

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **3 014 783**

51 Int. Cl.:

C12Q 1/6806 (2008.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **05.06.2018** **E 23182878 (1)**

97 Fecha y número de publicación de la concesión europea: **12.02.2025** **EP 4293122**

54 Título: **Bibliotecas de genoma completo de célula individual para la secuenciación de metilación**

30 Prioridad:

07.06.2017 US 201762516324 P

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
24.04.2025

73 Titular/es:

OREGON HEALTH & SCIENCE UNIVERSITY
(50.00%)
3181 S.W. Sam Jackson Park Road
Portland, Oregon 97239, US y
ILLUMINA, INC. (50.00%)

72 Inventor/es:

ADEY, ANDREW C;
MULQUEEN, RYAN;
STEEMERS, FRANK J.;
POKHOLOK, DMITRY K. y
NORBERG, STEVEN

74 Agente/Representante:

DEL VALLE VALIENTE, Sonia

ES 3 014 783 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

Bibliotecas de genoma completo de célula individual para la secuenciación de metilación

5 **Campo**

Las realizaciones de la presente descripción se refieren a la secuenciación de ácidos nucleicos. En particular, las realizaciones de los métodos y composiciones proporcionados en la presente memoria se refieren a la producción de bibliotecas de secuenciación de bisulfito de célula individual y a la obtención de datos de secuencia de las mismas.

10

Antecedentes

La secuenciación de célula individual de alto recuento celular ha demostrado su eficacia en la separación de poblaciones dentro de tejidos complejos a través de los transcriptomas, la accesibilidad a la cromatina y las diferencias mutacionales. Además, la resolución de célula individual ha permitido evaluar las trayectorias de diferenciación celular en patrones genómicos específicos, tales como la metilación del ADN. La metilación del ADN es una adición covalente a la citosina; una marca con especificidad de tipo celular que es objeto de modificación activa en los tejidos en desarrollo. La metilación del ADN puede sondearse a resolución de pares de bases utilizando la química de desaminación del tratamiento con bisulfito sódico.

20

El trabajo reciente ha optimizado la secuenciación por bisulfito hasta el punto de requerir entradas de una célula individual en la secuenciación de bisulfito de representación reducida de una célula individual (*scRRBS, single cell reduced representation bisulfite sequencing*) o la secuenciación de bisulfito del genoma completo de una célula individual (*scWGBS, single cell whole genome bisulfite sequencing*). Sin embargo, estos métodos carecen de escalabilidad y se basan en la desconvolución de célula individual a través de la generación de bibliotecas aisladas y en paralelo en que las reacciones de célula individual se realizan de forma aislada. Para la secuenciación de cada célula se necesita un conjunto de reactivos totalmente nuevo, lo que da lugar a un escalado lineal de los costes por cada célula adicional. Debido a las dificultades de la conversión del ADN con bisulfito, no se han utilizado sistemas microfluídicos basados en gotas o en microplacas (*chips*) para la secuenciación con bisulfito de células individuales, ni existen estrategias teóricamente viables que utilicen plataformas alternativas.

25

30

Smallwood y col. (2014), “Single-Cell Genome-Wide Bisulfite Sequencing for Assessing Epigenetic Heterogeneity”, exponen un método de secuenciación con bisulfito de célula individual.

35

Vitak y col. (2017), “Sequencing thousands of single-cell genomes with combinatorial indexing”, exponen un método de secuenciación indexada combinatoria de célula individual.

Resumen de la solicitud

40

La invención se define mediante las reivindicaciones adjuntas.

En la presente memoria se proporcionan composiciones y ensayos de perfilado del metiloma de célula individual de alto recuento celular y escalables. La secuenciación del genoma completo de célula individual (*scWGBS*) se mejora con las estrategias de indexación combinatoria de célula individual proporcionadas en la presente memoria, de modo que las células se pueden procesar en masa y el resultado de la célula individual se demultiplexa por ordenador. En algunas realizaciones, los métodos proporcionados en la presente memoria hacen uso de la incorporación de adaptadores basados en transposasa, lo que da como resultado una mayor eficacia y tasas de alineación mucho más altas que los métodos existentes. El uso de transposasa para anexas uno de los dos adaptadores de secuenciación permite una construcción de bibliotecas mucho más eficaz con menos lecturas de ruido, lo que da como resultado una tasa de alineación de ~60 % (tasas similares a las de las estrategias de células a granel) en comparación con el 10-30 % utilizando métodos de pocillo individual y célula individual. Esto da como resultado lecturas de secuencias más utilizables y una drástica reducción de costes en la porción de secuenciación del ensayo. El uso de estrategias de indexación combinatoria de célula individual para producir bibliotecas de secuenciación de bisulfito de célula individual se demuestra en una mezcla de células humanas y de ratón con una velocidad de colisión mínima. También se demuestra el éxito de la desconvolución de una mezcla de tres tipos de células humanas y lograr una asignación de tipo celular utilizando datos disponibles públicamente.

45

50

55

Definiciones

60

Como se utiliza en la presente memoria, los términos “organismo”, “sujeto” se utilizan indistintamente y se refieren a animales y plantas. Un ejemplo de un animal es un mamífero, tal como un ser humano.

65

Como se utiliza en la presente memoria, la expresión “tipo celular” pretende identificar células basándose en la morfología, el fenotipo, el origen del desarrollo u otra característica celular distintiva conocida o reconocible. Se puede obtener una variedad de diferentes tipos celulares de un solo organismo (o de la misma especie de organismo). Como tipos celulares ilustrativos se incluyen, pero sin limitación, células de la vejiga urinaria, epitelio pancreático, alfa

pancreático, beta pancreático, endotelial pancreático, linfoblasto de médula ósea, linfoblasto B de médula ósea, macrófago de médula ósea, eritroblasto de médula ósea, células dendríticas de médula ósea, adipocito de médula ósea, osteocito de médula ósea, condrocito de médula ósea, promieloblasto, megacarioblasto de médula ósea, vejiga, linfocito B cerebral, neuroglíocito cerebral, neurona, astrocito cerebral, neuroectodermo, macrófago cerebral, microglíocito cerebral, epitelio cerebral, neurona cortical, fibroblasto cerebral, epitelial mamario, epitelial de colon, linfocito B de colon, epitelial mamario, mioepitelial mamario, fibroblasto mamario, enterocito de colon, epitelial de cuello uterino, epitelial de ovario, fibroblasto de ovario, epitelial de conducto mamario, epitelial de lengua, célula dendrítica de amígdala, linfocito B de amígdala, linfoblasto de sangre periférica, linfoblasto T de sangre periférica, linfocito T cutáneo de sangre periférica, linfocito citolítico de sangre periférica, linfoblasto B de sangre periférica, monocito de sangre periférica, mieloblasto de sangre periférica, monoblasto de sangre periférica, promieloblasto de sangre periférica, macrófago de sangre periférica, basófilo de sangre periférica, endotelio hepático, mastocito hepático, epitelial de hígado, linfocito B de hígado, endotelial de bazo, epitelial de bazo, linfocito B de bazo, hepatocito de hígado, Alexander de hígado, fibroblasto de hígado, epitelial de pulmón, epitelial de bronquio, fibroblasto de pulmón, linfocito B de pulmón, Schwann de pulmón, epidermoide de pulmón, macrófago de pulmón, osteoblasto de pulmón, neuroendocrina, alveolar de pulmón, epitelial de estómago y fibroblasto de estómago.

