esquema de la unión del transposoma al ADN genómico, el corte en fragmentos y la adición o inserción del ADN de transposón que incluye un sitio de unión a la transposasa (negro) y una secuencia de sitio de cebado única y diferente en cada transposón de cada transposoma, tal como se representa en cada transposoma mediante diferentes patrones.

La figura 3B es un esquema de la unión del transposoma al ADN genómico, el corte en fragmentos y la adición o inserción del ADN de transposón que incluye un sitio de unión a la transposasa (negro) y una secuencia del sitio de

cebado única y diferente representativa del transposoma, es decir, la misma secuencia del sitio de unión del cebador única y diferente está presente en cada transposón del transposoma, tal como se representa en cada transposoma mediante el mismo patrón. Las diferentes secuencias del sitio de unión al cebador entre cada transposoma están representadas por diferentes patrones.

La figura 4 es un esquema de la eliminación de la transposasa, el rellenado de espacios para formar productos de extensión de ácidos nucleicos que incluyen el ADN genómico, el sitio de unión de la transposasa y una secuencia de sitio de cebado única y diferente en cada extremo del producto de extensión.

La figura 5 es un esquema que muestra la amplificación por PCR múltiple de los fragmentos de la figura 4.

La figura 6 es un esquema que muestra el flujo de trabajo general del método de detección de metilación de bajo aporte.

La figura 7 es un esquema que muestra el flujo de trabajo de scEM -Seq.

La figura 8 es la cobertura genómica de scEM-seq en comparación con sc-COOL seq y scWGBs utilizando diferentes algoritmos de alineación (PE o PE+SE). PE: el uso de la alineación de los extremos del par. PE+SE: el uso de la alineación de un solo extremo para las lecturas no se alineó con el genoma de referencia. Las secuencias scEM-seq y sc-COOL se realizan en embriones de ratón en estadio de 4 células. El scWGBS se realiza en ESC de ratón.

La figura 9 es la comparación de la velocidad de mapeo entre scEM-seq, sc-COOL seq y scWGBs.

La figura 10 es la comparación de la distribución de la tasa de metilación entre sc-COOL seq y scEM-seq.

## Descripción detallada

Según un aspecto, se proporciona un método de amplificación, secuenciación y ensamblaje del genoma completo de una única célula que incluye poner en contacto el ADN genómico bicatenario de una única célula con transposasas, tales como las transposasas de Tn5, cada una unida a un ADN de transposón, en donde el ADN del transposón incluye un sitio de unión a la transposasa (Tnp) bicatenario de 19 pares de bases y una primera secuencia de ácido nucleico que incluye una secuencia de sitio de cebado única y diferente para formar un dímero de complejo de transposasa/ADN de transposón llamado transposoma. La primera secuencia de ácido nucleico puede estar en forma de una extensión monocatenaria. Según un aspecto, la primera secuencia de ácido nucleico puede ser un saliente, tal como un saliente en 5', en donde el saliente incluye una secuencia de sitio de cebado única y diferente. El saliente puede incluir otras secuencias funcionales, según se desee. El saliente puede ser de cualquier longitud adecuada para incluir una secuencia del sitio de cebado u otras secuencias funcionales, según se desee. El transposoma se une a ubicaciones diana a lo largo del ADN genómico bicatenario y escinde el ADN genómico bicatenario en una pluralidad de fragmentos bicatenarios, teniendo cada fragmento bicatenario un primer complejo unido a una cadena superior por el sitio de unión a Tnp y un segundo complejo unido a una cadena inferior por el sitio de unión a Tnp. El sitio de unión al transposón y, por lo tanto, el ADN de transposón junto con el sitio de unión al cebador, está unido a cada extremo 5' del fragmento bicatenario. Según un aspecto, las transposasas de Tn5 se eliminan del complejo.

Los fragmentos bicatenarios incluyen una o más citosina o una o más metilcitosina y se extienden a lo largo del ADN de transposón para formar un producto de extensión bicatenario que tiene secuencias de sitios de cebado diferentes, distintas o únicas en cada extremo del producto de extensión bicatenario. Según un aspecto, se puede llenar un espacio que puede resultar de la unión del sitio de unión de la transposasa Tn5 al fragmento de ADN genómico bicatenario.

El producto de extensión bicatenario llenado se somete después a un tratamiento químico en presencia de ADN portador para convertir la citosina en uracilo.

El producto de extensión bicatenario llenado se mezcla con reactivos de amplificación y el fragmento de ADN genómico bicatenario se amplifica. Los amplicones, que incluyen una secuencia de sitio de cebado diferente o diferente o única (que puede funcionar como una secuencia de código de barras) en cada extremo, se secuencian utilizando, por ejemplo, métodos de secuenciación de alto rendimiento conocidos por los expertos en la técnica.

En un aspecto particular, las realizaciones se refieren a métodos para la amplificación, secuenciación y ensamblaje de prácticamente todo el genoma sin pérdida de representación de sitios específicos (definidos en la presente memoria como "amplificación del genoma completo"). En una realización específica, la amplificación del genoma completo comprende la amplificación de sustancialmente todos los fragmentos o todos los fragmentos de una genoteca genómica. En una realización específica adicional, "sustancialmente la totalidad" o "sustancialmente todas" se refiere a aproximadamente el 80 %, aproximadamente el 85 %, aproximadamente el 90 %, aproximadamente el 95 %, aproximadamente el 97 % o aproximadamente el 99 % de todas las secuencias de un genoma.

La práctica de ciertas realizaciones o características de ciertas realizaciones puede emplear, a menos que se indique lo contrario, técnicas convencionales de biología molecular, microbiología, ADN recombinante, etc., que son de



conocimiento habitual en la materia. Tales técnicas se explican detalladamente en la bibliografía. Véase, p. ej., Sambrook, Fritsch y Maniatis, MOLECULAR CLONING: A LABORATORY MANUAL, segunda edición (1989), OLIGONUCLEOTIDE SYNTHESIS (M. J. Gait Ed., 1984), ANIMAL CELL CULTURE (R. I. Freshney, Ed., 1987), la serie METHODS IN ENZYMOLOGY (Academic Press, Inc.); GENE TRANSFER VECTORS FOR MAMMALIAN CELLS (J. M. Miller y M. P. Calos eds. 1987), HANDBOOK OF EXPERIMENTAL IMMUNOLOGY, (D. M. Weir y C. C. Blackwell, Eds.), CURRENT PROTOCOLS IN MOLECULAR BIOLOGY (F. M. Ausubel, R. Brent, R. E. Kingston, D. D. Moore, J. G. Siedman, J. A. Smith, y K. Struhl, eds., 1987), CURRENT PROTOCOLS IN IMMUNOLOGY (J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach y W. Strober, eds., 1991); ANNUAL REVIEW OF IMMUNOLOGY; así como monografías en revistas como ADVANCES IN IMMUNOLOGY. Todas las patentes, solicitudes de patente y publicaciones mencionadas en la presente memoria, tanto anteriormente como más adelante, se incorporan en la presente memoria como referencia.

Los términos y símbolos de la química, la bioquímica, la genética y la biología molecular de los ácidos nucleicos utilizados en la presente memoria siguen los de los tratados y textos estándar en el campo, p. ej., Komberg y Baker, DNA Replication, segunda edición (W.H. Freeman, Nueva York, 1992); Lehninger, Biochemistry, segunda edición (Worth Publishers, Nueva York, 1975); Strachan y Read, Human Molecular Genetics, segunda edición (Wiley-Liss, New York, 1999); Eckstein, editor, Oligonucleotides and Analogs: A Practical Approach (Oxford University Press, Nueva York, 1991); Gait, editor, Oligonucleotide Synthesis: A Practical Approach (IRL Press, Oxford, 1984); y similares.

#### Definiciones

Como se usa en la presente memoria, el término “muestra biológica” pretende incluir, pero no se limita a, tejidos, células, fluidos biológicos y aislados de los mismos, aislados de un sujeto, así como tejidos, células y fluidos presentes en un sujeto.

El término “*in vitro*” tiene un significado reconocido en la técnica, p. ej., que implica reactivos o extractos purificados, p. ej., extractos celulares. El término “*in vivo*” también tiene un significado reconocido en la técnica, p. ej., que implica células vivas, p. ej., células inmortalizadas, células primarias, líneas celulares y/o células de un organismo.

Como se usa en la presente memoria, los términos “complementario” y “complementariedad” se utilizan en referencia a secuencias de nucleótidos relacionadas por las reglas de apareamiento de bases. Por ejemplo, la secuencia 5'-AGT-3' es complementaria a la secuencia 5'-ACT-3'. La complementariedad puede ser parcial o total. La complementariedad parcial se produce cuando una o más bases de ácidos nucleicos no coinciden según las reglas de apareamiento de bases. La complementariedad total o completa entre los ácidos nucleicos se produce cuando cada base de ácido nucleico se empareja con otra base según las reglas de apareamiento de bases. El grado de complementariedad entre las cadenas de ácido nucleico tiene efectos significativos sobre la eficacia y la fuerza de la hibridación entre las cadenas de ácido nucleico.

El término “hibridación” se refiere al apareamiento de ácidos nucleicos complementarios. La hibridación y la fuerza de la hibridación (es decir, la fuerza de la asociación entre los ácidos nucleicos) se ven afectadas por factores tales como el grado de complementariedad entre los ácidos nucleicos, la rigurosidad de las condiciones implicadas, la  $T_m$  del híbrido formado y la relación G: C dentro de los ácidos nucleicos. Se dice que una única molécula que contiene un apareamiento de ácidos nucleicos complementarios dentro de su estructura se “autohibrida”.

El término “ $T_m$ ” se refiere a la temperatura de fusión de un ácido nucleico. La temperatura de fusión es la temperatura a la que una población de moléculas de ácido nucleico bicatenario se disocia a la mitad en cadenas sencillas. La ecuación para calcular la  $T_m$  de los ácidos nucleicos es bien conocida en la técnica. Como indican las referencias estándar, una estimación simple del valor  $T_m$  puede calcularse mediante la ecuación:  $T_m = 81,5 + 0,41 (\% G + C)$ , cuando un ácido nucleico está en solución acuosa con NaCl 1 M (véase, p. ej., Anderson and Young, Quantitative Filter Hybridization, in Nucleic Acid Hybridization (1985)). Otras referencias incluyen cálculos más sofisticados que tienen en cuenta las características estructurales y de secuencia para el cálculo de  $T_m$ .

El término “rigurosidad” se refiere a las condiciones de temperatura, fuerza iónica y presencia de otros compuestos, tales como disolventes orgánicos, en las que se llevan a cabo las hibridaciones de ácidos nucleicos.

Las “condiciones de baja rigurosidad”, cuando se utilizan en referencia a la hibridación de ácidos nucleicos, comprenden condiciones equivalentes a la unión o hibridación a 42 °C en una solución que consiste en 5 veces SSPE (43,8 g/l de NaCl, 6,9 g/l de  $\text{NaH}_2\text{PO}_4(\text{H}_2\text{O})$  y 1,85 g/l de EDTA, pH ajustado a 7,4 con NaOH), SDS al 0,1 %, 5 veces el reactivo de Denhardt (50 veces el de Denhardt) contiene por cada 500 ml: 5 g de Ficoll (tipo 400, Pharmacia), 5 g de BSA (fracción V; Sigma) y 100 mg/ml de ADN de esperma de salmón desnaturalizado, seguido de lavado en una solución que comprende 5 veces SSPE y SDS al 0,1 % a 42 °C cuando se emplea una sonda de aproximadamente 500 nucleótidos de longitud.

Las “condiciones de rigurosidad media”, cuando se utilizan en referencia a la hibridación de ácidos nucleicos, comprenden condiciones equivalentes a la unión o hibridación a 42 °C en una solución que consiste en 5 veces SSPE

(43,8 g/l de NaCl, 6,9 g/l de  $\text{NaH}_2\text{PO}_4(\text{H}_2\text{O})$  y 1,85 g/l de EDTA, pH ajustado a 7,4 con NaOH), 0,5 % de SDS, 5 veces de reactivo de Denhardt y 100 mg/ml de desnaturalizado ADN de esperma de salmón seguido de lavado en una solución que comprende 1,0x SSPE y 1,0 % de SDS a 42 °C cuando se emplea una sonda de aproximadamente 500 nucleótidos de longitud.

Las “condiciones de alta rigurosidad”, cuando se utilizan en referencia a la hibridación de ácidos nucleicos, comprenden condiciones equivalentes a la unión o hibridación a 42 °C en una solución que consiste en 5 veces SSPE (43,8 g/l de NaCl, 6,9 g/l de  $\text{NaH}_2\text{PO}_4(\text{H}_2\text{O})$  y 1,85 g/l de EDTA, pH ajustado a 7,4 con NaOH), SDS al 0,5 %, reactivo de Denhardt 5 veces y 100 mg/ml de desnaturalizado ADN de esperma de salmón seguido de lavado en una solución que comprende 0,1 x SSPE y SDS al 1,0 % a 42 °C cuando se emplea una sonda de aproximadamente 500 nucleótidos de longitud.

El término “genoma”, como se usa en la presente memoria, se define como el conjunto de genes colectivo portado por un individuo, célula u orgánulo. El término “ADN genómico”, como se usa en la presente memoria, se define como un material de ADN que comprende el conjunto de genes colectivo parcial o completo portado por un individuo, célula u orgánulo.

Como se usa en la presente memoria, el término “nucleósido” se refiere a una molécula que tiene una base de purina o pirimidina unida covalentemente a un azúcar ribosa o desoxirribosa. Los nucleósidos ilustrativos incluyen adenosina, guanosina, citidina, uridina y timidina. Nucleósidos ilustrativos adicionales incluyen inosina, 1-metilguanosina, pseudouridina, 5,6-dihidrouridina, ribotimidina, 2N-metilguanosina y 2,2N,N-dimetilguanosina (también denominados nucleósidos “raros”). El término “nucleótido” se refiere a un nucleósido que tiene uno o más grupos fosfato unidos en enlaces éster al resto de azúcar. Los nucleótidos ilustrativos incluyen monofosfatos, difosfatos y trifosfatos de nucleósidos. Los términos “polinucleótido”, “oligonucleótido” y “molécula de ácido nucleico” se utilizan indistintamente en la presente memoria y se refieren a un polímero de nucleótidos, ya sea desoxirribonucleótidos o ribonucleótidos, de cualquier longitud unidos por un enlace fosfodiéster entre los átomos de carbono 5' y 3'. Los polinucleótidos pueden tener cualquier estructura tridimensional y pueden realizar cualquier función, conocida o desconocida. Los siguientes son ejemplos no limitativos de polinucleótidos: un gen o fragmento génico (por ejemplo, una sonda, cebador, etiqueta EST o SAGE), exones, intrones, ARN mensajero (ARNm), ARN de transferencia, ARN ribosómico, ribozimas, ADNc, polinucleótidos recombinantes, polinucleótidos ramificados, plásmidos, vectores, ADN aislado de cualquier secuencia, sondas de ácidos nucleicos y cebadores. Un polinucleótido puede comprender nucleótidos modificados, tales como nucleótidos metilados y análogos de nucleótidos. El término también se refiere a moléculas bicatenarias y monocatenarias. A menos que se especifique o se requiera lo contrario, cualquier realización de esta invención que comprenda un polinucleótido abarca tanto la forma bicatenaria como cada una de las dos formas monocatenarias complementarias que se sabe o se prevé que formen la forma bicatenaria. Un polinucleótido está compuesto por una secuencia específica de cuatro bases nucleotídicas: adenina (A); citosina (C); guanina (G); timina (T); y uracilo (U) para la timina cuando el polinucleótido es ARN. Por lo tanto, el término secuencia de polinucleótidos es la representación alfabética de una molécula de polinucleótido. Esta representación alfabética puede introducirse en bases de datos en un ordenador que tenga una unidad central de procesamiento y utilizarse para aplicaciones bioinformáticas tales como genómica funcional y búsqueda de homología.

Los términos “ADN”, “molécula de ADN” y “molécula de ácido desoxirribonucleico” se refieren a un polímero de desoxirribonucleótidos. El ADN se puede sintetizar de forma natural (p. ej., mediante la replicación del ADN). El ARN puede modificarse postranscripcionalmente. El ADN también se puede sintetizar químicamente. El ADN puede ser monocatenario (es decir, ADNmc) o multicatenario (p. ej., bicatenario, es decir, ADNbc).

Los términos “análogo de nucleótido”, “nucleótido alterado” y “nucleótido modificado” se refieren a un nucleótido no estándar, incluidos los ribonucleótidos o desoxirribonucleótidos de origen no natural. En ciertas realizaciones ilustrativas, los análogos de nucleótidos se modifican en cualquier posición para alterar ciertas propiedades químicas del nucleótido y conservar la capacidad del análogo de nucleótidos para realizar su función prevista. Los ejemplos de posiciones del nucleótido que pueden derivatizarse incluyen la posición 5, p. ej., 5- (2-amino)propil uridina, 5-bromouridina, 5-propino uridina, 5-propenil uridina, etc.; la posición 6, p. ej., 6- (2-amino) propil uridina; la posición 8 para adenosina y/o guanosinas, p. ej., 8-bromoguanosina, 8-cloro-guanosina, 8-fluoroguanosina, etc. Los análogos de nucleótidos también incluyen nucleótidos deaza-, p. ej., 7-deaza-adenosina; nucleótidos modificados con O y N (p. ej., alquilados, p. ej., N6-metil adenosina, o como se conoce de otro modo en la técnica); y otros análogos de nucleótidos modificados heterocíclicamente tales como los descritos en Herdewijn, Antisense Nucleic Acid Drug Dev., 10 de agosto de 2000 (4):297-310.

Los análogos de nucleótidos también pueden comprender modificaciones en la porción de azúcar de los nucleótidos. Por ejemplo, el grupo OH 2' puede reemplazarse por un grupo seleccionado entre H, OR, R, F, Cl, Br, I, SH, SR,  $\text{NH}_2$ , NHR,  $\text{NR}_2$ , COOR u OR, en donde R es alquilo, alquenoilo, alquinilo, arilo, etc.  $\text{C}_1\text{-C}_6$  sustituido o no sustituido. Otras posibles modificaciones incluyen las descritas en las patentes n.º US-5.858.988 y US-6.291.438.

El grupo fosfato del nucleótido también puede modificarse, p. ej., sustituyendo uno o más de los oxígenos del grupo fosfato por azufre (p. ej., fosforotioatos), o realizando otras sustituciones que permitan que el nucleótido desempeñe su función prevista, tal como se describe, p. ej., en Eckstein, Antisense Nucleic Acid Drug Dev. 10 de abril de 2000

(2):117-21, Rusckowski y col. Antisense Nucleic Acid Drug Dev. 10 de octubre de 2000 (5):333-45, Stein, Antisense Nucleic Acid Drug Dev. 11 de octubre de 2001 (5): 317-25, Vorobjev y col. Antisense Nucleic Acid Drug Dev. 11 de abril de 2001 (2):77-85, y la patente n.º US-5.684.143. Algunas de las modificaciones mencionadas anteriormente (p. ej., modificaciones del grupo fosfato) disminuyen la velocidad de hidrólisis de, por ejemplo, los polinucleótidos que comprenden dichos análogos in vivo o in vitro.

#### Obtención de muestras de adn

Los ácidos nucleicos procesados mediante los métodos descritos en la presente memoria pueden ser ADN y pueden obtenerse de cualquier fuente útil, tal como, por ejemplo, una muestra humana. En realizaciones específicas, una molécula de ADN bicatenario se define además como que comprende un genoma, tal como, por ejemplo, uno obtenido de una muestra de un ser humano. La muestra puede ser cualquier muestra de un ser humano, tal como sangre, suero, plasma, líquido cefalorraquídeo, raspaduras de mejillas, aspiración de pezones, biopsia, semen (que puede denominarse eyaculado), orina, heces, folículo piloso, saliva, sudor, cromatina inmunoprecipitada o aislada físicamente, etc. En realizaciones específicas, la muestra comprende una única célula. En realizaciones específicas, la muestra incluye solamente una única célula.

Según un aspecto, la muestra de ADN es ADN genómico, ADN cromosómico microdisecado, ADN de cromosoma artificial de levadura (YAC), ADN plasmídico, ADN cósmido, ADN de fago, ADN de cromosoma artificial derivado de P1 (PAC) o ADN de cromosoma artificial bacteriano (BAC), ADN mitocondrial, ADN de cloroplasto, ADN de muestra forense u otro ADN de fuentes naturales o artificiales a analizar. En otra realización preferida, la muestra de ADN es ADN de mamífero, ADN vegetal, ADN de levadura, ADN viral o ADN procariótico. En otra realización preferida, la muestra de ADN se obtiene de un ser humano, bovino, porcino, ovino, equino, roedor, ave, pez, camarón, planta, levadura, virus o bacteria. Preferiblemente, la muestra de ADN es ADN genómico.

En ciertas realizaciones ilustrativas, se identifican las células y, a continuación, se aísla una única célula o una pluralidad de células. Las células dentro del alcance de la presente descripción incluyen cualquier tipo de célula en la que los expertos en la técnica consideren útil comprender el contenido de ADN. Una célula según la presente descripción incluye una célula cancerosa de cualquier tipo, hepatocito, ovocito, embrión, célula madre, célula iPS, célula ES, neurona, eritrocito, melanocito, astrocito, célula germinal, oligodendrocito, célula renal y similares. Según un aspecto, los métodos de la presente invención se ponen en práctica con el ADN celular de una única célula. Una pluralidad de células incluye de aproximadamente 2 a aproximadamente 1.000.000 de células, de aproximadamente 2 a aproximadamente 10 células, de aproximadamente 2 a aproximadamente 100 células, de aproximadamente 2 a aproximadamente 1.000 células, de aproximadamente 2 a aproximadamente 10.000 células, de aproximadamente 2 a aproximadamente 100.000 células, de aproximadamente 2 a aproximadamente 10 células o de aproximadamente 2 a aproximadamente 5 células.

Como se usa en la presente memoria, una "célula única" se refiere a una célula. Las células solas útiles en los métodos descritos en la presente memoria pueden obtenerse de un tejido de interés o de una biopsia, muestra de sangre o cultivo celular. Además, se pueden obtener y utilizar células de órganos, tejidos, tumores, neoplasias o similares específicos en los métodos descritos en la presente memoria. Además, en general, se pueden utilizar células de cualquier población en los métodos, tales como una población de organismos unicelulares procariotas o eucariotas, incluidas bacterias o levaduras. Se puede obtener una suspensión de células individuales utilizando métodos estándar conocidos en la técnica que incluyen, por ejemplo, el uso enzimático de tripsina o papaína para digerir proteínas que conectan las células en muestras de tejido o la liberación de células adherentes en cultivo, o la separación mecánica de las células en una muestra. Las células solas se pueden colocar en cualquier recipiente de reacción adecuado en el que las células solas se pueden tratar individualmente. Por ejemplo, una placa de 96 pocillos, de modo que cada célula única se coloque en un único pocillo.

Los métodos para manipular células solas son conocidos en la técnica e incluyen la clasificación celular activada por fluorescencia (FACS), la citometría de flujo (Herzenberg., PNAS USA 76:1453-55 1979), la micromanipulación y el uso de recolectores de células semiautomáticos (p. ej., el sistema de transferencia celular Quixell™ de Stoelting Co.). Las células individuales pueden, por ejemplo, seleccionarse individualmente en función de las características detectables mediante observación microscópica, como la ubicación, la morfología o la expresión del gen indicador. Además, también se puede utilizar una combinación de centrifugación en gradiente y citometría de flujo para aumentar la eficiencia de aislamiento o clasificación.

Una vez que se ha identificado una célula deseada, la célula se lisa para liberar el contenido celular, incluido el ADN, utilizando métodos conocidos por los expertos en la técnica. El contenido celular está contenido dentro de un recipiente o un volumen de recolección. En algunos aspectos de la invención, el contenido celular, tal como el ADN genómico, puede liberarse de las células mediante la lisis de las células. La lisis se puede lograr, por ejemplo, calentando las células, o mediante el uso de detergentes u otros métodos químicos, o mediante una combinación de estos. Sin embargo, se puede utilizar cualquier método de lisis adecuado conocido en la técnica. Por ejemplo, calentar las células a 72 °C durante 2 minutos en presencia de Tween-20 es suficiente para lisar las células. Alternativamente, las células pueden calentarse a 65 °C durante 10 minutos en agua (Esumi y col., Neurosci Res 60 (4):439-51 (2008)); o 70 °C durante 90 segundos en el regulador II de PCR (Applied Biosystems) suplementado con un 0,5 % de NP-40 (Kurimoto y col., Nucleic Acids Res 34 (5):e42 (2006));

o la lisis se puede lograr con una proteasa tal como la proteinasa K o mediante el uso de sales caotrópicas tales como el isotiocianato de guanidina (publicación n.º US-2007/0281313). La amplificación del ADN genómico según los métodos descritos en la presente memoria se puede realizar directamente en los lisados celulares, de tal modo que se puede añadir una mezcla de reacción a los lisados celulares. Alternativamente, el lisado celular se puede separar en dos o más volúmenes, tal como en dos o más recipientes, tubos o regiones utilizando métodos conocidos por los expertos en la técnica con una porción del lisado celular contenida en cada recipiente, tubo o región de volumen. El ADN genómico contenido en cada recipiente, tubo o región puede amplificarse entonces mediante los métodos descritos en la presente memoria o los métodos conocidos por los expertos en la técnica.

Un ácido nucleico utilizado en la invención también puede incluir bases nativas o no nativas. A este respecto, un ácido desoxirribonucleico nativo puede tener una o más bases seleccionadas del grupo que consiste en adenina, timina, citosina o guanina y un ácido ribonucleico puede tener una o más bases seleccionadas del grupo que consiste en uracilo, adenina, citosina o guanina. Las bases no nativas ilustrativas que pueden incluirse en un ácido nucleico, ya sea que tengan una estructura básica nativa o analógica, incluyen, sin limitación, inosina, xantina, hipoxantina, isocitosina, isoguanina, 5-metilcitosina, 5-hidroximetilcitosina, 2-aminoadenina, 6-metiladenina, 6-metilguanina, 2-propilguanina, 2-propiladenina, 2-tioLiracilo, 2-tiotimina, 2-tiocitosina, 15-halouracilo, 15-halocitosina, 5-propiniluracilo, 5-propinilcitosina, 6-azouracilo, 6-azocitosina, 6-azotimina, 5-uracilo, 4-tiouracilo, 8-halo adenina o guanina, 8-aminoadenina o guanina, 8-tioladenina o guanina, 8-tioalquiladenina o guanina, 8-hidroxiadenina o guanina, uracilo o citosina sustituidos con 5-halo, 7-metilguanina, 7-metiladenina, 8-azaguanina, 8-azaadenina, 7-deazaguanina, 7-deazaadenina, 3-deazaguanina, 3-deazaadenina o similares. Una realización particular puede utilizar isocitosina e isoguanina en un ácido nucleico para reducir la hibridación no específica, como se describe generalmente en la patente n.º US-5.681.702.

#### Uso de transposomas para generar fragmentos

La presente invención se basa en parte en métodos para fabricar plantillas de fragmentos de ácido nucleico, tales como a partir de ADN o ADN genómico, utilizando una transposasa o transposoma para fragmentar la secuencia de ácido nucleico original o de partida, tal como el ADN genómico, y para unir una secuencia de sitio de cebado diferente a cada extremo de un sitio de corte o fragmentación para producir así un conjunto de fragmentos en el que cada miembro del conjunto tiene dos secuencias de sitios de cebado únicas y diferentes. Las plantillas de fragmentos de ácido nucleico se amplifican para producir amplicones. Los amplicones de las plantillas de fragmentos de ácido nucleico pueden recogerse y secuenciarse. Los amplicones recolectados forman una genoteca de amplicones de los fragmentos del ácido nucleico original, tal como el ADN genómico. Según un aspecto ilustrativo, se describen métodos para fabricar fragmentos de ácido nucleico utilizando una enzima tal como Tn5. Tales métodos son conocidos en la técnica e incluyen los que se practican utilizando el kit Illumina Nextera. Según un aspecto ilustrativo, los métodos descritos en la presente memoria utilizan una genoteca de transposomas y un método denominado "tagmentación" en la medida en que los fragmentos se crean a partir de una secuencia de ADNbc más grande en la que los fragmentos se etiquetan con cebadores para su uso en la extensión y amplificación de un solo cebador.

Según un aspecto, se obtiene un ADN genómico, tal como el ácido nucleico genómico obtenido de una célula única lisada. Se utiliza una pluralidad o genoteca de transposomas para cortar el ADN genómico en fragmentos bicatenarios. Cada transposoma de la pluralidad o genoteca es un dímero de una transposasa unida a un ADN de transposón, es decir, cada transposoma incluye dos ADN de transposón separados. Cada ADN de transposón de un transposoma incluye un sitio de unión a la transposasa y una secuencia del sitio de unión al cebador. La secuencia del sitio de unión del cebador es exclusiva del transposoma. Según un aspecto, la secuencia del sitio de cebado de cada transposón de un transposoma podría ser única y/o diferente. Según un aspecto, la secuencia del sitio de cebado de cada transposón de un transposoma podría ser la misma. Según un aspecto, la mayoría del transposoma tiene dos ADN de transposón que tienen diferentes secuencias de sitios de cebado y solo una pequeña fracción del transposoma tiene dos ADN de transposón que tienen la misma secuencia de sitios de cebado. Según un aspecto, la secuencia del sitio de cebado de los dos ADN de transposón de cada miembro del transposoma puede ser la misma, pero la secuencia o secuencias del sitio de cebado del ADN de transposón de diferentes miembros del transposoma son únicas y diferentes. Las secuencias de ácido nucleico del transposón que tienen secuencias del sitio de cebado que se unen a la transposasa pueden metilarse en cada citosina, de modo que la secuencia del sitio de cebado no cambie después de la conversión de citosina requerida para la detección de la metilación.

Según un aspecto, las secuencias del sitio de cebado de cada ADN de transposón de un transposoma son únicas y diferentes. Según un aspecto, la secuencia o secuencias del sitio de cebado del ADN de transposón de un transposoma son únicas y diferentes de los miembros restantes de la pluralidad o genoteca de transposomas. Según un aspecto, cada transposoma de la pluralidad o genoteca de transposomas tiene sus propias secuencias de sitios de cebado únicas y diferentes que son diferentes de los miembros restantes de la pluralidad o genoteca de transposomas y pueden tener dos secuencias de sitios de cebado únicas y diferentes que son diferentes de los miembros restantes de la pluralidad o genoteca de transposomas. El ADN de transposón se une a las cadenas superior e inferior de cada fragmento bicatenario en cada sitio de corte o fragmentación. Dado que la secuencia del sitio de cebado puede ser diferente para cada ADN de transposón, el sitio de corte o fragmentación se etiqueta con diferentes secuencias del sitio de cebado. Dado que la secuencia del sitio de cebado puede ser la misma para cada ADN de transposón, el sitio de corte o fragmentación se etiqueta con la misma secuencia del sitio de cebado. Cuando los transposomas

adyacentes utilizados para generar un fragmento tienen cada uno diferentes secuencias de sitios de unión al cebador asociadas al mismo, el fragmento tiene diferentes secuencias de sitios de unión al cebador en cada extremo del fragmento. En consecuencia, el fragmento tendrá dos secuencias de sitios de unión al cebador únicas y diferentes. Dado que cada transposoma tiene su propia secuencia de sitios de cebado única y/o diferente asociada a él (y puede tener dos secuencias de sitios de cebado únicas y/o diferentes asociadas a la misma), y se utiliza una genoteca de transposomas para crear muchos sitios de corte o fragmentación, cada sitio de corte o fragmentación tendrá una secuencia de sitios de cebado diferente y única unida en cada extremo del sitio de corte y cada fragmento tendrá secuencias de sitios de cebado diferentes y/o únicas en cada extremo del fragmento. En consecuencia, la genoteca de transposomas crea muchos fragmentos de la secuencia de ácido nucleico original y cada fragmento tiene una secuencia de sitio de cebado diferente en cada extremo del fragmento. Los fragmentos bicatenarios se procesan después para llenar los espacios. Los fragmentos se amplifican utilizando reactivos de amplificación adecuados, tales como secuencias cebadoras, ADN polimerasa y nucleótidos para la amplificación por PCR, y se secuencian utilizando métodos conocidos por los expertos en la técnica.

Según ciertos aspectos, un sistema de transposones ilustrativo incluye la transposasa Tn5, la transposasa Mu, la transposasa Tn7 o la transposasa IS5 y similares. Los expertos en la técnica conocen otros sistemas de transposones útiles e incluyen el sistema de transposones Tn3 (véase Maekawa, T., Yanagihara, K. y Ohtsubo, E. (1996), A cellfree system of Tn3 transposition and transposition immunity, *Genes Cells* 1, 1007-1016), el sistema de transposones Tn7 (véase Craig, N.L. (1991), Tn7: a target site-specific transposon, *Mol. Microbiol.* 5, 2569-2573), sistema de transposones Tn10 (véase Chalmers, R., Sewitz, S., Lipkow, K. y Crellin, P. (2000), Complete nucleotide sequence of Tn10, *J. Bacteriol* 182, 2970-2972), sistema de transposones Piggybac (véase Li, X., Burnight, E.R., Cooney, A. L., Malani, N., Brady, T., Sander, J.D., Staber, J., Wheelan, S.J., Joung, J.K., McCray, P.B., Jr., y col. (2013), PiggyBac transposase tools for genome engineering, *Proc. Natl. Acad. Sci. USA* 110, E2279-2287), sistema de transposones Sleeping beauty (véase Ivics, Z., Hackett, P. B., Plasterk, R.H., e Izsvak, Z. (1997), Molecular reconstruction of Sleeping Beauty, a Tc1-like transposon from fish, and its transposition in human cells, *Cell* 91, 501-510), sistema de transposones Tol2 (véase Kawakami, K. (2007), Tol2: a versatile gene transfer vector in vertebrates, *Genome Biol.* 8 Suppl. 1, S7.)