Como se utiliza en la presente memoria, el término “tejido” pretende significar una colección o agregación de células que actúan juntas para realizar una o más funciones específicas en un organismo. Como opción, las células pueden ser morfológicamente similares. Como tejidos ilustrativos se incluyen, pero sin limitación, tejido de ojo, músculo, piel, tendón, vena, arteria, sangre, corazón, bazo, ganglio linfático, hueso, médula ósea, pulmón, bronquios, tráquea, intestino delgado, intestino grueso, colon, recto, glándula salival, lengua, vesícula biliar, apéndice, hígado, páncreas, cerebro, estómago, piel, riñón, uréter, vejiga, uretra, gónada, testículo, ovario, útero, trompa de Falopio, timo, pituitaria, tiroides, suprarrenal o paratiroides. El tejido puede proceder de cualquiera de una variedad de órganos de un ser humano o de otro organismo. Un tejido puede ser un tejido sano o un tejido enfermo. Como ejemplos de tejidos enfermos se incluyen, pero sin limitación, una variedad de tumores malignos con metilación aberrante, por ejemplo, tumores malignos en pulmón, mama, colon y recto, próstata, nasofaringe, estómago, testículos, piel, sistema nervioso, hueso, ovario, hígado, tejidos hemáticos, páncreas, útero, riñón, tejidos linfoides, etc. Los tumores malignos pueden ser de una variedad de subtipos histológicos, por ejemplo, carcinomas, adenocarcinomas, sarcomas, fibroadenocarcinoma, neuroendocrinos o indiferenciados.

Como se utiliza en la presente memoria, el término “compartimento” pretende significar un área o un volumen que separa o aísla algo de otras cosas. Los compartimentos ilustrativos incluyen, pero sin limitación, viales, tubos, pocillos, gotitas, bolos, perlas, recipientes, casillas o áreas o volúmenes de superficie separados por fuerzas físicas tales como flujo de líquido, magnetismo, corriente eléctrica o similares. En una realización, un compartimento es un pocillo de una placa multipocillo, tal como una placa de 96 o 384 pocillos.

Como se utiliza en la presente memoria, un “complejo transposómico” se refiere a una enzima de integración y a un ácido nucleico que incluye un sitio de reconocimiento de integración. Un “complejo transposómico” es un complejo funcional formado por una transposasa y un sitio de reconocimiento de transposasa que es capaz de catalizar una reacción de transposición (véase, por ejemplo, Gunderson y col., WO 2016/130704). Como ejemplos de enzimas de integración se incluyen, pero sin limitación, tales como una integrasa o una transposasa. Como ejemplos de sitios de reconocimiento de integración se incluyen, pero sin limitación, un sitio de reconocimiento de transposasa.

Como se utiliza en la presente memoria, se pretende que el término “ácido nucleico” sea coherente con su uso en la técnica e incluye ácidos nucleicos en anillo de origen natural o análogos funcionales de los mismos. Los análogos funcionales particularmente útiles pueden hibridarse a un ácido nucleico de una manera específica de secuencia o pueden utilizarse como un molde para la replicación de una secuencia de nucleótidos particular. Los ácidos nucleicos de origen natural generalmente tienen una cadena principal que contiene enlaces fosfodiéster. Una estructura análoga puede tener un enlace de cadena principal alternativo que incluye cualquiera de una variedad de los conocidos en la técnica. Los ácidos nucleicos naturales generalmente tienen un azúcar desoxirribosa (p. ej., en el ácido desoxirribonucleico (ADN)) o un azúcar ribosa (p. ej., en el ácido ribonucleico (ARN)). Un ácido nucleico puede contener cualquiera de una variedad de análogos de estos restos de azúcar que se conocen en la técnica. Los ácidos nucleicos pueden incluir bases naturales o no naturales. En este sentido, un ácido desoxirribonucleico natural 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. En la técnica se conocen bases no naturales y útiles que pueden incluirse en un ácido nucleico. Como ejemplos de bases no naturales se incluye un ácido nucleico bloqueado (LNA, locked nucleic acid) y un ácido nucleico con puente (BNA, bridged nucleic acid). En un oligonucleótido de ADN pueden incorporarse bases de LNA y BNA y aumentar la fuerza y especificidad de hibridación del oligonucleótido. El experto en la técnica conoce bases de LNA y BNA y usos de las mismas, y son bases y usos habituales.

Como se utiliza en la presente memoria, cuando el término “diana” se utiliza en referencia a un ácido nucleico, pretende ser un identificador semántico para el ácido nucleico en el contexto de un método o composición expuesto en la presente memoria y no limita necesariamente la estructura o función del ácido nucleico más allá de lo que se indica explícitamente. Un ácido nucleico diana puede ser esencialmente cualquier ácido nucleico de secuencia conocida o desconocida. Puede ser, por ejemplo, un fragmento de ADN genómico o ADNc. La secuenciación puede

dar lugar a la determinación de la secuencia de toda la molécula diana o de una parte de ella. Las dianas pueden proceder de una muestra de ácido nucleico primario, tal como un núcleo. En una realización, las dianas pueden procesarse en moldes adecuados para la amplificación colocando secuencias universales en los extremos de cada fragmento diana. Las dianas también pueden obtenerse a partir de una muestra de ARN primario mediante transcripción inversa en ADNc.

Como se utiliza en la presente memoria, cuando la expresión “secuencia universal” se utiliza para describir una secuencia de nucleótidos, se refiere a una región de secuencia que es común a dos o más moléculas de ácido nucleico donde las moléculas también tienen regiones de secuencia que difieren entre sí. Una secuencia universal que está presente en diferentes miembros de una colección de moléculas puede permitir la captura de múltiples ácidos nucleicos distintos utilizando una población de ácidos nucleicos de captura universal, p. ej., oligonucleótidos de captura, que sean complementarios a una porción de la secuencia universal, p. ej., una secuencia de captura universal. Los ejemplos no limitativos de secuencias de captura universales incluyen secuencias que son idénticas o complementarias a los cebadores P5 y P7. De forma similar, una secuencia universal presente en distintos miembros de una colección de moléculas puede permitir la replicación o amplificación de múltiples ácidos nucleicos distintos utilizando una población de cebadores universales que sean complementarios a una porción de la secuencia de universal, p. ej., una secuencia de anclaje universal. Por lo tanto, un oligonucleótido de captura o un cebador universal incluye una secuencia que puede hibridar específicamente con una secuencia universal.