Se describen sistemas de transposición de Tn5 particulares y están disponibles para los expertos en la técnica. Véase Goryshin, I.Y. y W.S. Reznikoff, Tn5 in vitro transposition. *The Journal of biological chemistry*, 1998. 273(13): págs. 7367-74; Davies, D.R., y col., Three-dimensional structure of the Tn5 synaptic complex transposition intermediate. *Science*, 2000. 289(5476): págs. 77 a 85; Goryshin, I.Y., y col. Insertional transposon mutagenesis by electroporation of released Tn5 transposition complexes. *Nature biotechnology*, 2000. 18(1): págs. 97-100 y Steiniger-White, M., I. Rayment y W.S. Reznikoff, Structure/function insights into Tn5 transposition. *Current opinion in structural biology*, 2004. 14(1): págs. 50-7, cada uno de los cuales se incorpora en la presente memoria como referencia en su totalidad para todos los propósitos. Se conocen kits que utilizan un sistema de transposición de Tn5 para la preparación de genotecas de ADN y otros usos. Véase Adey, A., y col., Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density in vitro transposition. *Genome biology*, 2010. 11(12): pág. R119; Marine, R., y col., Evaluation of a transposase protocol for rapid generation of shotgun high-throughput sequencing libraries from nanogram quantities of DNA. *Applied and environmental microbiology*, 2011. 77(22): págs. 8071-9; Parkinson, N.J., y col., Preparation of high-quality next-generation sequencing libraries from picogram quantities of target DNA. *Genome research*, 2012. 22(1): p. 125-33; Adey, A. t J. Shendure, Ultra-low-input, tagmentation-based whole-genome bisulfite sequencing. *Genome research*, 2012. 22(6): p. 1139-43; Picelli, S., y col., Full-length RNA-seq from single cells using Smart-seq2. *Nature protocols*, 2014. 9(1): págs. 171-81 y Buenrostro, J.D., y col., Transposition of native chromatin for fast and sensitive epigenomic profiling of open cromatina, DNA-binding proteins and nucleosome position. *Nature methods*, 2013, cada uno de los cuales se incorpora en la presente memoria como referencia en su totalidad para todos los fines. Véanse también las patentes WO 98/10077, EP 2527438 y EP 2376517, cada una de las cuales se incorpora en la presente memoria como referencia en su totalidad. Un kit de transposición disponible en el mercado se comercializa con el nombre de NEXTERA y está disponible en Illumina.

#### Rellenado de espacios

Los fragmentos bicatenarios producidos por los métodos de transposomas descritos en la presente memoria se procesan a continuación para llenar los espacios. Según un aspecto, el transposón puede incluir una o más citosinas metiladas, tales como adaptadores o cebadores de citosina metilada, tal como se describe en la presente memoria. Para los sistemas que utilizan adaptadores de citosina metilados para realizar la transposición de Tn5, la etapa de relleno de espacios incluye el uso de dCTP metilado en lugar de dCTP en la mezcla de dNTP para crear adaptadores bicatenarios completamente metilados.

#### Adn portador y purificación opcional

Según ciertos aspectos, los fragmentos llenos de espacios se combinan después con el ADN portador. El ADN portador puede ser cualquier fragmento de ADNbc que tenga una longitud de entre 100 pares de bases (pb) y pares de bases de 4 kilos. Según un aspecto, el ADN portador puede ser un tipo de ADN que es diferente del ADN diana. Según un aspecto, el ADN portador puede ser un tipo de ADN que es el mismo que el ADN diana. Según un aspecto,

el ADN portador es ADN lambda sonificado. Según un aspecto, el ADN portador no incluye los adaptadores de secuenciación de Illumina.

El ADN portador sirve para proteger el ADN objetivo de las duras condiciones del tratamiento químico y para reducir el daño al ADN objetivo o la pérdida del ADN diana. Con respecto al daño del ADN, la conversión ilustrativa sin bisulfito utiliza un reactivo de oxidación fuerte Fe (II), que provocará daños en el ADN al producir radicales hidroxilo. El ADN portador, que se añade al medio de reacción en una cantidad que es de 100 a 10.000 veces (p. ej., de 100 a 1000 veces) mayor que la del ADN de la muestra, sirve para ocupar el exceso de radicales hidroxilo y prevenir o limitar su interacción con el ADN diana. Según un aspecto ilustrativo, para una muestra de 6 pg de ADN de una única célula, se proporcionan 20 ng de ADN portador lambda sonificado.

En cuanto a la pérdida de ADN, la reacción de conversión utiliza un regulador especializado y purificaciones (intercambio de regulador mientras se conserva el ADN) antes o incluso entre los diferentes pasos. Además, puede ser necesaria una etapa de purificación para eliminar los reactivos químicos antes de la amplificación por PCR. Se pueden llevar a cabo dos o tres etapas de purificación como resultado del proceso de conversión de citosina. Cada paso de purificación del ADN puede provocar una pérdida de ADN del 50 al 90 % si la cantidad de ADN de entrada es baja, como el ADN de una única célula. Pero si se añade ADN portador a la muestra de ADN, como 20 ng de ADN lambda a 6 pg de ADN de muestra, la purificación del ADN solo provocará una pérdida del 10 %, lo que resultará en la preservación del 90 % del ADN de la muestra aunque esté mezclado con el ADN portador. Según la presente descripción, la eficiencia de la purificación del ADN aumenta a medida que aumenta la cantidad de ADN de entrada.

Según ciertos aspectos, el medio de reacción que incluye los segmentos bicatenarios llenos de espacios y el ADN portador puede purificarse mediante la purificación del ADN en columnas giratorias o mediante la purificación del ADN basada en glóbulos, u otros métodos de purificación conocidos por los expertos en la técnica, antes de la conversión del ADN requerida para la detección de la metilación. Alternativamente, el medio de reacción puede proceder directamente a la conversión química.

#### Tratamiento químico

Los fragmentos llenados combinados con el ADN portador se mezclan con reactivos químicos que transforman químicamente una citosina en un uracilo. Tales reactivos químicos se describen en la patente US 2013/0244237 y pueden estar disponibles en New England Biolabs. Otros reactivos enzimáticos resultarán evidentes para los expertos en la técnica basándose en la presente descripción. Según un aspecto, el reactivo no es bisulfito o excluye el bisulfito o el reactivo convierte la citosina en uracilo con la condición de que el reactivo no sea bisulfito.

#### Purificación opcional

Según ciertos aspectos, el medio de reacción que incluye los fragmentos convertidos químicamente puede purificarse mediante purificación de ADN en columnas giratorias o glóbulos, u otros métodos de purificación conocidos por los expertos en la técnica, antes de la amplificación. Alternativamente, el medio de reacción puede proceder directamente a la amplificación.

#### Amplificación

Según un aspecto, la amplificación de solo los fragmentos diana se lleva a cabo mediante el uso de cebadores que se dirigen a las secuencias adaptadoras incorporadas en los fragmentos por los transposomas. El ADN portador no se amplifica y se convierte en ADNmc debido a la desnaturalización.

La expresión “amplificación” o “amplificar” se refiere a un proceso mediante el cual se forman copias adicionales o múltiples de un polinucleótido particular. El ADN a amplificar se puede obtener de una única célula o de una pequeña población de células. Los métodos descritos en la presente memoria permiten amplificar el ADN de cualquier especie u organismo en una mezcla de reacción, tal como una mezcla de reacción única llevada a cabo en un único recipiente de reacción. En un aspecto, los métodos descritos en la presente memoria incluyen la amplificación independiente de la secuencia del ADN de cualquier fuente, incluyendo, pero sin limitarse a, ADN humano, animal, vegetal, de levadura, viral, eucariótico y procariótico.

Las plantillas de fragmentos de ADN preparadas utilizando los métodos de transposasa descritos en la presente memoria pueden amplificarse en microgotas utilizando métodos conocidos por los expertos en la técnica. Las microgotas se pueden formar como una emulsión de una fase oleosa y una fase acuosa. Una emulsión puede incluir gotitas acuosas o volúmenes acuosos aislados dentro de una fase oleosa continua. Los métodos de amplificación del genoma completo en emulsión se describen utilizando gotitas acuosas de pequeño volumen en aceite para aislar cada fragmento para la amplificación uniforme del genoma de una única célula. Distribuyendo cada fragmento en su propia gota o volumen de reacción acuoso aislado, se permite que cada gota alcance la saturación de la amplificación del ADN. Los amplicones dentro de cada gotita se fusionan después mediante demulsificación, lo que da como resultado una amplificación uniforme de todos los fragmentos del genoma completo de la célula individual.

En ciertos aspectos, la amplificación se logra mediante PCR. La PCR es una reacción en la que se hacen copias replicadas de un polinucleótido diana utilizando un par de cebadores o un conjunto de cebadores que consiste en un cebador aguas arriba y uno aguas abajo, y un catalizador de polimerización, tal como una ADN polimerasa, y normalmente una enzima polimerasa térmicamente estable. Los métodos para la PCR son bien conocidos en la técnica y se enseñan, por ejemplo, en MacPherson y col. (1991) PCR 1: A Practical Approach (IRL Press at Oxford University Press). El término “reacción en cadena de la polimerasa” (“PCR”, por sus siglas en inglés) de Mullis (patentes n.º US-4.683.195, US-4.683.202 y US-4.965.188) se refiere a un método para aumentar la concentración de un segmento de una secuencia diana sin clonación o purificación. Este proceso para amplificar la secuencia diana incluye proporcionar a los cebadores oligonucleotídicos la secuencia diana y los reactivos de amplificación deseados, seguidos de una secuencia precisa de ciclos térmicos en presencia de una polimerasa (p. ej., la ADN polimerasa). Los cebadores son complementarios a sus respectivas cadenas (“secuencias de unión a cebadores”) de la secuencia diana bicatenaria. Para efectuar la amplificación, la secuencia diana bicatenaria se desnatura y los cebadores se hibridan a continuación con sus secuencias complementarias dentro de la molécula diana. Tras la hibridación, los cebadores se extienden con una polimerasa para formar un nuevo par de cadenas complementarias. Las etapas de desnaturalización, hibridación con cebadores y extensión de la polimerasa pueden repetirse muchas veces (es decir, la desnaturalización, la hibridación y la extensión constituyen un “ciclo”; puede haber numerosos “ciclos”) para obtener una alta concentración de un segmento amplificado de la secuencia diana deseada.” La longitud del segmento amplificado de la secuencia diana deseada está determinada por las posiciones relativas de los cebadores entre sí y, por lo tanto, esta longitud es un parámetro controlable. En virtud del aspecto repetitivo del proceso, el método se denomina “reacción en cadena de la polimerasa” (en lo sucesivo, “PCR”) y se dice que la secuencia diana está “amplificada por PCR”.

Con la PCR, es posible amplificar una única copia de una secuencia diana específica en el ADN genómico hasta un nivel detectable mediante varias metodologías diferentes (p. ej., hibridación con una sonda marcada; incorporación de cebadores biotinilados seguida de detección de conjugados avidina-enzima; incorporación de trifosfatos desoxinucleotídicos marcados con <sup>32</sup>P, tales como dCTP o dATP, en el segmento amplificado). Además del ADN genómico, cualquier secuencia de oligonucleótidos o polinucleótidos se puede amplificar con el conjunto apropiado de moléculas cebadoras. En particular, los segmentos amplificados creados por el propio proceso de PCR dentro de cada microgota son, en sí mismos, plantillas eficaces para las amplificaciones por PCR posteriores. Los métodos y kits para realizar la PCR son bien conocidos en la técnica. Todos los procesos de producción de copias replicadas de un polinucleótido, tales como la PCR o la clonación de genes, se denominan colectivamente en la presente memoria replicación. Un cebador también se puede utilizar como sonda en reacciones de hibridación, como los análisis de transferencia Southern o Northern.

La expresión “amplificación” o “amplificar” se refiere a un proceso mediante el cual se forman copias adicionales o múltiples de un polinucleótido particular. La amplificación incluye métodos tales como la PCR, la amplificación por ligación (o reacción en cadena de la ligasa, LCR, por sus siglas en inglés) y otros métodos de amplificación. Estos métodos son conocidos y se practican ampliamente en la técnica. Véanse, p. ej., las patentes n.º US-4.683.195 y US-4.683.202 e Innis y col., “PCR protocols: a guide to method and applications” Academic Press, Incorporated (1990) (para PCR); y Wu y col. (1989) Genomics 4:560-569 (para LCR). En general, el procedimiento de PCR describe un método de amplificación génica que comprende (i) la hibridación de cebadores de secuencia específica con genes específicos dentro de una muestra (o genoteca) de ADN, (ii) la amplificación posterior que implica múltiples rondas de hibridación, alargamiento y desnaturalización utilizando una ADN polimerasa, y (iii) la selección de los productos de PCR para detectar una banda del tamaño correcto. Los cebadores utilizados son oligonucleótidos de longitud suficiente y secuencia apropiada para proporcionar el inicio de la polimerización, es decir, cada cebador está diseñado específicamente para ser complementario a cada cadena del locus genómico que se va a amplificar.

Los reactivos y el hardware para llevar a cabo reacciones de amplificación están disponibles en el mercado. Los cebadores útiles para amplificar secuencias de una región génica particular son preferiblemente complementarios y se hibridan específicamente con secuencias en la región diana o en sus regiones flanqueantes y se pueden preparar utilizando métodos conocidos por los expertos en la técnica. Las secuencias de ácidos nucleicos generadas por amplificación pueden secuenciarse directamente.

Cuando la hibridación se produce en una configuración antiparalela entre dos polinucleótidos monocatenarios, la reacción se denomina “hibridación” y esos polinucleótidos se describen como “complementarios”. Un polinucleótido bicatenario puede ser complementario u homólogo a otro polinucleótido, si la hibridación puede producirse entre una de las cadenas del primer polinucleótido y del segundo. La complementariedad u homología (el grado en que un polinucleótido es complementario con otro) es cuantificable en términos de la proporción de bases en cadenas opuestas que se espera que formen enlaces de hidrógeno entre sí, según las reglas de apareamiento de bases generalmente aceptadas.

Los términos “producto de PCR”, “fragmento de PCR” y “producto de amplificación” se refieren a la mezcla resultante de compuestos después de completar dos o más ciclos de las etapas de desnaturalización, hibridación y extensión de la PCR. Estos términos abarcan el caso donde ha habido amplificación de uno o más segmentos de una o más secuencias diana.

El término “reactivos de amplificación” puede referirse a aquellos reactivos (trifosfatos de desoxirribonucleótidos, regulador, etc.) necesarios para la amplificación, excepto los cebadores, la plantilla de ácido nucleico y la enzima de amplificación. Normalmente, los reactivos de amplificación junto con otros componentes de reacción se colocan y se contienen en un recipiente de reacción (tubo de ensayo, micropocillo, etc.). Los métodos de amplificación incluyen los métodos de PCR conocidos por los expertos en la técnica y también incluyen la amplificación en círculo rodante (Blanco y col., J. Biol. Chem., 264, 8935-8940, 1989), amplificación por círculo rodante hiperramificado (Lizard y col., Nat. Genetics, 19, 225-232, 1998) y la amplificación isotérmica mediada por bucles (Notomi y col., Nuc. Acids Res., 28, e63, 2000), cada uno de los cuales se incorpora en la presente memoria como referencia en su totalidad.

Para la PCR en emulsión, se crea una reacción de PCR en emulsión agitando o agitando vigorosamente una mezcla de “agua en aceite” para generar millones de compartimentos acuosos de tamaño micrométrico. Los chips microfluídicos pueden estar equipados con un dispositivo para crear una emulsión agitando o agitando una fase oleosa y una fase acuosa. Alternativamente, las gotas acuosas pueden formarse espontáneamente combinando un determinado aceite con una fase acuosa o introduciendo una fase acuosa en una fase oleosa. La genoteca de ADN a amplificar se mezcla en una dilución limitante antes de la emulsificación. La combinación del tamaño del compartimento, es decir, el tamaño de las microgotas, y la cantidad de microgotas creadas que limitan la dilución de la genoteca de fragmentos de ADN a amplificar se utiliza para generar compartimentos que contienen, en promedio, solo una molécula de ADN. Dependiendo del tamaño de los compartimentos acuosos generados durante la etapa de formación de microgotas o emulsificación, se pueden realizar hasta  $3 \times 10^9$  reacciones de PCR individuales por  $\mu\text{l}$  simultáneamente en el mismo tubo. Esencialmente, cada microgota pequeña del compartimento acuoso de la emulsión forma un microrreactor de PCR. El tamaño promedio de un compartimento en una emulsión varía desde un diámetro inferior a la micras hasta más de 100 micras, o de 1 picolitro a 1000 picolitros o de 1 nanolitro a 1000 nanolitros o de 1 picolitro a 1 nanolitro o de 1 picolitro a 1000 nanolitros, dependiendo de las condiciones de emulsificación.

Otros métodos de amplificación, tal como se describen en la solicitud de patente británica n.º GB 2.202.328 y en la solicitud de patente de PCT n.º PCT/US89/01025, cada una incorporada en la presente memoria como referencia, pueden utilizarse según la presente descripción. En la primera aplicación, los cebadores “modificados” se utilizan en una plantilla similar a la PCR y en una síntesis dependiente de enzimas. Los cebadores pueden modificarse marcando con un resto de captura (p. ej., biotina) y/o un resto detector (p. ej., una enzima). En esta última aplicación, se añade un exceso de sondas marcadas a una muestra. En presencia de la secuencia diana, la sonda se une y se escinde catalíticamente. Tras la escisión, la secuencia diana se libera intacta para unirse al exceso de sonda. La escisión de la sonda marcada señala la presencia de la secuencia diana.