Los términos “P5” y “P7” pueden utilizarse cuando se hace referencia a cebadores de amplificación, p. ej., un oligonucleótido de captura. Los términos “P5” (P5 prima) y “P7” (P7 prima) se refieren al complemento de P5 y P7, respectivamente. Se entenderá que en los métodos presentados en la presente memoria puede utilizarse cualquier cebador de amplificación adecuado, y que el uso de P5 y P7 son solo realizaciones ilustrativas. Los usos de cebadores de amplificación, tales como P5 y P7, en cubetas de lectura son conocidos en la técnica, como se ilustra en las exposiciones de WO 2007/010251, WO 2006/064199, WO 2005/065814, WO 2015/106941, WO 1998/044151 y WO 2000/018957. Por ejemplo, cualquier cebador de amplificación directo adecuado, ya sea inmovilizado o en solución, puede ser útil en los métodos presentados en la presente memoria para la hibridación a una secuencia complementaria y para la amplificación de una secuencia. De forma similar, cualquier cebador de amplificación inverso adecuado, ya sea inmovilizado o en solución, puede ser útil en los métodos presentados en la presente memoria para la hibridación a una secuencia complementaria y para la amplificación de una secuencia. Un experto en la técnica sabrá cómo diseñar y utilizar secuencias cebadoras que sean adecuadas para la captura y/o amplificación de ácidos nucleicos que se muestran en la presente memoria.

Como se utiliza en la presente memoria, el término “cebador” y sus derivados se refieren en general a cualquier ácido nucleico que pueda hibridar con una secuencia diana de interés. Normalmente, el cebador funciona como un sustrato sobre el que una polimerasa puede polimerizar nucleótidos; sin embargo, en algunas realizaciones, el cebador puede incorporarse a la cadena de ácido nucleico sintetizado y proporcionar un sitio al que pueda hibridarse otro cebador para iniciar la síntesis de una nueva cadena que sea complementaria a la molécula de ácido nucleico sintetizada. El cebador puede incluir cualquier combinación de nucleótidos o análogos de los mismos. En algunas realizaciones, el cebador es un oligonucleótido o polinucleótido monocatenario. Los términos “polinucleótido” y “oligonucleótido” se utilizan indistintamente en la presente memoria en referencia a una forma polimérica de nucleótidos de cualquier longitud, y pueden incluir ribonucleótidos, desoxirribonucleótidos, análogos de los mismos o mezclas de los mismos. Debe entenderse que los términos incluyen, como equivalentes, análogos de ADN o ARN elaborados a partir de análogos de nucleótidos y que son aplicables a polinucleótidos monocatenarios (tales como sentido o antisentido) y bicatenarios. El término, como se utiliza en la presente memoria, también abarca el ADNc, es decir, el ADN complementario o de copia producido a partir de un molde de ARN, por ejemplo, por la acción de la transcriptasa inversa. Este término se refiere únicamente a la estructura primaria de la molécula. Por lo tanto, el término incluye ácido desoxirribonucleico (“ADN”) tri-, bi- y mono-catenario así como ácido ribonucleico (“ARN”) tri-, bi- y mono-catenario.

Como se utiliza en la presente memoria, el término “adaptador” y sus derivados, p. ej., adaptador universal, se refiere en general a cualquier oligonucleótido lineal que pueda ligarse a una molécula de ácido nucleico de la exposición. En algunas realizaciones, el adaptador es sustancialmente no complementario al extremo 3' o al extremo 5' de cualquier secuencia diana presente en la muestra. En algunas realizaciones, las longitudes adecuadas del adaptador están en el intervalo de aproximadamente 10-100 nucleótidos, aproximadamente 12-60 nucleótidos y aproximadamente 15-50 nucleótidos de longitud. En general, el adaptador puede incluir cualquier combinación de nucleótidos y/o ácidos nucleicos. En algunos aspectos, el adaptador puede incluir uno o más grupos escindibles en una o más ubicaciones. En otro aspecto, el adaptador puede incluir una secuencia que sea sustancialmente idéntica, o sustancialmente complementaria, a al menos una porción de un cebador, por ejemplo, un cebador universal. En algunas realizaciones, el adaptador puede incluir un código de barras o una etiqueta para ayudar en la corrección de errores, en la identificación o en la secuenciación posteriores. Los términos “adaptador” y “adaptador” se utilizan indistintamente.

Como se utiliza en la presente memoria, cuando el término “cada uno(a)” se utiliza en referencia a una colección de artículos, pretende identificar un artículo individual en la colección, pero no se refiere necesariamente a cada artículo de la colección, a menos que el contexto indique lo contrario con claridad.

Como se utiliza en la presente memoria, el término “transporte” se refiere al movimiento de una molécula a través de un líquido. El término puede incluir transporte pasivo, tal como el movimiento de moléculas a lo largo de su gradiente de concentración (p. ej., difusión pasiva). El término también puede incluir transporte activo mediante el cual las moléculas pueden moverse a lo largo de su gradiente de concentración o en contra de su gradiente de concentración. Por lo tanto, el transporte puede incluir la aplicación de energía para mover una o más moléculas en una dirección deseada o en una ubicación deseada, tal como un sitio de amplificación.

Como se utiliza en la presente memoria, “amplificar”, “amplificando” o “reacción de amplificación” y sus derivados, se refieren, en general, a cualquier acción o proceso mediante el cual al menos una porción de una molécula de ácido nucleico se replica o copia en al menos una molécula de ácido nucleico adicional. La molécula de ácido nucleico adicional incluye opcionalmente una secuencia que es sustancialmente idéntica o sustancialmente complementaria a al menos alguna porción de la molécula de ácido nucleico molde. La molécula de ácido nucleico molde puede ser monocatenaria o bicatenaria y la molécula de ácido nucleico adicional puede ser, independientemente, monocatenaria o bicatenaria. La amplificación incluye opcionalmente la replicación lineal o exponencial de una molécula de ácido nucleico. En algunas realizaciones, dicha amplificación puede realizarse utilizando condiciones isotérmicas; en otras realizaciones, dicha amplificación puede incluir termociclado. En algunas realizaciones, la amplificación es una amplificación multiplexada que incluye la amplificación simultánea de múltiples secuencias diana en una única reacción de amplificación. En algunas realizaciones, la “amplificación” incluye la amplificación de al menos alguna porción de los ácidos nucleicos basados en ADN y ARN solos o en combinación. La reacción de amplificación puede incluir cualquiera de los procesos de amplificación conocidos por un experto en la técnica. En algunas realizaciones, la reacción de amplificación incluye la reacción en cadena de la polimerasa (PCR, por sus siglas en inglés).

Como se utiliza en la presente memoria, “condiciones de amplificación” y sus derivados, se refiere en general a condiciones adecuadas para amplificar una o más secuencias de ácido nucleico. Dicha amplificación puede ser lineal o exponencial. En algunas realizaciones, las condiciones de amplificación pueden incluir condiciones isotérmicas o de manera alternativa pueden incluir condiciones de termociclado, o una combinación de condiciones isotérmicas y de termociclado. En algunas realizaciones, las condiciones adecuadas para amplificar una o más secuencias de ácido nucleico incluyen condiciones de reacción en cadena de la polimerasa (PCR). Normalmente, las condiciones de amplificación se refieren a una mezcla de reacción que es suficiente para amplificar ácidos nucleicos, tal como una o más secuencias diana, o para amplificar una secuencia diana amplificada ligada a uno o más adaptadores, p. ej., una secuencia diana amplificada ligada a adaptador. En general, las condiciones de amplificación incluyen un catalizador para la amplificación o para la síntesis de ácido nucleico, por ejemplo, una polimerasa; un cebador que posea cierto grado de complementariedad con el ácido nucleico a amplificar; y nucleótidos, tales como desoxirribonucleótido trifosfatos (dNTP) para promover la extensión del cebador una vez hibridado con el ácido nucleico. Las condiciones de amplificación pueden requerir hibridación o emparejamiento de un cebador con un ácido nucleico, extensión del cebador y una etapa de desnaturalización en que el cebador extendido se separa de la secuencia de ácido nucleico que se somete a amplificación. Normalmente, pero no necesariamente, las condiciones de amplificación pueden incluir termociclado; en algunos casos, las condiciones de amplificación incluyen una pluralidad de ciclos donde se repiten las etapas de emparejamiento, extensión y separación. Habitualmente, las condiciones de amplificación incluyen cationes tales como Mg^{2+} o Mn^{2+} y también pueden incluir diversos modificadores de la fuerza iónica.