Otros métodos de amplificación adecuados incluyen la “PCR race” y la “PCR unilateral”. (Frohman, en: PCR Protocols: A Guide To Methods And Applications, Academic Press, N.Y., 1990, incorporado en la presente memoria como referencia). Los métodos basados en la ligación de dos (o más) oligonucleótidos en presencia de un ácido nucleico que tiene la secuencia del “dioligonucleótido” resultante, amplificando así el dioligonucleótido, también se pueden utilizar para amplificar el ADN según la presente descripción (Wu y col., Genomics 4:560-569, 1989, incorporado en la presente memoria como referencia).

Como se usa en la presente memoria, el término “cebador” generalmente incluye un oligonucleótido, ya sea natural o sintético, que es capaz, tras formar un dúplex con una plantilla de polinucleótido, de actuar como un punto de inicio de la síntesis de ácidos nucleicos, tal como un cebador de secuenciación, y extenderse desde su extremo 3' a lo largo de la plantilla de modo que se forme un dúplex extendido. La secuencia de nucleótidos añadidos durante el proceso de extensión está determinada por la secuencia del polinucleótido molde. Por lo general, los cebadores se extienden mediante una ADN polimerasa. Los cebadores suelen tener una longitud en el intervalo de entre 3 y 36 nucleótidos, también de 5 a 24 nucleótidos, también de 14 a 36 nucleótidos. Los cebadores dentro del alcance de la invención incluyen cebadores ortogonales, cebadores de amplificación, cebadores de construcción y similares. Los pares de cebadores pueden flanquear una secuencia de interés o un conjunto de secuencias de interés. Los cebadores y las sondas pueden tener una secuencia degenerada o cuasidegenerada. Los cebadores dentro del alcance de la presente invención se unen de forma adyacente a una secuencia diana. Un “cebador” puede considerarse un polinucleótido corto, generalmente con un grupo 3' -OH libre que se une a una diana o plantilla potencialmente presente en una muestra de interés al hibridarse con la diana y, posteriormente, promover la polimerización de un polinucleótido complementario a la diana. Los cebadores de la presente invención están compuestos por nucleótidos que varían de 17 a 30 nucleótidos. En un aspecto, el cebador tiene al menos 17 nucleótidos, o alternativamente, al menos 18 nucleótidos, o alternativamente, al menos 19 nucleótidos, o alternativamente, al menos 20 nucleótidos, o alternativamente, al menos 21 nucleótidos, o alternativamente, al menos 22 nucleótidos, o alternativamente, al menos 23 nucleótidos, o alternativamente, al menos 24 nucleótidos, o alternativamente, al menos 25 nucleótidos, o alternativamente, al menos 26 nucleótidos, o alternativamente, al menos 27 nucleótidos, o alternativamente, al menos 28 nucleótidos, o alternativamente, al menos 29 nucleótidos, o alternativamente, al menos 30 nucleótidos, o alternativamente al menos 50 nucleótidos, o alternativamente al menos 75 nucleótidos o alternativamente al menos 100 nucleótidos.

#### Purificación

Según ciertos aspectos, el medio de reacción que incluye los fragmentos amplificados se puede purificar mediante purificación de ADN basada en columnas giratorias o glóbulos, u otros métodos de purificación conocidos por los



expertos en la técnica, antes de la secuenciación. La purificación del ADN seguida de la amplificación, por ejemplo mediante una reacción de PCR, eliminará la mayor parte del ADN portador monocatenario, lo que da como resultado una genoteca de ADN diana amplificado puro lista para la secuenciación.

## 5 Secuenciación

El ADN amplificado según los métodos descritos en la presente memoria puede secuenciarse y analizarse utilizando métodos conocidos por los expertos en la técnica. La determinación de la secuencia de una secuencia de ácido nucleico de interés se puede realizar utilizando una variedad de métodos de secuenciación conocidos en la técnica que incluyen, pero no se limitan a, la secuenciación por hibridación (SBH), la secuenciación por ligación (SBL) (Shendure y col. (2005) *Science* 309:1728), secuenciación cuantitativa incremental por adición de nucleótidos fluorescentes (QIFNAS), ligadura y escisión escalonadas, transferencia de energía por resonancia fluorescente (FRET), balizas moleculares, digestión con sonda reportera TaqMan, pirosecuenciación, secuenciación fluorescente *in situ* (FISSEQ), glóbulos FISSEQ (patente n.º US-7.425.431), secuenciación oscilante (PCT/US05/27695), secuenciación múltiple (serie n.º US-12/027.039, presentada el 6 de febrero de 2008; Porreca y col. (2007) *Nat. Métodos* 4:931), secuenciación de colonias polimerizadas (POLONY) (patentes n.º US-6.432.360, US-6.485.944 y US-6.511.803, y PCT/US05/06425); secuenciación circular de nanomallas (ROLONY) (serie n.º US-12/120.541, presentada el 14 de mayo de 2008), ensayos de ligación de oligo específicos de alelos (p. ej., ensayo de ligación de oligos [OLA], molécula plantilla única OLA que utiliza una sonda lineal ligada y una lectura de amplificación por círculo rodante [RCA], sondas de candado ligadas y/o molécula plantilla única OLA utilizando una sonda de candado circular ligada y una lectura de amplificación por círculo rodante [RCA]) y similares. También se pueden utilizar métodos de secuenciación de alto rendimiento, p. ej., utilizando plataformas tales como las plataformas Roche 454, Illumina Solexa, AB-SOLiD, Helicos, Polonator y similares. En la técnica se conocen diversas tecnologías de secuenciación basadas en la luz (Landegren y col. (1998) *Genome Res.* 8:769-76; Kwok (2000) *Pharmacogenomics* 1:95-100; y Shi (2001) *Clin. Chem.* 47:164-172).

El ADN amplificado se puede secuenciar mediante cualquier método adecuado. En particular, el ADN amplificado se puede secuenciar utilizando un método de cribado de alto rendimiento, como la tecnología de secuenciación SOLiD de Applied Biosystems o el analizador del genoma de Illumina. En un aspecto de la invención, el ADN amplificado puede secuenciarse de forma automática. El número de lecturas puede ser de al menos 10.000, al menos 1 millón, al menos 10 millones, al menos 100 millones o al menos 1000 millones. En otro aspecto, el número de lecturas puede ser de 10.000 a 100.000, o alternativamente de 100.000 a 1 millón, o alternativamente de 1 millón a 10 millones, o alternativamente de 10 millones a 100 millones, o alternativamente de 100 millones a 1000 millones. Una "lectura" es una longitud de secuencia continua de ácidos nucleicos obtenida mediante una reacción de secuenciación.

La "secuenciación con escopeta" se refiere a un método utilizado para secuenciar una gran cantidad de ADN (como el genoma completo). En este método, el ADN que se va a secuenciar se tritura primero en fragmentos más pequeños que se pueden secuenciar individualmente. A continuación, las secuencias de estos fragmentos se vuelven a ensamblar en su orden original basándose en sus secuencias superpuestas, produciendo así una secuencia completa. La "trituration" del ADN se puede realizar utilizando varias técnicas diferentes, incluida la digestión con enzimas de restricción o el cizallamiento mecánico. Las secuencias superpuestas se alinean normalmente mediante un ordenador programado adecuadamente. Los métodos y programas para la secuenciación automática de una genoteca de ADNc son bien conocidos en la técnica.

Los métodos de amplificación y secuenciación son útiles en el campo de la medicina predictiva, en la que los ensayos de diagnóstico, los ensayos de pronóstico, la farmacogenómica y los ensayos clínicos de monitorización se utilizan con fines pronósticos (predictivos) para tratar así a un individuo de forma profiláctica. En consecuencia, un aspecto de la presente invención se refiere a ensayos de diagnóstico para determinar el ADN genómico con el fin de determinar si un individuo está en riesgo de desarrollar un trastorno y/o enfermedad. Tales ensayos se pueden utilizar con fines pronósticos o predictivos para tratar profilácticamente a un individuo antes del inicio del trastorno y/o enfermedad. En consecuencia, en ciertas realizaciones ilustrativas, se proporcionan métodos para diagnosticar y/o pronosticar una o más enfermedades y/o trastornos utilizando uno o más de los métodos de elaboración de perfiles de expresión descritos en la presente memoria.

## 55 Realizaciones electrónicas

En ciertas realizaciones ilustrativas, se proporcionan medios legibles por aparatos electrónicos que comprenden una o más secuencias de ADN genómico descritas en la presente memoria. Como se usa en la presente memoria, "medio legible por aparato electrónico" se refiere a cualquier medio adecuado para almacenar, retener o contener datos o información que pueda leerse y accederse directamente mediante un aparato electrónico. Tales medios pueden incluir, pero no se limitan a: medios de almacenamiento magnéticos, tales como disquetes, medios de almacenamiento en disco duro y cinta magnética; medios de almacenamiento óptico, como discos compactos; medios de almacenamiento electrónico tales como RAM, ROM, EPROM, EEPROM y similares; discos duros generales e híbridos de estas categorías, como los medios de almacenamiento magnético/óptico. El medio está adaptado o configurado para registrar en él uno o más perfiles de expresión descritos en la presente memoria.

Como se usa en la presente memoria, el término “aparato electrónico” pretende incluir cualquier aparato informático o de procesamiento adecuado u otro dispositivo configurado o adaptado para almacenar datos o información. Los ejemplos de aparatos electrónicos adecuados para su uso con la presente invención incluyen aparatos informáticos independientes; redes, incluidas una red de área local (LAN), una red de área amplia (WAN) Internet, Intranet y Extranet; aparatos electrónicos tales como asistentes digitales personales (PDA, por sus siglas en inglés), teléfonos móviles, buscapersonas y similares; y sistemas de procesamiento locales y distribuidos.

Como se usa en la presente memoria, “grabado” se refiere a un proceso para almacenar o codificar información en el medio legible del aparato electrónico. Los experimentados en la técnica pueden adoptar fácilmente cualquiera de los métodos actualmente conocidos para registrar información en medios conocidos para generar productos que comprendan uno o más perfiles de expresión descritos en la presente memoria.

Se pueden utilizar una variedad de programas y formatos de software para almacenar la información del ADN genómico de la presente invención en el medio legible por aparatos electrónicos. Por ejemplo, la secuencia de ácido nucleico puede representarse en un archivo de texto de procesamiento de texto, formatearse en software disponible comercialmente, como WordPerfect y MicroSoft Word, o representarse en forma de un archivo ASCII, almacenado en una aplicación de base de datos, como DB2, Sybase, Oracle o similares, así como en otras formas. Se puede emplear cualquier número de formatos de estructuración de procesadores de datos (p. ej., archivo de texto o base de datos) para obtener o crear un medio que tenga registrados en el mismo uno o más perfiles de expresión descritos en la presente memoria.

Debe entenderse que las realizaciones de la presente invención que se han descrito son meramente ilustrativas de algunas de las aplicaciones de los principios de la presente invención. Los experimentados en la técnica pueden realizar numerosas modificaciones basándose en las enseñanzas presentadas en la presente memoria sin apartarse del verdadero espíritu y alcance de la invención. El contenido de todas las referencias, patentes y solicitudes de patente publicadas citadas a lo largo de esta solicitud se incorpora aquí como referencia en su totalidad para todos los fines.

Los siguientes ejemplos se exponen como representativos de la presente invención. Estos ejemplos no deben interpretarse como limitativos del alcance de la invención, ya que estas y otras realizaciones equivalentes resultarán evidentes a la vista de la presente descripción, las figuras y las reivindicaciones adjuntas.

#### Ejemplo i

Según ciertos aspectos ilustrativos, se utiliza un sistema de transposición para fabricar fragmentos de ácido nucleico para el tratamiento de metilación química, incluido el ADN portador, la amplificación y la secuenciación, según se desee. Según un aspecto, se utiliza un sistema de transposición para fragmentar el ADN genómico en fragmentos de ADN genómico bicatenario, teniendo el ADN de transposición diferentes secuencias de sitios de cebado insertadas en el mismo. Como se ilustra en la figura 1, un ADN de transposición incluye un sitio de unión a la transposasa bicatenaria y una secuencia de sitios de cebado única y diferente M. Aunque no se ilustra en la figura 1, el ADN de transposición puede incluir una o más 5-metilcitosinas. El sitio de unión a la transposasa bicatenaria puede ser un sitio de unión a la transposasa (Tnp) Tn5 bicatenario de 19 pares de bases que está unido o conectado, tal como mediante un enlace covalente, a un saliente monocatenario que incluye una secuencia del sitio de cebado, tal como en un extremo del saliente. El ADN de transposición se inserta en el ADN genómico de una única célula mientras se crean fragmentos utilizando una transposasa. Tras eliminar la transposasa y rellenar los espacios, los fragmentos de ADN genómico que tienen secuencias de sitios de cebado diferentes, distintas o únicas en cada extremo del fragmento se amplifican utilizando cebadores junto con una ADN polimerasa, nucleótidos y reactivos de amplificación para amplificar por PCR todo el genoma de la célula única.

Según ciertos aspectos, cuando se amplifican cantidades bajas o pequeñas de ADN, como el ADN de unas pocas células, es decir, de 2 a 5 o de 2 a 10 células o de 2 a 100 o de una única célula, no se lleva a cabo una etapa de purificación de la columna de ADN para maximizar la pequeña cantidad (~6 pg) de ADN genómico que se puede obtener del interior de una única célula antes de la amplificación. El ADN se puede amplificar directamente a partir de un lisado celular u otra afección impura. En consecuencia, la muestra de ADN puede ser impura, no purificada o no aislada. En consecuencia, los aspectos del presente método permiten maximizar el ADN genómico para la amplificación y reducir la pérdida debida a los fragmentos que tienen la misma secuencia del sitio de cebado en cada extremo que con otros métodos, es decir, métodos no múltiples. Según un aspecto adicional, los métodos descritos en la presente memoria pueden utilizar métodos de amplificación distintos de la PCR.

Según un aspecto y como se ilustra en general en la figura 2, la transposasa (Tnp, círculos grises) y el ADN de transposición, cada uno con secuencias de sitio de cebado únicas y diferentes ilustradas por secuencias salientes de diferentes patrones, se combinan para formar una pluralidad de transposomas. Cada transposoma tiene dos secuencias de sitios de cebado diferentes y únicas. Cada transposoma tiene dos secuencias de sitios de cebado diferentes y únicas en comparación con cada otro transposoma dentro de la pluralidad. Para hacer un grupo de mezclas de transposones de 20 secuencias de transposones, se mezclan secuencias molares iguales de cada tipo de transposones en un regulador que contiene Tris 10 mM, pH 8, NaCl 50 mM y EDTA 1 mM. Para ensamblar los

complejos de transposomas, el grupo de 20 transposones se mezcla con la transposasa Tn5 en una proporción molar igual y se incuba a temperatura ambiente durante 30 minutos.

Como se muestra en la figura 3A, los transposomas de la genoteca de transposomas capturan aleatoriamente o se unen de otro modo al ADN genómico unicelular diana como dímeros. Los transposomas representativos están numerados 1, 2 y 3, aunque el número de miembros del transposoma puede ser mayor dependiendo de la aplicación deseada. Un número representativo de transposones que tienen secuencias de sitios de unión al cebador diferentes y/o únicas está entre 5 y 50. Cada transposoma incluye dos secuencias de sitios de cebado únicas y/o diferentes. Por ejemplo, el transposoma 1 incluye dos secuencias de sitios de cebado únicas y/o diferentes, el transposoma 2 incluye dos secuencias de sitios de cebado únicas y/o diferentes, el transposoma 3 incluye dos secuencias de sitios de cebado únicas y/o diferentes, etc. La secuencia de sitio de cebado única y/o diferente está dentro de cada ADN de transposón del transposoma. Las transposasas en el transposoma cortan el ADN genómico con una transposasa que corta una cadena superior y una transposasa que corta una cadena inferior para crear un fragmento de ADN genómico. Aunque no se muestra en la figura 3A, el fragmento puede incluir una o más citosinas o una o más 5-metilcitosinas. La pluralidad de transposomas crea una pluralidad de fragmentos de ADN genómico, con uno o más fragmentos que incluyen una o más citosinas o una o más 5-metilcitosinas.