Como se utiliza en la presente memoria, “re-amplificación” y sus derivados, se refiere en general, a cualquier proceso mediante el cual al menos una porción de una molécula de ácido nucleico amplificada se amplifica adicionalmente a través de cualquier proceso de amplificación adecuado (denominado en algunas realizaciones amplificación “secundaria”) produciendo así una molécula de ácido nucleico reamplificada. La amplificación secundaria no necesita ser idéntica al proceso de amplificación original mediante el cual se produjo la molécula de ácido nucleico amplificada; ni es necesario que la molécula de ácido nucleico reamplificada sea completamente idéntica o completamente complementaria a la molécula de ácido nucleico amplificada; todo lo que se requiere es que la molécula de ácido nucleico reamplificada incluya al menos una porción de la molécula de ácido nucleico amplificada o su complemento. Por ejemplo, la reamplificación puede implicar el uso de diferentes condiciones de amplificación y/o cebadores diferentes, incluidos diferentes cebadores específicos de diana a los de la amplificación primaria.

Como se usa en la presente memoria, la expresión “reacción en cadena de la polimerasa” (“PCR”) se refiere al método de Mullis de las patentes US- 4,683,195 y 4,683,202, que describen un método para aumentar la concentración de un segmento de un polinucleótido de interés en una mezcla de ADN genómico sin clonación ni purificación. Este proceso para amplificar el polinucleótido de interés consiste en introducir un gran exceso de dos cebadores de oligonucleótidos a la mezcla de ADN que contiene el polinucleótido de interés deseado, seguido de una serie de ciclos térmicos en presencia de una polimerasa de ADN. Los dos cebadores son complementarios a sus respectivas cadenas del polinucleótido de doble cadena de interés. La mezcla se desnaturaliza a una temperatura más alta primero y los cebadores se aparean a continuación con secuencias complementarias dentro de la molécula de polinucleótido de interés. Después del apareamiento, los cebadores se extienden con una polimerasa para formar un nuevo par de cadenas complementarias. Las etapas de desnaturalización, emparejamiento de cebadores y extensión de polimerasa pueden repetirse muchas veces (en un proceso denominado termociclado) para obtener una alta concentración de un segmento amplificado del polinucleótido deseado de interés. La longitud del segmento amplificado del polinucleótido de interés deseado (amplicón) viene determinada por las posiciones relativas de los cebadores entre sí y, por lo tanto, esta longitud es un parámetro controlable. Debido a que el proceso se repite, el método se denomina “reacción en

cadena de la polimerasa” (a continuación en la memoria “PCR”). Debido a que los segmentos amplificados deseados del polinucleótido de interés se convierten en las secuencias de ácido nucleico predominantes (en términos de concentración) en la mezcla, se dice que son “amplificadas por PCR”. En una modificación del método descrito anteriormente, las moléculas de ácido nucleico pueden amplificarse por PCR utilizando una pluralidad de pares de cebadores diferentes, en algunos casos, uno o más pares de cebadores por molécula de ácido nucleico diana de interés, formando de este modo una reacción de PCR múltiple.

Como se define en la presente memoria, “amplificación multiplexada” se refiere a la amplificación selectiva y no aleatoria de dos o más secuencias diana dentro de una muestra utilizando al menos un cebador específico de diana. En algunos casos, la amplificación multiplexada se realiza de forma que algunas o todas las secuencias diana se amplifiquen en un solo recipiente de reacción. El “plexi” o “plex” de una amplificación multiplexada dada se refiere generalmente al número de secuencias específicas de diana diferentes que se amplifican durante esa amplificación multiplexada individual. En algunas realizaciones, el plexi puede ser aproximadamente 12-plex, 24-plex, 48-plex, 96-plex, 192-plex, 384-plex, 768-plex, 1536-plex, 3072-plex, 6144-plex o superior. También es posible detectar las secuencias diana amplificadas mediante varias metodologías distintas (por ejemplo, electroforesis en gel seguida de densitometría, cuantificación con un bioanalizador o PCR cuantitativa, hibridación con una sonda marcada; incorporación de cebadores biotinilados seguida de detección de conjugado de enzima y avidina; incorporación de desoxinucleótidos trifosfato marcados con ^{32}P en la secuencia diana amplificada).

Como se utiliza en la presente memoria, “secuencias diana amplificadas” y sus derivados, se refiere en general a una secuencia de ácido nucleico producida mediante la amplificación de las secuencias diana utilizando cebadores específicos de diana y los métodos proporcionados en la presente memoria. Las secuencias diana amplificadas pueden ser del mismo sentido (es decir, la cadena positiva) o de sentido contrario (es decir, la cadena negativa) con respecto a las secuencias diana.

Como se utiliza en la presente memoria, los términos “ligar”, “ligamiento” y sus derivados, se refieren en general, al proceso de unir de manera covalente dos o más moléculas entre sí, por ejemplo, unir de manera covalente entre sí dos o más moléculas de ácido nucleico. En algunos casos, el ligamiento incluye unir muescas entre nucleótidos adyacentes de ácidos nucleicos. En algunas realizaciones, el ligamiento incluye formar un enlace covalente entre un extremo de una primera molécula y un extremo de una segunda molécula de ácido nucleico. En algunas realizaciones, el ligamiento puede incluir formar un enlace covalente entre un grupo fosfato 5' de un ácido nucleico y un grupo hidroxilo 3' de un segundo ácido nucleico formando de este modo una molécula de ácido nucleico ligada. En general, a efectos de la presente exposición, una secuencia diana amplificada puede ligarse a un adaptador para generar una secuencia diana amplificada ligada a adaptador.

Como se utiliza en la presente memoria, “ligasa” y sus derivados, se refiere en general, a cualquier agente capaz de catalizar el ligamiento de dos moléculas de sustrato. En algunos casos, la ligasa incluye una enzima capaz de catalizar la unión de muescas entre nucleótidos adyacentes de un ácido nucleico. En algunas realizaciones, la ligasa incluye una enzima capaz de catalizar la formación de un enlace covalente entre un fosfato 5' de una molécula de ácido nucleico y un hidroxilo 3' de otra molécula de ácido nucleico, formando de este modo una molécula de ácido nucleico ligada. Ligasas adecuadas pueden incluir, aunque no de forma limitativa, ADN T4 ligasa, ARN T4 ligasa y ADN ligasa de *E. coli*.

Como se utiliza en la presente memoria, “condiciones de ligamiento” y sus derivados, se refiere en general, a condiciones adecuadas para ligar dos moléculas entre sí. En algunas realizaciones, las condiciones de ligamiento son adecuadas para sellar muescas o huecos entre los ácidos nucleicos. Como se utiliza en la presente memoria, el término muesca o hueco está en consonancia con el uso del término en la técnica. Habitualmente, una muesca o hueco puede ligarse en presencia de una enzima, tal como una ligasa, a una temperatura y un pH adecuados. En algunas realizaciones, la ADN T4 ligasa puede unir una muesca entre ácidos nucleicos a una temperatura de aproximadamente 70-72 °C.

El término “cubeta de lectura”, como se utiliza en la presente memoria, se refiere a una cámara que comprende una superficie sólida a través de la cual pueden fluir uno o más reactivos líquidos. Se exponen ejemplos de cubetas de lectura y de sistemas fluidos y plataformas de detección relacionados que pueden utilizarse fácilmente en los métodos de la presente exposición, por ejemplo, en Bentley y col., *Nature* 456:53-59 (2008), las patentes WO 04/018497; US-7.057.026; WO 91/06678; WO 07/123744; US-7.329.492; US-7.211.414; US-7.315.019; US-7.405.281, y US 2008/0108082.

Como se utiliza en la presente memoria, cuando el término “amplicón” se utiliza en referencia a un ácido nucleico, significa el producto de copiar el ácido nucleico, en donde el producto tiene una secuencia de nucleótidos que es igual o complementaria a al menos una porción de la secuencia de nucleótidos del ácido nucleico. Un amplicón puede producirse por cualquiera de una variedad de métodos de amplificación que utilicen el ácido nucleico, o un amplicón del mismo, tal como un molde, incluyendo, por ejemplo, extensión por polimerasa, reacción en cadena de la polimerasa (PCR), amplificación de círculo rodante (RCA), extensión por ligamiento o reacción en cadena de ligamiento. Un amplicón puede ser una molécula de ácido nucleico que tenga una sola copia de una secuencia de nucleótidos particular (p. ej., un producto de PCR) o múltiples copias de la secuencia de nucleótidos (p. ej., un producto

concatamérico de RCA). Un primer amplicón de un ácido nucleico diana es normalmente una copia complementaria. Los amplicones posteriores son copias que se crean, después de la generación del primer amplicón, a partir del ácido nucleico diana o del primer amplicón. Un amplicón posterior puede tener una secuencia que sea sustancialmente complementaria a la del ácido nucleico diana o sustancialmente idéntica a la del ácido nucleico diana.

5 Como se utiliza en la presente memoria, la expresión “sitio de amplificación” se refiere a un sitio en o sobre una matriz donde se pueden generar uno o más amplicones. Un sitio de amplificación puede configurarse además para contener, sujetar o unir al menos un amplicón que se genera en el sitio.

10 Como se utiliza en la presente memoria, el término “matriz” se refiere a una población de sitios que pueden diferenciarse entre sí según la ubicación relativa. Las diferentes moléculas que se encuentran en diferentes sitios de una matriz se pueden diferenciar entre sí según las ubicaciones de los sitios en la matriz. Un sitio individual de una matriz puede incluir una o más moléculas de un tipo particular. Por ejemplo, un sitio puede incluir una sola molécula de ácido nucleico diana que tenga una secuencia particular o un sitio puede incluir varias moléculas de ácido nucleico que tengan la misma secuencia (y/o secuencia complementaria, de la misma). Los sitios de una matriz pueden ser diferentes casillas situadas en el mismo sustrato. Como casillas ilustrativas se incluyen, sin limitación, pocillos en un sustrato, perlas (u otras partículas) en o sobre un sustrato, proyecciones desde un sustrato, crestas sobre un sustrato o canales en un sustrato. Los sitios de una matriz pueden ser diferentes sustratos, cada uno con una molécula diferente. Se pueden identificar diferentes moléculas unidas a diferentes sustratos según las ubicaciones de los sustratos en una superficie a la que están asociados los sustratos o según las ubicaciones de los sustratos en un líquido o gel. Como matrices ilustrativas en que se ubican distintos sustratos en una superficie se incluyen, sin limitación, las que tienen perlas en pocillos.

25 Como se utiliza en la presente memoria, cuando el término “capacidad” se utiliza en referencia a un sitio y a un material de ácido nucleico, significa la cantidad máxima de material de ácido nucleico que puede ocupar el sitio. Por ejemplo, el término puede referirse al número total de moléculas de ácido nucleico que pueden ocupar el sitio en una condición particular. También pueden utilizarse otras medidas incluyendo, por ejemplo, la masa total de material de ácido nucleico o el número total de copias de una secuencia de nucleótidos particular que puede ocupar el sitio en una condición particular. Normalmente, la capacidad de un sitio para un ácido nucleico diana será sustancialmente equivalente a la capacidad del sitio para amplicones del ácido nucleico diana.

30 Como se utiliza en la presente memoria, la expresión “agente de captura” se refiere a un material, sustancia química, molécula o resto de la misma, que es capaz de fijarse, retenerse o unirse a una molécula diana (p. ej., un ácido nucleico diana). Los ejemplos de agentes de captura incluyen, sin limitación, un ácido nucleico de captura (también denominado en la presente memoria oligonucleótido de captura) que es complementario a al menos una porción de un ácido nucleico diana, un miembro de un par de unión receptor-ligando (por ejemplo, avidina, estreptavidina, biotina, lectina, hidrato de carbono, proteína de unión a ácido nucleico, epítipo, anticuerpo, etc.) capaz de unirse a un ácido nucleico diana (o un resto de enlace unido al mismo), o un reactivo químico capaz de formar un enlace covalente con un ácido nucleico diana (o resto de enlace unido al mismo).

40 Como se utiliza en la presente memoria, la expresión “población clonal” se refiere a una población de ácidos nucleicos que es homogénea con respecto a una secuencia de nucleótidos particular. La secuencia homogénea tiene normalmente una longitud de al menos 10 nucleótidos, pero puede ser incluso más larga incluyendo, por ejemplo, una longitud de al menos 50, 100, 250, 500 o 1000 nucleótidos. Una población clonal puede proceder de un solo ácido nucleico diana o ácido nucleico molde. Normalmente, todos los ácidos nucleicos de una población clonal tendrán la misma secuencia de nucleótidos. Se entenderá que, en una población clonal, puede producirse un pequeño número de mutaciones (p. ej., debidas a artefactos de amplificación) sin apartarse de la clonalidad.

50 Como se utiliza en la presente memoria, en el contexto de una composición, un artículo, un ácido nucleico o un núcleo, “proporcionar” significa preparar la composición, el artículo, el ácido nucleico o el núcleo, adquirir la composición, el artículo, el ácido nucleico o el núcleo, u obtener de cualquier otra manera el compuesto, la composición, el artículo o el núcleo.

55 El término “y/o” significa uno o todos los elementos enumerados o una combinación de cualquiera de dos o más de los elementos enumerados.

60 Las palabras “preferido(a)” y “preferiblemente” se refieren a realizaciones de la invención que pueden proporcionar determinados beneficios, en determinadas circunstancias. Sin embargo, en las mismas circunstancias o en otras, también pueden preferirse otras realizaciones. Además, la mención de una o más realizaciones preferidas no implica que otras realizaciones no sean útiles y no pretende excluir otras realizaciones del alcance de la invención.

Los términos “comprende” y variaciones del mismo no tienen un significado limitativo cuando estos términos aparecen en la descripción y en las reivindicaciones.