De este modo, un ADN de transposón del dímero de ADN de transposón se une a cada extremo del sitio de corte o sitio de fragmentación, es decir, un ADN de transposón del transposoma 1 se une al sitio de corte de la izquierda y el otro ADN de transposón del transposoma 1 se une al sitio de corte de la derecha. Dado que la genoteca de transposomas corta el ácido nucleico en fragmentos, cada fragmento tendrá una secuencia de sitio de cebado diferente en cada extremo del fragmento. Esto está representado por los dos fragmentos ilustrativos donde el fragmento superior tiene una secuencia de sitios de cebado 1 única y diferente en un extremo y una secuencia de sitios de cebado 2 única y diferente en el otro extremo. Del mismo modo, el fragmento inferior tiene una secuencia de sitio de cebado única y diferente 2 en un extremo y una secuencia de sitio de cebado única y diferente 3 en el otro extremo. Como se ilustra, el sitio de corte entre los dos fragmentos es producido por el transposoma 2 y el sitio de corte a la izquierda (es decir, ver el lado derecho del fragmento superior en la figura 3) incluye un transposón con una secuencia de sitios de cebado 2 única y diferente, mientras que el sitio de corte a la derecha (es decir, ver el lado izquierdo del fragmento inferior en la figura 3) incluye la secuencia de sitios de cebado 2 única y diferente (donde "2" se refiere al transposoma 2). Según un aspecto, se añade un transposoma 100 nM al lisado celular y la mezcla de reacción de transposición se incuba a 55 grados durante 10 minutos con una concentración final de magnesio de 5 mM. Tras eliminar la transposasa, el ADN genómico se corta en millones de pequeños fragmentos de ADN, cada uno etiquetado con una de las 20 secuencias de transposones en cada extremo. (Figura 3A) De esta manera, la genoteca de transposomas puede incluir 20 secuencias de sitios de unión a cebadores diferentes y/o únicas, tal como se describe en la presente memoria, mientras que los miembros de la genoteca de transposomas pueden acercarse a millones de miembros.

Como se muestra en la figura 3B, los transposomas de la genoteca de transposomas capturan aleatoriamente o se unen de otro modo al ADN genómico unicelular diana como dímeros. Los transposomas representativos están numerados 1, 2 y 3, aunque el número de miembros del transposoma puede ser mayor dependiendo de la aplicación deseada. Un número representativo de transposones que tienen secuencias de sitios de unión al cebador diferentes y/o únicas está entre 5 y 50. Cada transposoma incluye la misma secuencia única y/o diferente del sitio de unión al cebador en cada transposón del transposoma. Por ejemplo, el transposoma 1 incluye la misma secuencia de sitios de unión al cebador en cada transposón, el transposoma 2 incluye la misma secuencia de sitios de unión al cebador en cada transposón, el transposoma 3 incluye la misma secuencia de sitios de unión al cebador en cada transposón, etc. Sin embargo, cada transposoma tiene un sitio de unión al cebador único y diferente asociado al mismo, de tal modo que cada transposoma tiene un sitio de unión al cebador diferente asociado al mismo en comparación con otros miembros del genoteca de transposomas. Las transposasas en el transposoma cortan el ADN genómico con una transposasa que corta una cadena superior y una transposasa que corta una cadena inferior para crear un fragmento de ADN genómico. La pluralidad de transposomas crea una pluralidad de fragmentos de ADN genómico con uno o más fragmentos que incluyen una o más citosinas o una o más 5-metilcitosinas para el tratamiento químico para cambiar las citosinas en uracilos. De este modo, un ADN de transposón del dímero de ADN de transposón se une a cada extremo del sitio de corte o sitio de fragmentación, es decir, un ADN de transposón del transposoma 1 se une al sitio de corte de la izquierda y el otro ADN de transposón del transposoma 1 se une al sitio de corte de la derecha. Dado que la genoteca de transposomas corta el ácido nucleico en fragmentos, cada fragmento tendrá una secuencia de sitio de cebado diferente en cada extremo del fragmento, ya que los transposomas adyacentes unidos al ácido nucleico que crean el fragmento tienen cada uno diferentes secuencias de sitios de unión al cebador. Esto está representado por los dos fragmentos ilustrativos donde el fragmento superior tiene una secuencia de sitios de cebado 1 única y diferente en un extremo y una secuencia de sitios de cebado 2 única y diferente en el otro extremo. Del mismo modo, el fragmento inferior tiene una secuencia de sitios de cebado 2 única y diferente en un extremo (que es la misma secuencia de sitios de unión al cebador que en el extremo derecho del fragmento superior) y una secuencia de sitios de cebado 3 única y diferente en el otro extremo. Como se ilustra, el sitio de corte entre los dos fragmentos es producido por el transposoma 2 y el sitio de corte a la izquierda (es decir, ver el lado derecho del fragmento superior en la figura 3) incluye un transposón con una secuencia de sitios de cebado 2 única y diferente, mientras que el sitio de corte a la derecha (es decir, ver el lado izquierdo del fragmento inferior en la figura 3) incluye la secuencia de sitios de cebado 2 única y diferente (donde "2" se refiere al transposoma 2). En consecuencia, incluso cuando el

transposoma tiene la misma secuencia del sitio de unión al cebador en cada transposón, el método da como resultado un fragmento que tiene diferentes secuencias del sitio de unión al cebador en cada extremo del fragmento.

Como se ilustra en la figura 4, la fragmentación del ADN genómico y la inserción de los transposones de metilo dejan un espacio en ambos extremos del sitio de transposición/inserción. El espacio puede tener cualquier longitud, pero un espacio de 9 bases es ilustrativo. El resultado es un fragmento de ADN genómico con un sitio de unión a Tnp del ADN de transposón unido a la posición 5' de una cadena superior y un sitio de unión a Tnp del ADN de transposón unido a la posición 5' de una cadena inferior. Se muestran las brechas resultantes de la unión o inserción del ADN de transposón. Tras la transposición, se retira la transposasa y se realiza la extensión del espacio para rellenar el espacio y complementar el saliente monocatenario diseñado originalmente en el ADN de transposón, tal como se muestra en la figura 4. Posteriormente, los fragmentos llenados se someten a un tratamiento químico para convertir la citosina en uracilo en presencia de ADN portador. Estos fragmentos procesados químicamente pueden después someterse a amplificación y purificación como se describe en la presente memoria.

Según un aspecto, después se añade una mezcla de reacción de ADN polimerasa que contiene 200 uM de dNTP, 1 regulador de reacción NEB Q5, 125 nM cada uno de los 20 cebadores y 0,02 U/ul de ADN polimerasa Q5 y se incuba a 72 °C durante 3 minutos para llenar el espacio dejado por la transposición (figura 4). Como se ilustra adicionalmente en la figura 5, los fragmentos mostrados en la figura 4 se someten a amplificación por PCR múltiple para producir amplicones. Se realizan 15 ciclos de reacciones de PCR como: 98 °C 30s, 65 °C 1 min, 72 °C 2 min para amplificar el ADN genómico diana. Los productos de amplificación se purifican después mediante una columna de purificación de ADN Zymo.

Ejemplo ii

Protocolo general

Se lisa una única célula en un regulador de lisis. La transposición de Tn5 se lleva a cabo utilizando una genoteca de transposomas que incluye transposomas, cada uno con una secuencia de sitios de unión a cebadores diferente y única (o cada uno con dos secuencias de sitios de unión a cebadores diferentes y únicas), tal como se describe en la presente memoria, e incluye transposones de metilo y regulador de transposición, que se mezcla bien y se incuba a 55 °C durante 10 minutos. Se añade 1 mg/ml de proteasa después de la transposición para evitar que la transposasa se una al ADN genómico de una única célula. La ADN polimerasa Q5, el dNTP, el regulador de reacción de PCR y los cebadores se añaden a la mezcla de reacción que se calienta a 72 °C durante 10 minutos para llenar el espacio generado por la inserción del transposón. Se añaden ADN portador, tal como ADN lambda sonificado de entre 100 pb y 4000 pb, y reactivos químicos o reactivos enzimáticos para convertir la citosina en uracilo y se lleva a cabo la conversión de citosina en uracilo. Esta etapa se puede llevar a cabo sin purificación previa. A continuación, los fragmentos procesados químicamente se someten a 5 a 25 ciclos de reacción de PCR para amplificar el ADN genómico de una única célula. Esta etapa se puede llevar a cabo sin purificación. Los productos de amplificación se purifican para su posterior análisis, tal como mediante secuenciación profunda de alto rendimiento.

Ejemplo iii

Lisis celular

Se selecciona una célula, se corta de una placa de cultivo y se dispensa en un tubo utilizando un microscopio de disección láser (LMD-6500, Leica) de la siguiente forma. Las células se colocan en placas de cultivo recubiertas de membrana y se observan mediante microscopía de campo brillante con un objetivo de 10x (Leica). Después se utiliza un láser UV para cortar la membrana alrededor de una célula seleccionada individualmente de tal modo que caiga en la tapa de un tubo de PCR. El tubo se centrifuga brevemente para llevar la célula al fondo del tubo. Se añaden de 3 a 5 µl de regulador de lisis (30 mM de Tris-Cl PH 7,8, EDTA 2 mM, KCl 20 mM, Triton X-100 al 0,2 %, 500 µg/ml de proteasa de Qiagen) al lateral del tubo de PCR y se extiende hacia abajo. A continuación, la célula capturada se lisa térmicamente utilizando el siguiente programa de temperatura en una máquina de PCR: 50 °C 3 horas, 75 °C 30 minutos. Alternativamente, pipetee por la boca una única célula en un regulador de lisis con bajo contenido de sal que contenga EDTA y proteasa, como la proteasa QIAGEN (QIAGEN), a una concentración de 10 a 5000 µg/ml. La condición de incubación varía en función de la proteasa que se utilice. En el caso de la proteasa QIAGEN, la incubación sería de 37 a 55 °C durante 1 a 4 h. A continuación, la proteasa se inactiva térmicamente hasta 80 °C y después se inactiva con inhibidores de la proteasa específicos, como el clorhidrato de fluoruro de 4-(2-aminoetil)bencenosulfonilo (AEBSF) o el fluoruro de fenilmetanosulfonilo (PMSF) (Sigma Aldrich). La lisis celular se conserva a -80 °C.

Un procedimiento de lisis de células individuales ilustrativo es el siguiente:

1. (1) Se prepara un regulador de lisis a partir de: a) 20 ul, Tris 1 M, pH 8.0 (Invitrogen 15568025; final: 20 mM); b) 4 ul de NaCl 5 M (Invitrogen AM9760G; final: 20 mM); c) 15 ul, 10 % Triton X-100 (Sigma 93443; final: 0,15 %); d) TDT de 150 ul a 100 mM (Sigma 43816; final: 15 mM); e) 2 ul 0,5 M de EDTA (Invitrogen AM9260G; final: 1 mM); f) 5 ul 100 uM de ADNmc portador (final: 500 nM); y g) 804 ul de agua. La combinación se mezcla y se almacena a -20 °C.

2. (2) Se prepara un regulador de transposición 2X (5 ul por célula; la receta que figura a continuación para 1 ml) es la siguiente: a) 20 ul 1 M TAPS pH 8,5 (Boston Bio Products BB-2375) (final: 20 mM); b) 10 ul de MgCl<sub>2</sub> 1 M (final: 10 mM); c) 320 ul 50 % PEG 8000 (Hampton Research HR2-535) (final: 16 %); d) 650 ul de agua. La combinación se mezcla y se almacena a -20 °C.

3. (3) La lisis se lleva a cabo de la siguiente forma. Se preparan 7,5 mg/ml de proteasa de Qiagen a partir de a) 1 ul de 60 mg/ml de proteasa de Qiagen y b) 7 ul de agua. Se añaden 2 ul de regulador de lisis por tubo. Se añaden 0,5 ul 7,5 mg/ml de QP por tubo. Las células se lisan en un volumen de 2,5 ul en una máquina de PCR hermética con los siguientes ciclos: a) 50 °C durante 1 h; b) 65 °C durante 1 h; b) 70 °C durante 15 min; c) 4 C constante.

Ejemplo iv

Transposición

La lisis de células individuales y la genoteca de transposomas se mezclan en un sistema regulador que contiene 1 - 100 mM de Mg<sup>2+</sup> y, opcionalmente, 1 - 100 mM de Mn<sup>2+</sup> o Co<sup>2+</sup> o Ca<sup>2+</sup> y se incuban a 37 - 55 °C durante 5 - 240 minutos. El volumen de reacción varía dependiendo del volumen de lisis celular. La cantidad de genoteca de transposomas añadida en la reacción podría ajustarse fácilmente dependiendo del tamaño de fragmentación deseado. La reacción de transposición se detiene quelando el Mg<sup>2+</sup> usando EDTA y opcionalmente EGTA u otros agentes quelantes para iones. Opcionalmente, se podría añadir ADN bicatenario corto a la mezcla como un complemento. El transposoma residual se inactiva mediante la digestión con proteasas, como la proteasa QIAGEN, a una concentración final de 1 a 500 µg/ml a 37 a 55 °C durante 10 a 60 minutos. La proteasa se inactiva entonces mediante calor y/o un inhibidor de proteasa, tal como la AEBSF.

Se proporcionan un método y constructos ilustrativos de la siguiente forma.

Constructo de Nextera:

Los transposones de Nextera tienen una cadena de 5'-/Phos/-CTGTCTCTTATACACATCT-3' (id. de sec. n.º: 1) y una cadena modificada en 5mC de cualquiera de 5'-TMGTMGGMAGMGTAGATGTGTATAAGAGAMAG-3' ("P5") (id. de sec. n.º: 2) o 5'-GTMTMGTGGGMTMGGAGATGTGTATAAGAGAMAG-3' ("P7") (id. de sec. n.º: 3) (IDT, purificación: PAGE). También se puede utilizar Nextera XT (Illumina). M significa citosina metilada.

Cebadores de índice Nextera (IDT), purificación: desalación estándar; después se disuelven en 0,1 X TE a 5 uM y se almacenan a -20 °C) están en el formato de

5'-CAAGCAGAAGACGGCATACGAGAT-[i7]-GTCTCGTGGGCTCGG-3' y

5'-AATGATACGGCGACCAACGAGATCTACAC-[i5]-TCGTCTCGGAGCGTC-3'.

Las secuencias son las siguientes:

701: CAAGCAGAAGACGGCATACGAGATTCGCCTTAGTCTCGTGGGCTCGG (id. de sec. n.º: 4)

702: CAAGCAGAAGACGGCATACGAGATCTAGTACGGTCTCGTGGGCTCGG (id. de sec. n.º: 5)

703: CAAGCAGAAGACGGCATACGAGATTTCTGCCTGTCTCGTGGGCTCGG (id. de sec. n.º: 6)

704: CAAGCAGAAGACGGCATACGAGATGCTCAGGAGTCTCGTGGGCTCGG (id. de sec. n.º: 7)

705: CAAGCAGAAGACGGCATACGAGATAGGAGTCCGTCTCGTGGGCTCGG (id. de sec. n.º: 8)

706: CAAGCAGAAGACGGCATACGAGATCATGCCTAGTCTCGTGGGCTCGG (id. de sec. n.º: 9)

707: CAAGCAGAAGACGGCATACGAGATGTAGAGAGTCTCGTGGGCTCGG (id. de sec. n.º: 10)

708: CAAGCAGAAGACGGCATACGAGATCCTCTCTGGTCTCGTGGGCTCGG (id. de sec. n.º: 11)

709: CAAGCAGAAGACGGCATACGAGATAGCGTAGCGTCTCGTGGGCTCGG (id. de sec. n.º: 12)

710: CAAGCAGAAGACGGCATACGAGATCAGCCTCGGTCTCGTGGGCTCGG (id. de sec. n.º: 13)

711: CAAGCAGAAGACGGCATACGAGATTGCCTCTTGTCTCGTGGGCTCGG (id. de sec. n.º: 14)

712: CAAGCAGAAGACGGCATACGAGATTCCTCTACGTCTCGTGGGCTCGG (id. de sec. n.º: 15)

501: AATGATACGGCGACCACCGAGATCTACACTAGATCGCTCGTCGGCAGCGTC (id. de sec. n.º: 16)

502: AATGATACGGCGACCACCGAGATCTACACCTCTCTATTTCGTCGGCAGCGTC (id. de sec. n.º: 17)

5 503: AATGATACGGCGACCACCGAGATCTACACTATCCTCTTCGTCGGCAGCGTC (id. de sec. n.º: 18)

504: AATGATACGGCGACCACCGAGATCTACACAGAGTAGATCGTCGGCAGCGTC (id. de sec. n.º: 19)

10 505: AATGATACGGCGACCACCGAGATCTACACGTAAGGAGTCGTCGGCAGCGTC (id. de sec. n.º: 20)

506: AATGATACGGCGACCACCGAGATCTACACACTGCATATCGTCGGCAGCGTC (id. de sec. n.º: 21)

507: AATGATACGGCGACCACCGAGATCTACACAAGGAGTATCGTCGGCAGCGTC (id. de sec. n.º: 22)

15 508: AATGATACGGCGACCACCGAGATCTACACCTAAGCCTTCGTCGGCAGCGTC (id. de sec. n.º: 23)

Los constructos utilizados para generar los datos descritos en la presente memoria son secuencias modificadas con 5 mC con n = 20 etiquetas de las siguientes secuencias. Debe entenderse que los expertos en la técnica pueden diseñar muchos otros conjuntos de secuencias de sitios de unión a cebadores de este tipo y las siguientes 20 secuencias de sitios de unión a cebadores de transposones no pretenden ser limitantes de ninguna forma.