Se entiende que siempre que en la presente memoria se describan realizaciones con las frases “incluye”, “que incluye” o “incluyendo” y similares, también se proporcionan realizaciones análogas descritas en términos de “que consiste en” y/o “que consiste esencialmente en”.

5 A menos que se especifique de cualquier otra manera, “un”, “uno”, “una”, “el”, “la” y “al menos un, uno, una”, se usan indistintamente y significan uno(a) o más de uno(a).

También en la presente memoria, las enumeraciones de intervalos numéricos mediante extremos incluyen todos los números comprendidos dentro de ese intervalo (p. ej., 1 a 5 incluye 1, 1,5, 2, 2,75, 3, 3,80, 4, 5, etc.).

10 En cualquier método expuesto en la presente memoria que incluya distintas etapas, las etapas pueden realizarse en cualquier orden factible. Y, según sea apropiado, puede realizarse cualquier combinación de dos o más etapas simultáneamente.

15 A lo largo de la presente memoria descriptiva, la referencia a “una realización”, a “determinadas realizaciones” o a “algunas realizaciones” etc., significa que, en al menos una realización de la exposición, se incluye un rasgo, una configuración, una composición o una característica particular descrita en relación con la realización. Por lo tanto, la aparición de dichas expresiones en diversos lugares a lo largo de esta memoria descriptiva, no se refieren necesariamente a la misma realización de la exposición. Además, los rasgos, configuraciones, composiciones o características particulares pueden combinarse de cualquier manera adecuada en una o más realizaciones.

Breve descripción de las figuras

25 La siguiente descripción detallada de las realizaciones ilustrativas de la presente exposición puede entenderse mejor cuando se lee junto con los siguientes dibujos.

La Figura 1 muestra un diagrama de bloque general de un método ilustrativo general de la indexación combinatoria de célula individual según la presente exposición.

30 La Figura 2 muestra un dibujo esquemático de una realización del método de indexación combinatoria de célula individual ilustrado en líneas generales en la Figura 1.

La Figura 3 muestra un dibujo esquemático de una realización ilustrativa de una molécula adaptadora de fragmentos después de la amplificación lineal.

35 La Figura 4 muestra un dibujo esquemático de una realización ilustrativa de una molécula adaptadora de fragmentos después de la adición de adaptadores universales.

40 Los dibujos esquemáticos no son necesariamente a escala. Los mismos números que se utilizan en las figuras se refieren a los mismos componentes, etapas y similares. Sin embargo, se entenderá que el uso de un número para referirse a un componente en una figura dada no pretende estar limitado al componente en otra figura etiquetada con el mismo número. Además, el uso de números distintos para referirse a componentes no pretende indicar que los distintos componentes numerados no puedan ser iguales o similares a otros componentes numerados.

45 Descripción detallada de realizaciones ilustrativas

El método proporcionado en la presente memoria incluye proporcionar núcleos aislados de una pluralidad de células (Figura 1, bloque 12). Las células pueden ser de cualquier organismo u organismos y de cualquier tipo de célula o de cualquier tejido del organismo u organismos. El método puede incluir además la disociación de las células (Figura 2, bloque i) y/o el aislamiento de los núcleos (Figura 2, bloque ii). El experto en la técnica conoce métodos para aislar los núcleos de las células y estos son métodos habituales. El número de núcleos puede ser de al menos 2. El límite superior depende de las limitaciones prácticas del equipo (p. ej., placas multipocillo) utilizadas en otras etapas del método descrito en la presente memoria. Por ejemplo, en una realización, el número de núcleos puede ser no superior a 1.000.000.000, no superior a 100.000.000, no superior a 10.000.000, no superior a 1.000.000, no superior a 10.000 o no superior a 1.000. La persona experta reconocerá que las moléculas de ácido nucleico en cada núcleo representan el complemento genético completo de un organismo, y son moléculas de ADN genómico que incluyen secuencias tanto intrónicas como exónicas, así como secuencias reguladoras no codificantes tales como secuencias promotoras y potenciadoras.

60 En una realización, los núcleos incluyen nucleosomas unidos a ADN genómico. Dichos núcleos pueden ser útiles en métodos que no determinan la secuencia de ADN del genoma completo de una célula, tal como sciATAC-seq. (single-cell ATAC sequencing using combinatorial indexing, secuenciación ATAC de célula individual mediante indexación combinatoria). En otra realización, los núcleos aislados se someten a condiciones que empobrecen los núcleos de nucleosomas, generando núcleos con empobrecimiento de nucleosomas (Figura 1, bloque 13 y Figura 2, bloque ii).

65 Dichos núcleos pueden ser útiles en métodos cuyo objeto es determinar la secuencia completa de ADN genómico de una célula. En una realización, las condiciones utilizadas para el empobrecimiento de nucleosomas mantienen la

integridad de los núcleos aislados. La persona experta conoce métodos para generar núcleos empobrecidos en nucleosomas (véase, por ejemplo, Vitak y col., 2017, *Nature Methods*, 14(3):302-308). En una realización, las condiciones son un tratamiento químico que incluye un tratamiento con un agente caotrópico capaz de interrumpir las interacciones ácido nucleico-proteína. Un ejemplo de un agente caotrópico útil incluye, pero no se limita a, 5 diyodosalicilato de litio. En otra realización, las condiciones son un tratamiento químico que incluye un tratamiento con un detergente capaz de interrumpir las interacciones ácido nucleico-proteína. Un ejemplo de un detergente útil incluye, pero no se limita a, dodecilsulfato de sodio (SDS). En algunas realizaciones, cuando se utiliza un detergente tal como SDS, las células de las que se aíslan los núcleos se tratan con un agente reticulante antes del aislamiento. Un ejemplo útil de un agente reticulante incluye, pero no se limita a, formaldehído.

El método proporcionado en la presente memoria incluye la distribución de subconjuntos de los núcleos, tal como núcleos con empobrecimiento de nucleosomas, en una primera pluralidad de compartimentos (Figura 1, bloque 14 y Figura 2, esquema de la izquierda). El número de núcleos presentes en un subconjunto, y por lo tanto en cada compartimento, puede ser de al menos 1. En una realización, el número de núcleos presentes en un subconjunto no es superior a 2.000. El experto en la técnica conoce métodos para distribuir los núcleos en subconjuntos y estos métodos son habituales. Los ejemplos incluyen, pero no se limitan a, clasificación de núcleos activados por fluorescencia (FANS).

Cada compartimento incluye un complejo transposómico. El complejo transposómico, una transposasa unida a un sitio de reconocimiento de transposasa, puede insertar el sitio de reconocimiento de transposasa en un ácido nucleico diana dentro de un núcleo en un proceso denominado a veces “tagmentación”. En algunos acontecimientos de inserción de este tipo, una cadena del sitio de reconocimiento de transposasa puede transferirse al ácido nucleico diana. Dicha cadena se denomina “cadena transferida”. En una realización, un complejo transposómico incluye una transposasa dimérica que tiene dos subunidades y dos secuencias transposónicas no contiguas. En otra realización, una transposasa incluye una transposasa dimérica que tiene dos subunidades y una secuencia transposónica contigua.