1. AGAAGMMGTGTGMMGGTMTA (id. de sec. n.º: 24)

25 2. ATMGTGMGGAMGAGAMAGMA (id. de sec. n.º: 25)

3. AATMMTAGMAMMGGTTMGMM (id. de sec. n.º: 26)

4. AMGTGTTGMAGGTGMAMTMG (id. de sec. n.º: 27)

30 5. AMAMMAMAMGGMMTAGAGTM (id. de sec. n.º: 28)

6. TGGAMAATMAMGMGAMMAGM (id. de sec. n.º: 29)

35 7. TMTATMTAAMGMGMAMMGTGM (id. de sec. n.º: 30)

8. TTMGMTMGMTMTMTMGAAMM (id. de sec. n.º: 31)

9. TGGTGGAGMGTGMAGAMTMT (id. de sec. n.º: 32)

40 10. TATMTTMMTGMMAGMGGM (id. de sec. n.º: 33)

11. MTGAMGTGTGAGGMGMTAGA (id. de sec. n.º: 34)

45 12. MMATMATMMAAMMGMMTTMG (id. de sec. n.º: 35)

13. MAMGAGAAGMMGTMMGMTTA (id. de sec. n.º: 36)

14. MGTAMGTGMAAMAMTMMGMT (id. de sec. n.º: 37)

50 15. MTTGGTMAGGMGAGAAGMAM (id. de sec. n.º: 38)

16. GGMGTGATMAGTGMGTGGAT (id. de sec. n.º: 39)

55 17. GAGMGTTTGGTGAMMGMMAT (id. de sec. n.º: 40)

18. GMMTGMMGGTMMATTGAMMTA (id. de sec. n.º: 41)

19. GTAAGMMAMTMMAGMGTMAM (id. de sec. n.º: 42)

60 20. GATMTGTTGMGMGMTMTGGTG (id. de sec. n.º: 43)

El transposón Tn5 se construye a partir de 5'-/Phos/ CTGTCTCTTATACACATCT-3', mientras que la otra cadena estaba en forma de 5'- [etiqueta] -AGATGTGTATAAGAGAMAG-3'. Cada uno de los oligos (IDT, purificación: PAGE) se disolvió en 0,1 X TE hasta una concentración final de 100 uM. Para cada una de las etiquetas n = 20, se hibridaron dos cadenas a una concentración final de 5 uM cada una. Los 20 transposones hibridados se agruparon después con volúmenes iguales. En segundo lugar, la transposasa se purificó después de la expresión del plásmido pTXB1-Tn5

(Addgene). El transposoma se ensambló a una concentración final de 1,25 uM de dímero (2,5 uM de monómero), se diluyó 1:10 (dímero de 125 nM o monómero de 250 nM) y se dividió en alícuotas para usos individuales y se almacenó a -80 °C.

5 La mezcla de 20 cebadores (para su uso en la mezcla 1 de PCR) estaba en forma de 5'-[etiqueta] AGATGTGTATAAG-3'. Cada uno de los oligos (IDT, purificación: desalación estándar) se disolvió en 0,1 X TE hasta una concentración final de 100 pM, y se combinó con volúmenes iguales (100 pM en total, o 5 pM cada uno). Se almacenó a -20 °C. La mezcla de 40 cebadores (para su uso en la mezcla 2 de PCR) estaba en forma de

10 5'ACACTCTTTCCCTACACGACGCTCTTCCGATCT-[etiquetaMETA]AGATGTGTATAAG-3' para un lado del adaptador Illumina, y

5'GACTGGAGTTCAGACGTGTGCTCTTCCGATCT-[etiquetaMETA]AGATGTGTATAAG-3' para el otro. Cada uno de los oligos (IDT, purificación: PAGE) se disolvió en 0,1 X TE hasta una concentración final de 50 uM, y se combinó con volúmenes iguales (50 uM en total, o 1,25 uM cada uno) y se almacenó a -20 °C.

20 Para la inserción del adaptador utilizando un transposoma, ya sea Nextera, o los transposomas descritos en la presente memoria en una reacción de 10 ul, para cada célula individual, añadir cada uno de los siguientes reactivos en tubos de PCR de baja unión: A) Muestra de lisis de 2,5 ul; B) 5 ul 2x de Trans regulador; C) 2,5 ul de complejo de Tn5 diluido y mantener a 55 °C durante 10 minutos y a 4 °C constantes a partir de entonces.

25 La reacción del transposoma se detiene utilizando un regulador de parada de la siguiente forma: Preparar 0,2 mg/ml de QP; preparar el regulador de parada; 1ul 2x de regulador de parada; 1 ul de 2 mg/ml de QP (100 ug/ml finales). La transposición se detiene o se detiene ejecutándola (un volumen de 12 ul) en condiciones de a) 50 °C durante 40 minutos; b) 70 °C durante 20 minutos, y c) 4 °C constantes a partir de entonces.

Ejemplo v

30 Llenado de espacios y ADN portador

Tras la transposición y la eliminación de la transposasa, se añade a la solución una mezcla de reacción de PCR que incluye Mg<sup>2+</sup>, mezcla de dNTP, cebadores y una ADN polimerasa termoestable, como la exo-ADN polimerasa de Deepvent (New England Biolabs), a una temperatura adecuada y durante un período de tiempo adecuado para llenar el espacio de 9 pb dejado por la reacción de transposición. La temperatura y el tiempo de incubación para rellenar el espacio dependen de la ADN polimerasa específica utilizada. Tras la reacción, la ADN polimerasa se inactiva opcionalmente mediante calentamiento y/o tratamiento con proteasas, tal como la proteasa QIAGEN. La proteasa, si se utiliza, se inactiva entonces mediante calor y/o un inhibidor de proteasa. El ADN portador (ADNbc) se añade a continuación al medio de reacción. El ADN portador presente incluye fragmentos de ADN de entre 100 pb y 4000 pb de longitud, tal como los que se crean al sonicar el ADN lambda. El ADN portador está presente en el medio de reacción en una cantidad de al menos 20 ng y entre 20 ng y 50 ng.

Un procedimiento ilustrativo de rellenado de espacio es el siguiente:

45 1. Preparar la mezcla de dNTP dentro de 5-metil-dCTP (2,5 mM cada nt). Añadir 100 ul de 10 mM de dTTP (NEB N0443S), 100 ul de 10 mM de dGTP (NEB N0442S), 100 ul de 10 mM de dATP (NEB N0440S) a 100 ul de 10 mM de 5-metil-dCTP (NEB N0356S) y, a continuación, mezcle bien agitando en vórtice y almacenar a -20 °C.

50 2. Hacer una mezcla de PCR (23 ul por célula); a) 7 ul de regulador de reacción Q5 (incluido con Q5); b) Potenciador Q5 High GC de 7 ul (incluido con el Q5); c) Mezcla de dNTP de 2,8 ul dentro de 5-metil-dCTP (2,5 mM cada nt); d) 0,7 ul 100 mM de MgCl<sub>2</sub> (Invitrogen AM9530G); e) 0,35 ul Q5 (NEB M0491S); f) 1 ul de 20 ng/ul de ADN portador lambda (sonicado de 200 a 300 pb); g) 4,15ul de H<sub>2</sub>O. agité en vórtice para mezclar.

3. Añadir 23 ul de mezcla de PCR por tubo, evitando tocar el líquido. Agite en vórtice y gire hacia abajo.

55 Rellenar los espacios al ejecutar (volumen de 35 ul) en las siguientes condiciones. a) 4 °C durante 3 minutos (para permitir que la tapa se precaliente); b) 65 °C durante 3 minutos; c) almacenar a 4 C.

60 Para la limpieza: 1. Añadir 200 ul de regulador de unión directamente a cada tubo, mezcle 10 veces y transferir a la columna (ZYMO DCC); 2. 200 ul de lavado dos veces; 3. Eluir en 17,8 ul de regulador de elución (sin EDTA, botella NEB con tapa blanca del kit EM-Seq).

Ejemplo vi

65 Conversión química o enzimática de citosina en uracilo

A continuación se añade ADN portador (ADNbc) al medio de reacción junto con reactivos químicos para convertir las citosinas en uracilos. El ADN portador presente durante la conversión química o enzimática de la citosina en uracilo incluye fragmentos de ADN de entre 100 pb y 4000 pb de longitud, tal como los que se crean al sonicar el ADN lambda. El ADN portador está presente en el medio de reacción en una cantidad de al menos 20 ng y entre 20 ng y 50 ng.

#### Ejemplo vii

##### Amplificación de fragmentos de ADN

Según un aspecto, los métodos generales conocidos por los expertos en la técnica se utilizan para amplificar un fragmento de ADN. Los fragmentos convertidos químicamente del ejemplo anterior se combinan con reactivos de reacción de PCR en un medio acuoso. El medio acuoso se somete entonces a condiciones de PCR para amplificar por PCR cada fragmento de ADN.

#### Ejemplo viii

##### Secuenciación de amplicones de fragmentos de ADN

Según un aspecto, los fragmentos se secuencian utilizando métodos conocidos por los expertos en la técnica y las secuencias se almacenan en una memoria legible por ordenador. A continuación, las secuencias pueden compararse y ensamblarse en secuencias genómicas utilizando métodos, incluidos métodos de software, conocidos por los expertos en la técnica.

#### Ejemplo ix

##### Detección de metilación unicelular

Un aspecto de la presente descripción se muestra en la figura 7.

(1) El ADN diana se extrae de células únicas, o de 2 células, o de 4 células, o... 100 células. Se considera que tal ADN extraído está en bajo volumen.

(2) El ADN diana está fragmentado y los transposones de metilo se insertan o unen mediante un transposoma Tn5. El cebador del transposón que se une a Tn5 está completamente metilado en cada citosina. El resultado son fragmentos de ADN diana que incluyen adaptadores de PCR completamente metilados.

(3) Los fragmentos están llenos de espacios. Después de rellenar los espacios, se añade el ADN portador a la reacción. El ADN portador se crea a partir de ADN lambda que se ha sonicado en fragmentos que tienen de 100 pb a 400 pb.

(4) El medio de reacción se purifica mediante columnas giratorias de ADN. La conversión de ADN se llevó a cabo con el kit EM-Seq de NEB. La etapa ABOPEC del kit EM-Seq se modifica para reducir el volumen a 40 ul.

(5) Después de la conversión, la polimerasa Q5U y el regulador se añaden directamente a la reacción sin purificación del ADN. La amplificación del genoma completo se realiza mediante el uso de cebadores de PCR de Nextera (si se utiliza el sistema de transposomas de Nextera) o cebadores de PCR como se describe en la presente memoria.

(6) Después de la reacción de PCR, el medio de reacción se purifica para eliminar el ADN portador monocatenario y la genoteca purificada está lista para la secuenciación del ADN.

Como se muestra en la figura 8, el método descrito en la presente memoria que utiliza el enfoque del transposoma con el ADN portador dio como resultado una mayor cobertura genómica en comparación con una técnica de secuenciación del metiloma completo de una única célula existente. Como se muestra en la figura 9, el método descrito en la presente memoria que utiliza el enfoque del transposoma con el ADN portador dio como resultado una tasa de mapeo más alta en comparación con una técnica de secuenciación del metiloma completo de una única célula existente. Como se muestra en la figura 10, el método descrito en la presente memoria que utiliza el enfoque del transposoma con el ADN portador dio como resultado una mayor precisión en comparación con una técnica de secuenciación del metiloma completo de una única célula existente. Las técnicas incluyen scWGBS "Smallwood, S. A., Lee, H. J., Angermueller, C., Krueger, F., Saadeh, H., Peat, J.,... y Kelsey, G. (2014). Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. Nature methods, 11 (8), 817." y scCOOL-seq "Guo, F., Li, L., Li, J., Wu, X., Hu, B., Zhu, P.,... y Tang, F. (2017). Single-cell multi-omics sequencing of mouse early embryos and embryonic stem cells. Cell research, 27(8), 967."

#### Ejemplo x



## Conversión enzimática del kit Metil-seq

### Oxidación de 5-metilcitosinas y 5-hidroximetilcitosinas

- 5 1. Preparar el regulador TET2.

Añadir 100 µl de regulador de reacción TET2 (incluido con TET2) a un tubo de suplemento de regulador de reacción TET2 (incluido con TET2) y mezcle bien, después almacenar a -20 °C.

- 10 2. Preparar hierro recién diluido (II).

Diluir la solución de hierro (II) de 500 mM (incluida con TET2) a 0,4 mM con agua.

- 15 3. En hielo, añadir la premezcla de oxidación de 9,6 µl (6 µl de regulador de reacción TET2 reconstituido (incluido con el TET2), un suplemento de oxidación de 0,6 µl (incluido con el TET2), un potenciador de oxidación de 0,6 µl (incluido con el TET2) y 2,4 µl de TET2 (E7120S)) directamente al ADN etiquetado de 17,4 µl. Mezclar bien agitando en vórtex, centrifugar brevemente y, a continuación, añadir 3 µl de hierro (II) recién diluido y 0,4 mM hasta un volumen total de 30 µl. Mezclar bien agitando en vórtice y después centrifugar brevemente. Incubar a 37 °C durante 1 hora y después a 4 °C en un termociclador. El volumen total de la reacción TET es de 30 µl, lo que es suficiente para un ADN de bajo aporte. El volumen total puede estar entre 20 µl y 40 µl. El volumen en este intervalo permite ventajosamente llevar a cabo la siguiente etapa de purificación en el mismo tubo único para reducir la pérdida de ADN.

- 20 4. Añadir 0,6 µl de reactivo de parada (incluido con TET2). Mezclar bien agitando en vórtice y después centrifugar brevemente. Incubar a 37 °C durante 30 minutos y después a 4 °C en un termociclador.

### 25 Limpieza del ADN oxidado

- 30 1. Añadir 200 µl de regulador de unión directamente al tubo de PCR, mezcle 10 veces y transferir a la columna (ZYMO DCC)

2. 200 µl, lavar dos veces

- 35 3. Eluir en 12,3 µl de regulador de elución (sin EDTA), botella NEB con tapa blanca del kit EM-Seq. La purificación en columna del ADN se lleva a cabo preferiblemente para purificar las glóbulos, a fin de minimizar la pérdida de ADN debida a la purificación.

### Desnaturalización del ADN con hidróxido de sodio

- 40 1. Preparar NaOH 0,1 N (Sigma) recién diluido añadiendo 1 µl de NaOH 10 M a 99 µl de agua.

2. Añadir 3 µl de NaOH 0,1 N a los 12 µl de ADN oxidado. Agitar en vórtice para mezclar y después centrifugar brevemente. Incubar a 50 °C durante 10 minutos en el termociclador precalentado. Después, colocar en hielo y continuar con la siguiente sección.

### 45 Desaminación de citosinas

1. Diluir el ácido clorhídrico al 37 % (Sigma 30721) a 0,018 M añadiendo 1,48 µl a 998,52 µl de agua.

- 50 2. En hielo, añadir 7 µl de HCl 0,018 M a los 15 µl de ADN desnaturalizado.

3. En hielo, añadir la premezcla de desaminación de 18 µl (13,2 µl de agua, 4 µl de regulador de reacción APOBEC (incluido con APOBEC), 0,4 µl de BSA (incluido con APOBEC), 0,4 µl de APOBEC (E7120S)). Mezclar bien agitando en vórtex, centrifugar brevemente.

- 55 4. Incubar a 37 °C durante 3 horas y después a 4 °C en un termociclador.

El volumen de la reacción para la etapa de desnaturalización y la reacción ABOPEC es de 40 µl y puede estar entre 30 µl y 60 µl. Este volumen permite ventajosamente evitar una etapa de purificación del ADN después del tratamiento con APOBEC. La PCR se puede realizar directamente añadiendo 40 µl de mezcla de regulador de PCR, incluida la ADN polimerasa y los cebadores, a la reacción ABOPEC de 40 µl.

## Ejemplo xi

### Preparación de genotecas mediante PCR

### 65 Constructo de Nextera

Combinar un sistema de desaminación de 40 ul, una mezcla maestra Q5U 2x de 40 ul y un índice nextera de 100 mM de 0,4 ul (P5 y P7). Amplificar haciendo funcionar (volumen de 80,8 ul): a) 4 °C durante 3 minutos (para permitir que la tapa se precaliente), b) 98 °C durante 20 s, c) 12 ciclos de 98 °C durante 10 s, 62 °C durante 30 s, 65 °C durante 1 minuto, e) 65 °C durante 5 minutos y f) 4 °C constantes. Selección de tamaño 1,2 de perlas con glóbulos AMPure.

Construido de META (amplificación de etiquetado de múltiples extremos):

Combinar lo siguiente: Mezcla de imprimación META 20, 2 ul; 40 ul de elución de ADN; Mezcla maestra Q5U 2x de 40 ul; incubación a 98 °C durante 20 s, 10 ciclos de [98 °C durante 10 s, 62 °C durante 30 s, 65 °C durante 1 min] y 65 °C durante 5 min. El producto de amplificación se purifica en esta etapa con 13,8 ul de regulador de elución y las genotecas de secuenciación se preparan mediante dos etapas de PCR adicionales. En la primera etapa de la PCR, la PCR se realizó mediante la adición de 16,5 ul de PCR Mix 2 (15 ul de Q5U 2x mezcla maestra y 1,5 ul de 40 cebadores) y la incubación a 98 °C durante 30 s, 2 ciclos de 98 °C durante 10 s + 62 °C durante 30 s + 65 °C durante 1 min y 65 °C durante 5 min. En la segunda etapa de PCR, los cebadores se eliminaron de modo similar mediante la adición de 1 ul 20 U/ul de Exol (NEB M0293S) y la incubación a 37 °C durante 30 min, 72 °C durante 20 min. La PCR se realizó de modo similar mediante la adición de 2,5 ul de cebador índice UI-Neb (NEB E7335S, E7500S, E7710S, E7730S) y 6,5 ul de PCR Mix 3 (5 ul de Q5U 2x mezcla maestra NEB), 1,25 ul de agua, 0,25 ul de imprimador universal (IDT), purificación: PAGE)) e incubación a 98 °C durante 30 s, 2 o más ciclos de 98 °C durante 10 s + 62 °C durante 30 s + 65 °C durante 1 min y 65 °C durante 5 min. Las genotecas se pueden agrupar en esta etapa o en cualquier etapa posterior. Se utilizan 1,2 glóbulos AMPure para seleccionar el tamaño.

Ejemplo ix

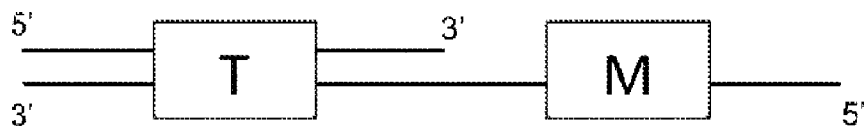
Kits

Los materiales y reactivos requeridos para el método de amplificación descrito pueden ensamblarse juntos en un kit. Los kits de la presente descripción generalmente incluirán al menos el transposoma (que consiste en la enzima transposasa y el ADN de transposón), los nucleótidos, la ADN polimerasa, el ADN portador y los reactivos químicos para convertir la citosina en uracilo o la 5-metilcitosina en uracilo, necesarios para llevar a cabo el método reivindicado junto con conjuntos de cebadores según sea necesario. En una realización preferida, el kit también contendrá instrucciones para amplificar el ADN a partir de muestras de ADN. Los kits ilustrativos son aquellos adecuados para su uso en la amplificación del ADN genómico completo. En cada caso, los kits tendrán preferiblemente recipientes distintos para cada reactivo, enzima o reactivo individual. En general, cada agente se dividirá en alícuotas adecuadas en sus respectivos recipientes. Los medios de contenedor de los kits generalmente incluirán al menos un vial o tubo de ensayo. También son posibles frascos, botellas y otros medios de recipiente en los que se colocan los reactivos y se colocan en alícuotas. Los recipientes individuales del kit se mantendrán preferiblemente confinados para su venta comercial. Los recipientes más grandes adecuados pueden incluir recipientes de plástico moldeados por inyección o soplado en los que se retienen los viales deseados. Las instrucciones se proporcionan preferiblemente con el kit.

## REIVINDICACIONES

1. Un método para analizar las características de metilación del ADN genómico diana que comprende
  - 5 poner en contacto el ADN genómico con una genoteca de transposomas con cada transposoma de la genoteca que tiene dos transposasas y dos ADN de transposón, en donde cada ADN de transposón incluye un sitio de unión a la transposasa y una secuencia del sitio de unión al cebador, en donde la secuencia del sitio de unión al cebador es diferente del sitio de unión al cebador de otros miembros de la genoteca de transposomas, en donde cada ADN de
    - 10 transposón incluye una o más 5-metilcitosinas, en donde la genoteca de transposomas se une a ubicaciones diana a lo largo del ADN genómico y la transposasa escinde el ADN genómico en una pluralidad de fragmentos de ADN genómico bicatenario que representan una genoteca de fragmentos de ADN genómico, donde cada fragmento de ADN genómico bicatenario incluye una o más citosinas y/o una o más 5-metilcitosinas y una secuencia de sitios de unión al cebador única y/o diferente en cada extremo del fragmento de ADN genómico, rellenar un espacio entre el ADN de transposón y el fragmento de ADN genómico para formar una genoteca de productos de extensión de fragmentos de ADN genómico bicatenario que tienen secuencias de sitios de unión al cebador únicas y/o diferentes en cada extremo,
      - 20 tratar la genoteca de productos de extensión de fragmentos de ADN genómico bicatenario para convertir la citosina en uracilo en presencia de ADN portador, y amplificar los productos de extensión de fragmentos de ADN genómico bicatenario para producir amplicones, y secuenciar los amplicones.
  - 25 2. El método de la reivindicación 1, en donde cada transposoma dentro de la genoteca de transposomas incluye dos secuencias de sitios de unión a cebadores diferentes.
  - 30 3. El método de la reivindicación 1, en donde cada transposoma dentro de la genoteca de transposomas incluye dos secuencias idénticas del sitio de unión al cebador en cada transposón del transposoma, que son diferentes de las secuencias del sitio de unión al cebador en otros transposomas de la genoteca de transposomas.
  - 35 4. El método de la reivindicación 1, en donde el ADN genómico es ADN genómico completo obtenido de una sola célula.
  5. El método de la reivindicación 1, en donde la transposasa es la transposasa Tn5, la transposasa Mu, la transposasa Tn7 o la transposasa IS5.
  - 40 6. El método de la reivindicación 1, en donde el ADN de transposón incluye un sitio de unión a Tnp bicatenario de 19 pares de bases y un saliente, en donde el saliente incluye una secuencia de sitios de unión al cebador única y/o diferente en el extremo 5' del saliente.
  - 45 7. El método de la reivindicación 1, en donde las transposasas unidas se eliminan de los fragmentos bicatenarios antes de rellenar los espacios y extender los fragmentos de ADN genómico bicatenario.
  8. El método de la reivindicación 1, en donde el ADN genómico proviene de una célula prenatal, una célula cancerosa o una célula tumoral circulante; preferiblemente, el ADN genómico proviene de una sola célula prenatal, una sola célula cancerosa o una
    - 50 sola célula tumoral circulante.
  9. El método de la reivindicación 1, en donde la secuencia única y diferente del sitio de unión al cebador es un sitio de unión al cebador de la PCR específico.
  - 55 10. El método de la reivindicación 1, en donde la genoteca de transposomas incluye de 1 a 100, 1 a 10, 5 a 50, 30 a 100, 15 a 25, 100 a 1000, 1000 a 10.000 o 10.000 a 100.000 secuencias de sitios de unión a cebadores únicas y diferentes.
  - 60 11. El método de la reivindicación 1, en donde las diferentes secuencias del sitio de unión al cebador son ortogonales.
  12. El método de la reivindicación 1, en donde el ADN de transposón está metilado en cada citosina.
  - 65 13. El método de la reivindicación 1, en donde el ADN de transposón incluye adaptadores de citosina metilada; preferentemente, la etapa de rellenado de espacios incluye el uso de dCTP metilado en lugar de dCTP en la mezcla de dNTP.

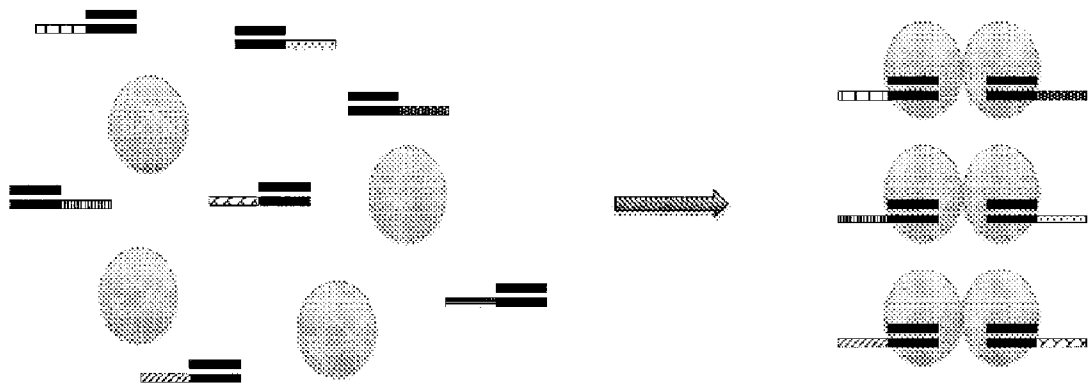
14. El método de la reivindicación 1, que se **caracteriza por** uno o más de los siguientes elementos:
- (i) el ADN portador se selecciona del grupo que consiste en fragmentos de ADNbc que tienen una longitud de entre 100 pares de bases (pb) y pares de bases de 4 kilos;
  - (ii) el ADN portador es un tipo de ADN diferente de o igual al ADN diana;
  - (iii) el ADN portador es ADN lambda sonificado;
  - (iv) el ADN portador no incluye los adaptadores de secuenciación de Illumina;
  - (v) el ADN portador se añade al medio de reacción en una cantidad que es de 100 a 10.000 veces mayor que la cantidad de ADN de la muestra.
15. El método de la reivindicación 1, que incluye además una etapa después de la etapa de rellenado de espacios pero antes de la etapa de conversión: purificar el medio de reacción que incluye los segmentos bicatenarios rellenos y el ADN portador; o, el medio de reacción que incluye los segmentos bicatenarios llenados y el ADN portador pasa directamente a la etapa de conversión sin purificación; preferentemente, la etapa de purificación se realiza mediante purificación de ADN basada en columnas giratorias o glóbulos de ADN.
16. El método de la reivindicación 1, en donde el reactivo para convertir la citosina en uracilo no es bisulfito o excluye el bisulfito.
17. El método de la reivindicación 1, que incluye además una etapa después de la etapa de conversión pero antes de la etapa de amplificación: purificar el medio de reacción que incluye los fragmentos convertidos químicamente; o, el medio de reacción que incluye los fragmentos convertidos químicamente pasa directamente a la etapa de amplificación sin purificación.