Algunas realizaciones pueden incluir el uso de una transposasa Tn5 hiperactiva y un sitio de reconocimiento de transposasa de tipo Tn5 (Goryshin y Reznikoff, *J. Biol. Chem.*, 273:7367 (1998)), o una transposasa MuA y un sitio de reconocimiento de transposasa Mu que comprende secuencias terminales R1 y R2 (Mizuuchi, K., *Cell*, 35: 785, 1983; Savilahti, H, y col., *EMBO J.*, 14: 4893, 1995). El experto en la técnica también puede utilizar secuencias de Tn5 de extremos en mosaico (EM).

Más ejemplos de sistemas de transposición que se pueden utilizar con determinadas realizaciones de las composiciones y métodos proporcionados en la presente memoria incluyen Tn552 de *Staphylococcus aureus* (Colegio y col., *J. Bacteriol.*, 183: 2384-8, 2001; Kirby C y col., *Mol. Microbiol.*, 43: 173-86, 2002), Ty1 (Devine y Boeke, *Nucleic Acids Res.*, 22: 3765-72, 1994 y la Publicación Internacional WO 95/23875), el Transposón Tn7 (Craig, N L, *Science*. 271: 1512, 1996; Craig, N L, Revisado en: *Curr Top Microbiol Immunol.*, 204:27-48, 1996), Tn/O e IS10 (Kleckner N, y col, *Curr Top Microbiol Immunol.*, 204:49-82, 1996), la transposasa Mariner (Lampe D J, y col., *EMBO J.*, 15: 5470-9, 1996), Tc1 (Plasterk R H, *Curr. Topics Microbiol. Immunol.*, 204: 125-43, 1996), elemento P (Gloor, G B, *Methods Mol. Biol.*, 260: 97-114, 2004), Tn3 (Ichikawa y Ohtsubo, *J Biol. Chem.* 265:18829-32, 1990), secuencias de inserción bacterianas (Ohtsubo y Sekine, *Curr. Top. Microbiol. Immunol.* 204: 1-26, 1996), retrovirus (Brown, y col., *Proc Natl Acad Sci USA*, 86:2525-9, 1989), y retrotransposición de levadura (Boeke y Corces, *Annu Rev Microbiol.* 43:403-34, 1989). Algunos ejemplos adicionales incluyen IS5, Tn10, Tn903, IS911, y versiones modificadas por ingeniería genética de enzimas de la familia de las transposasas (Zhang y col., (2009) *PLoS Genet.* 5:e1000689. Epub 16 de oct de 2009; Wilson C. y col. (2007) *J. Microbiol. Methods* 71:332-5).

Otros ejemplos de integrasas que pueden utilizarse con los métodos y las composiciones que se proporcionan en la presente memoria incluyen integrasas retrovíricas y secuencias de reconocimiento de integrasas de dichas integrasas retrovíricas, tales como integrasas de HIV-1, HIV-2, SIV, PFV-1, RSV.

Las secuencias transposónicas útiles con los métodos y composiciones descritos en la presente memoria se proporcionan en la publicación de solicitud de patente US- 2012/0208705, publicación de solicitud de patente US- 2012/0208724 y publicación de solicitud de patente internacional WO 2012/061832. En algunas realizaciones, una secuencia transposónica incluye un primer sitio de reconocimiento de transposasa, un segundo sitio de reconocimiento de transposasa y un índice presente entre los dos sitios de reconocimiento de transposasa.

Algunos complejos transposómicos útiles en la presente memoria incluyen una transposasa que tiene dos secuencias transposónicas. En algunas de estas realizaciones, las dos secuencias transposónicas no están ligadas entre sí, en otras palabras, las secuencias transposónicas no son contiguas entre sí. En la técnica se conocen ejemplos de dichos transposomas (véase, por ejemplo, la publicación de solicitud de patente US- 2010/0120098).

En algunas realizaciones, un complejo transposómico incluye un ácido nucleico de secuencia transposónica que se une a dos subunidades de transposasa para formar un “complejo en bucle” o un “transposoma en bucle”. En un ejemplo, un transposoma incluye una transposasa dimérica y una secuencia transposónica. Los complejos en bucle pueden garantizar que los transposones se inserten en el ADN diana mientras se mantiene la información de

ordenación del ADN diana original y sin fragmentar el ADN diana. Como se apreciará, en un ácido nucleico diana, las estructuras en bucle pueden insertar secuencias de ácido nucleico deseadas, tales como índices, manteniendo la conectividad física del ácido nucleico diana. En algunas realizaciones, la secuencia transposónica de un complejo transposómico en bucle puede incluir un sitio de fragmentación de modo que la secuencia transposónica se pueda fragmentar para crear un complejo transposómico que comprenda dos secuencias transposónicas. Dichos complejos transposómicos son útiles para garantizar que los fragmentos de ADN diana adyacentes, en los cuales se insertan los transposones, reciban combinaciones de códigos que puedan ensamblarse sin ambigüedades en una fase posterior del ensayo.

Un complejo transposómico también incluye al menos una secuencia índice, también denominada índice de transposasa. La secuencia índice está presente como parte de la secuencia transposónica. En una realización, la secuencia índice puede estar presente en una cadena transferida, la cadena del sitio de reconocimiento de la transposasa que se transfiere al ácido nucleico diana. Una secuencia índice, también denominada etiqueta o código de barras, es útil como marcador característico del compartimento en que estaba presente un ácido nucleico diana particular. La secuencia índice de un complejo transposómico es diferente para cada compartimento. Por consiguiente, en esta realización, un índice es una etiqueta de secuencia de ácido nucleico que se une a cada uno de los ácidos nucleicos diana presentes en un compartimento particular, cuya presencia es indicativa del compartimento en que se encuentra una población de núcleos en esta fase del método, o se utiliza para identificarlo.

Una secuencia índice puede tener hasta 20 nucleótidos de longitud, p. ej., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20. Una etiqueta de cuatro nucleótidos ofrece la posibilidad de multiplexar 256 muestras en la misma matriz, mientras que una etiqueta de seis bases permite procesar 4096 muestras en la misma matriz.

En una realización, la cadena transferida también puede incluir una secuencia universal, una primera secuencia cebadora de secuenciación o una combinación de las mismas. En la presente memoria se describen secuencias universales y secuencias cebadoras de secuenciación. Por lo tanto, en algunas realizaciones en donde la cadena transferida se transfiere a los ácidos nucleicos diana, los ácidos nucleicos diana incluyen un índice de transposasa y también incluyen una secuencia universal, una primera secuencia cebadora de secuenciación o una combinación de las mismas.

En una realización, los nucleótidos de citosina de una cadena transferida están metilados. En otra realización, los nucleótidos de una cadena transferida no contienen citosina. Dicha cadena transferida, y cualquier secuencia presente en la cadena transferida, incluyendo una secuencia índice de transposasa, una secuencia universal y/o una primera secuencia cebadora de secuenciación, puede denominarse empobrecida en citosina. El uso de secuencias de nucleótidos empobrecidas en citosina en un complejo transposómico no tiene ningún impacto significativo sobre la eficacia de la transposasa.