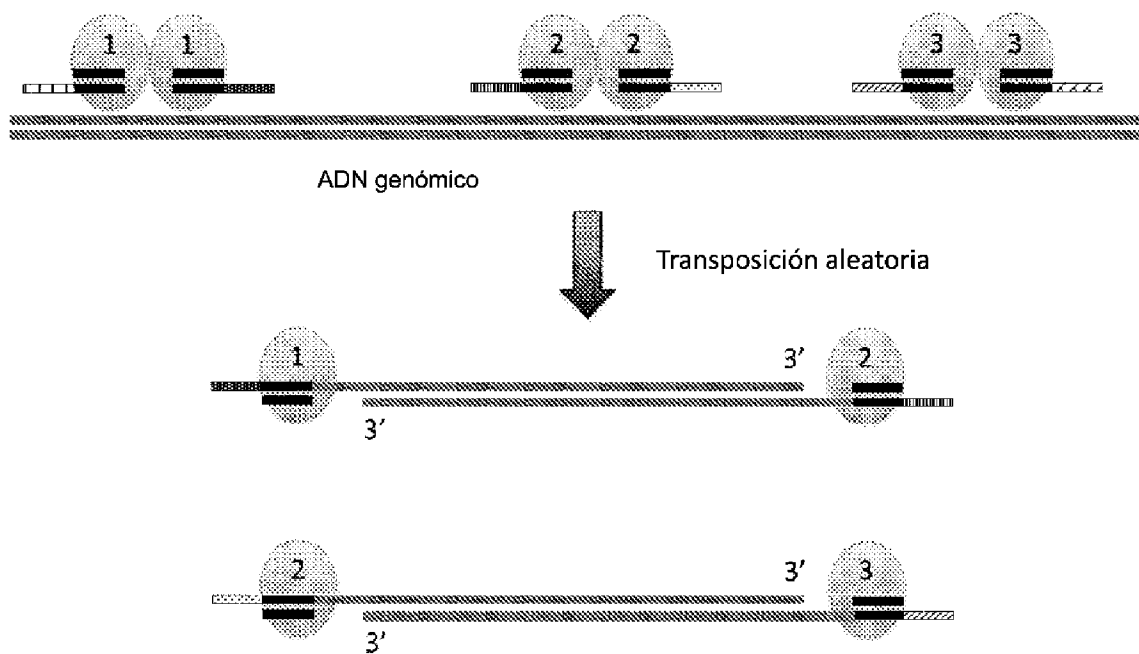
T: Sitio de unión a la transposasa

M: Sitio de imprimación multiplex

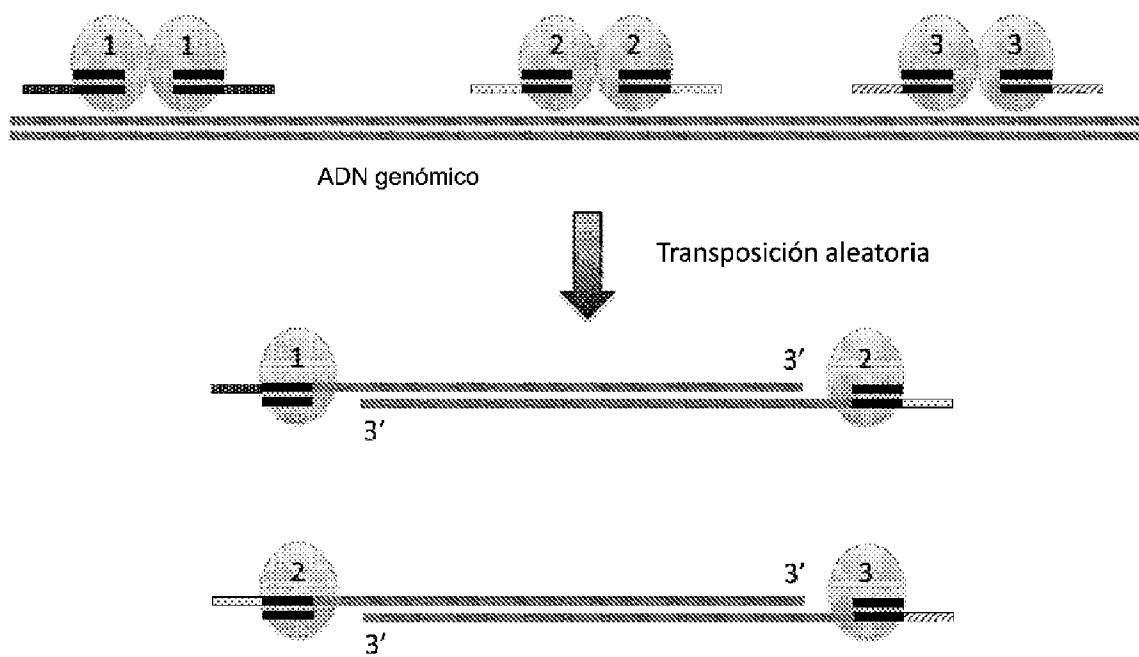
**Figura 1**



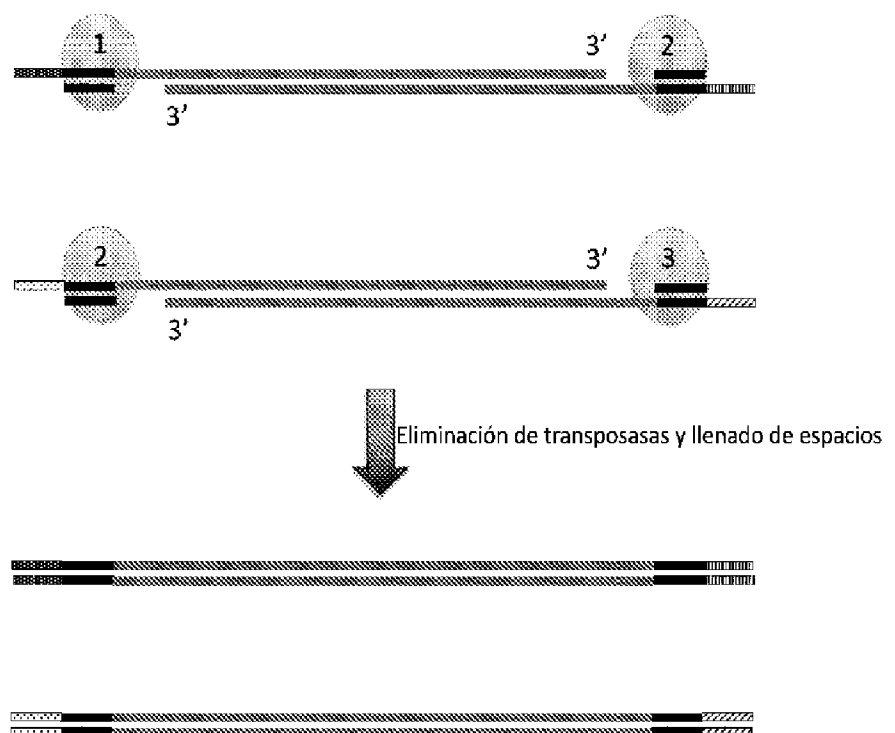
**Figura 2**



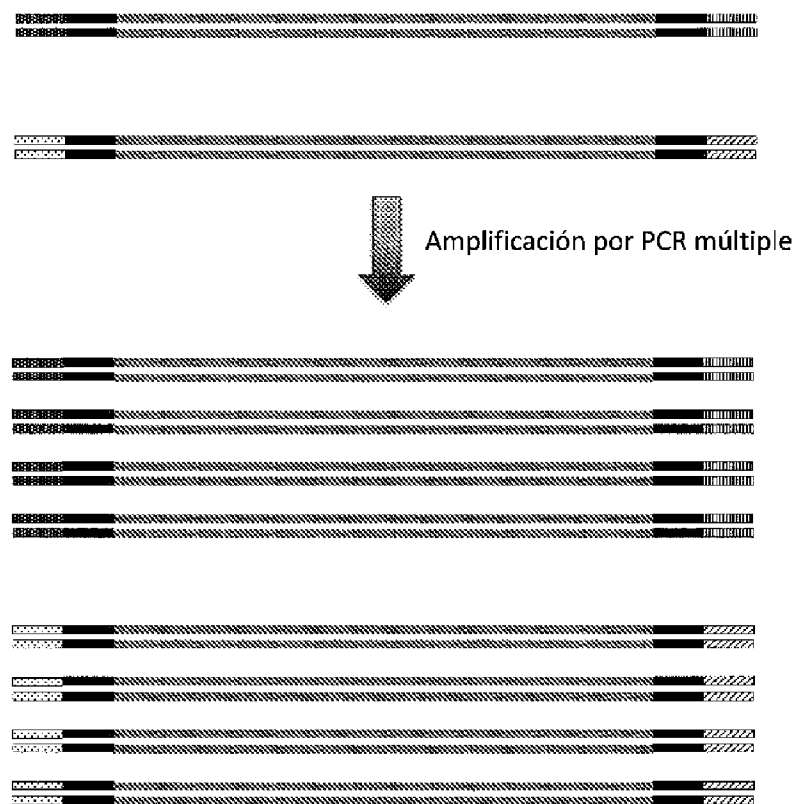
**Figura 3A**



**Figura 3B**

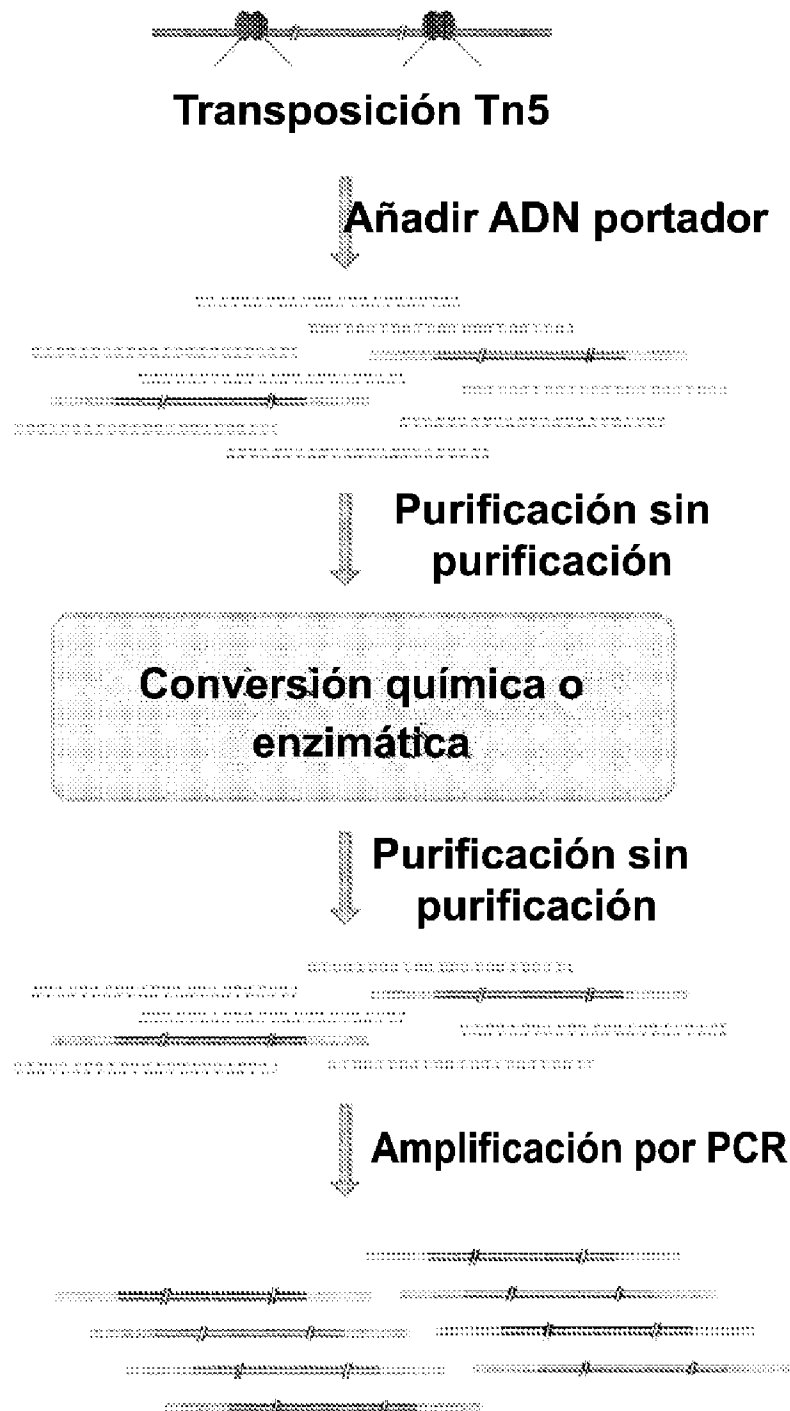


**Figura 4**



**Figura 5**





**Figura 6**

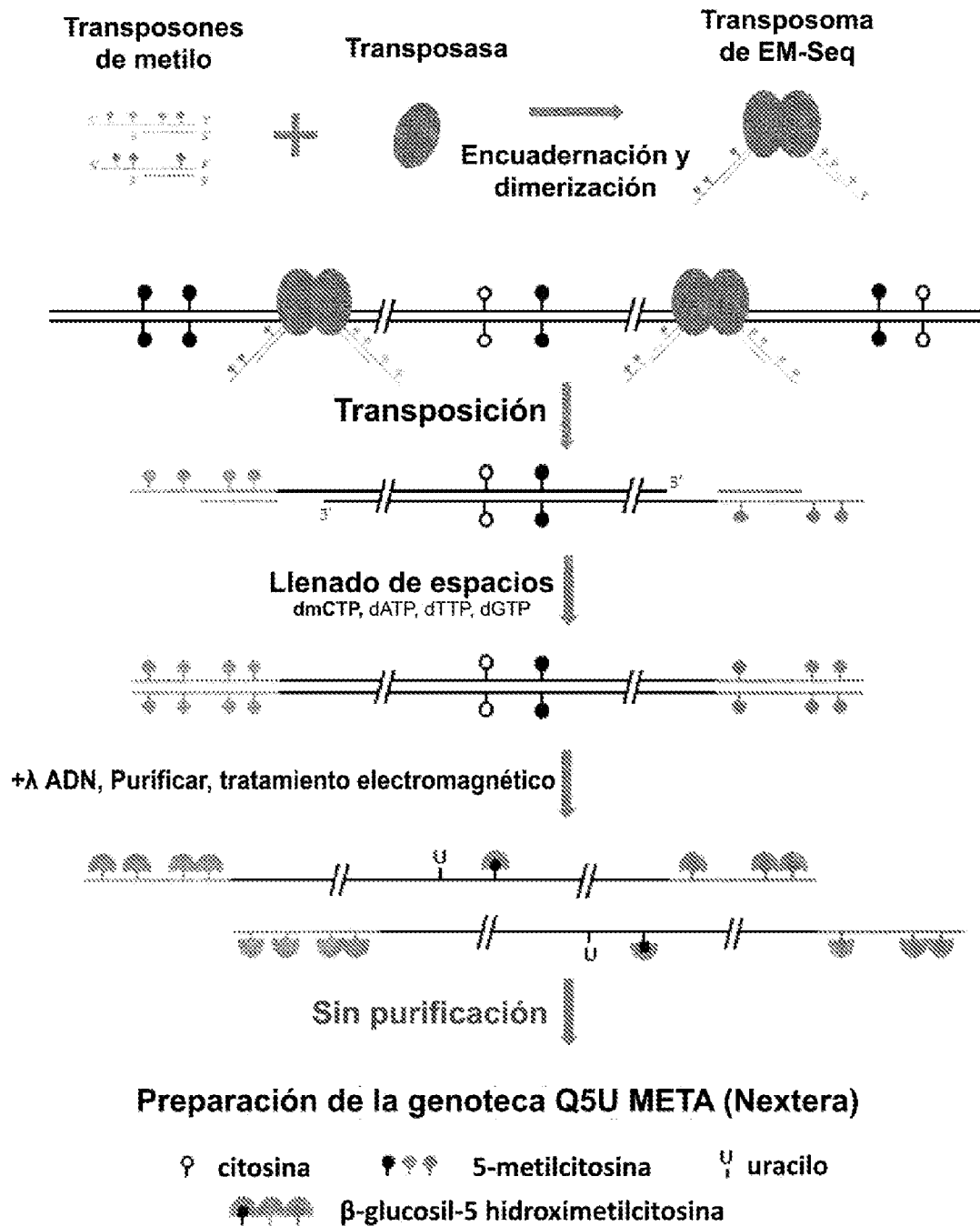
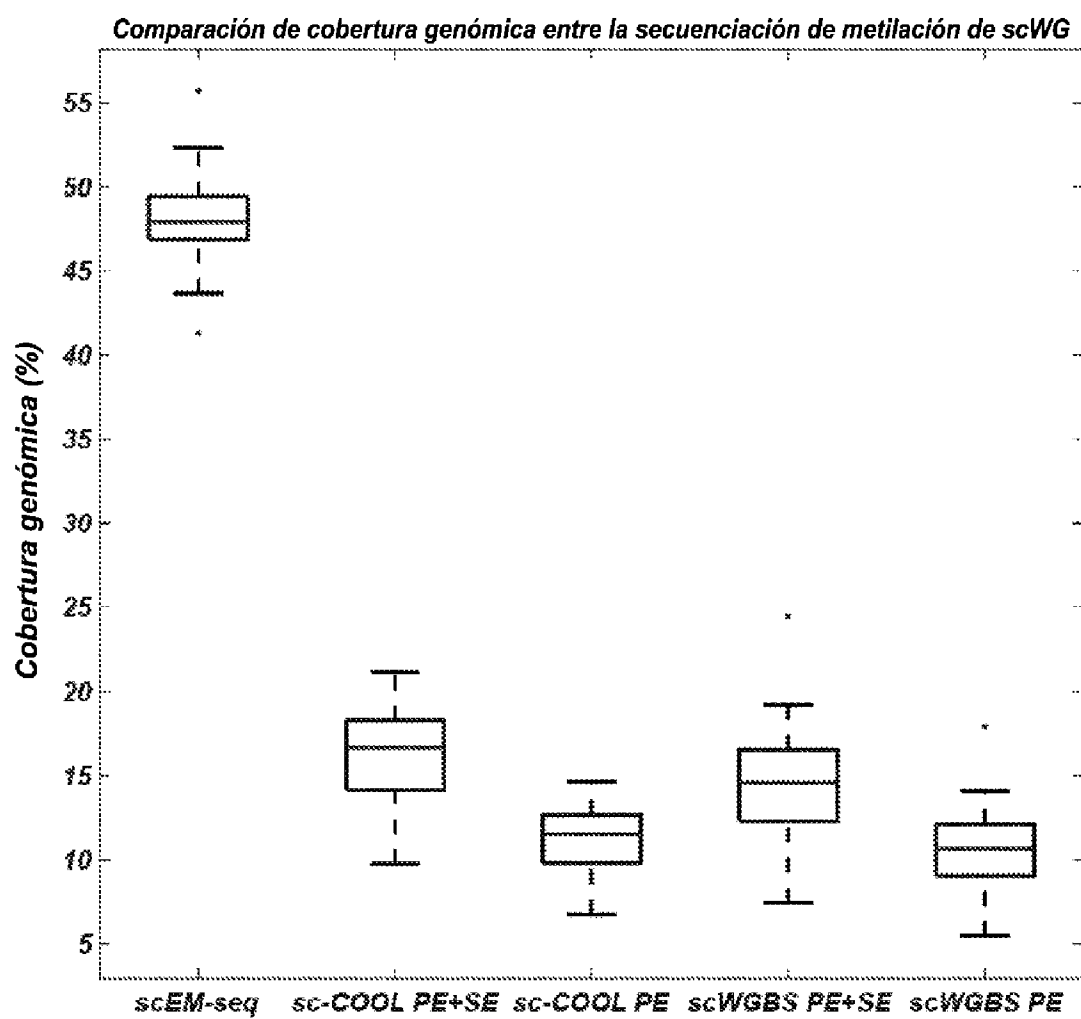


Figura 7

**Figura 8**

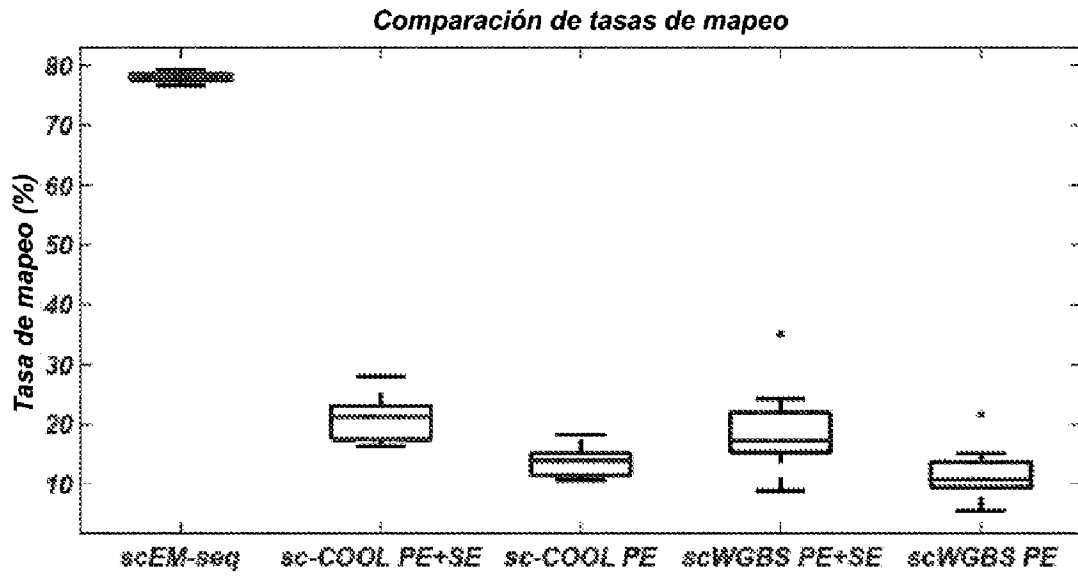


Figura 9

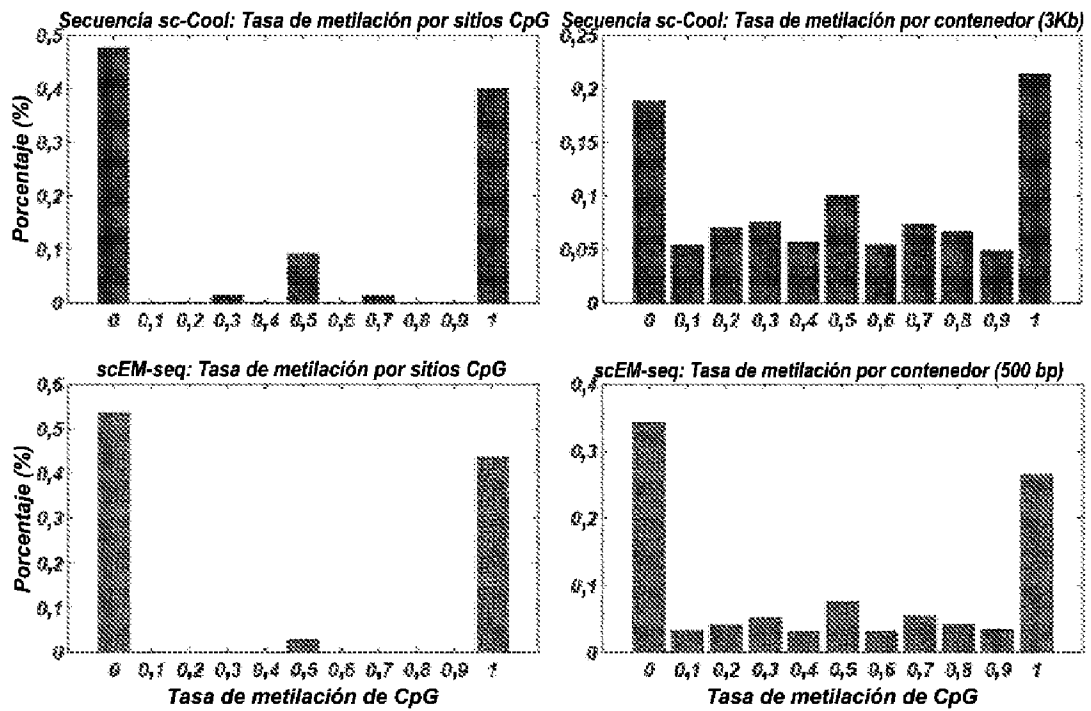


Figura 10