El método también incluye la generación de núcleos indexados (Figura 1, bloque 15 y Figura 2, bloque iii). En una realización, la generación de núcleos indexados incluye la fragmentación de ácidos nucleicos presentes en los subconjuntos de núcleos con empobrecimiento de nucleosomas (p. ej., los ácidos nucleicos presentes en cada compartimento) en una pluralidad de fragmentos de ácido nucleico. En una realización, la fragmentación de ácidos nucleicos se consigue utilizando un sitio de fragmentación presente en los ácidos nucleicos. Normalmente, los sitios de fragmentación se introducen en los ácidos nucleicos diana utilizando un complejo transposómico. Por ejemplo, un complejo transposómico en bucle puede incluir un sitio de fragmentación. Puede utilizarse un sitio de fragmentación para escindir la asociación física, pero no la informativa, entre las secuencias índice que se han insertado en un ácido nucleico diana. La escisión puede ser por medios bioquímicos, químicos u otros. En algunas realizaciones, un sitio de fragmentación puede incluir un nucleótido o una secuencia de nucleótidos que puede fragmentarse por diversos medios. Los ejemplos de sitios de fragmentación incluyen, aunque no de forma limitativa, un sitio de endonucleasa de restricción, al menos un ribonucleótido escindible con una ARNasa, análogos de nucleótidos escindibles en presencia de un determinado agente químico, un enlace diol escindible por tratamiento con peryodato, un grupo disulfuro escindible con un agente reductor químico, un resto escindible que se puede someter a escisión fotoquímica y un péptido escindible mediante una enzima peptidasa u otros medios adecuados (véase, por ejemplo, la publicación de solicitud de patente US- 2012/0208705, publicaciones de solicitud de patente US- 2012/0208724 y WO 2012/061832. El resultado de la fragmentación es una población de núcleos indexados, conteniendo cada núcleo fragmentos de ácido nucleico, donde los fragmentos de ácido nucleico se incluyen en al menos una cadena de la secuencia índice indicativa del compartimento particular.

Los núcleos indexados de múltiples compartimentos pueden combinarse (Figura 1, bloque 16 y Figura 2, esquema de la izquierda). Por ejemplo, se combinan los núcleos indexados de 2 a 96 compartimentos (cuando se utiliza una placa de 96 pocillos) o de 2 a 384 compartimentos (cuando se utiliza una placa de 384 pocillos). Los subconjuntos de estos núcleos indexados combinados, denominados en la presente memoria núcleos indexados agrupados, se distribuyen después en una segunda pluralidad de compartimentos. El número de núcleos presentes en un subconjunto y, por lo tanto, en cada compartimento, se basa en parte en el deseo de reducir las colisiones de índice, que es la presencia de dos núcleos que tienen el mismo índice de transposasa que terminan en el mismo compartimento en esta etapa del proceso. En esta realización, el número de núcleos presentes en un subconjunto puede ser de 2 a 30, tal como de 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29 o 30. En una realización,

el número de núcleos presentes en un subconjunto es de 20 a 24, tal como de 22. El experto en la técnica conoce métodos para distribuir los núcleos en subconjuntos y estos métodos son habituales. Los ejemplos incluyen, pero no se limitan a, clasificación de núcleos activados por fluorescencia (FANS).

5 Los núcleos indexados distribuidos se tratan para identificar nucleótidos metilados (Figura 1, bloque 17 y Figura 2, bloque iv). La metilación de sitios, tales como secuencias de dinucleótidos CpG, puede medirse utilizando cualquiera de las diversas técnicas utilizadas en la materia para el análisis de tales sitios. Un método útil es la identificación de secuencias de dinucleótidos CpG metiladas. La identificación de secuencias de dinucleótidos CpG metiladas se determina utilizando tecnologías basadas en la conversión de citosina, que se basan en la modificación química dependiente del estado de metilación de las secuencias CpG dentro del ADN genómico aislado, o fragmentos del mismo, seguido de un análisis de secuencia de ADN. Los reactivos químicos que pueden diferenciar entre secuencias de dinucleótidos CpG metilados y no metilados incluyen hidracina, que escinde el ácido nucleico, y bisulfito. El tratamiento con bisulfito seguido de hidrólisis alcalina convierte específicamente la citosina no metilada en uracilo, dejando la 5-metilcitosina sin modificar como se describe en Olek A., 1996, *Nucleic Acids Res.* 24:5064-6 o Frommer y col., 1992, *Proc. Natl. Acad. Sci. USA* 89:1827-1831. El ADN tratado con bisulfito se puede analizar posteriormente mediante técnicas moleculares, tales como amplificación, secuenciación y detección por PCR, incluida la hibridación de oligonucleótidos (p. ej., utilizando micromatrices de ácidos nucleicos). En una realización, los núcleos indexados en cada compartimento se exponen a condiciones para el tratamiento con bisulfito. El tratamiento de ácidos nucleicos con bisulfito es conocido por el experto en la técnica y es un tratamiento habitual. En una realización, el tratamiento con bisulfito convierte los residuos de citosina no metilados de los dinucleótidos CpG en residuos de uracilo y deja los residuos de 5-metilcitosina inalterados. El tratamiento con bisulfito da como resultado fragmentos de ácido nucleico tratados con bisulfito.

25 Después de la generación de los fragmentos de ácido nucleico tratados con bisulfito, los fragmentos se modifican para incluir nucleótidos adicionales en uno o ambos extremos (Figura 1, bloque 18 y Figura 2, bloques v y vi). En una realización, la modificación incluye someter los fragmentos de ácido nucleico tratados con bisulfito a una amplificación lineal utilizando una pluralidad de cebadores. Cada cebador incluye al menos dos regiones; una secuencia de nucleótidos universal en el extremo 5' y una secuencia de nucleótidos aleatoria en el extremo 3'. La secuencia de nucleótidos universal es idéntica en cada cebador y, en una realización, incluye una segunda secuencia cebadora de secuenciación (también denominada cebador 2 de lectura en la Figura 2 (bloque vii)). La región de secuencia de nucleótidos aleatoria se utiliza para que haya al menos un cebador que sea complementario a cada secuencia de los fragmentos de ácido nucleico tratados con bisulfito. El número de nucleótidos aleatorios que puede utilizarse para aumentar la probabilidad de cobertura completa a un nivel deseado puede determinarse utilizando métodos habituales y puede ser de 6 a 12 nucleótidos aleatorios, tal como de 9 nucleótidos aleatorios. En una realización, el número de ciclos está limitado a no más de 10 ciclos, tal como a 9 ciclos, 8 ciclos, 7 ciclos, 6 ciclos, 5 ciclos, 4 ciclos, 3 ciclos, 2 ciclos o 1 ciclo. El resultado de la amplificación lineal es moléculas adaptadoras de fragmentos amplificadas. En la Figura 3 se muestra un ejemplo de una molécula adaptadora de fragmentos. La molécula adaptadora de fragmentos 30 incluye nucleótidos que se originan a partir de la cadena transferida del complejo transposómico 31 y 32, que incluye un índice de transposasa y una secuencia universal que puede utilizarse para la amplificación y/o secuenciación. La molécula adaptadora de fragmentos también incluye los nucleótidos que se originan a partir del ADN genómico de un núcleo 33, la región de secuencia de nucleótidos aleatoria 34 y la secuencia de nucleótidos universal 35.

45 A la amplificación lineal le sigue una reacción de amplificación exponencial, tal como una PCR, para modificar después los extremos de la molécula adaptadora de fragmentos antes de la inmovilización y la secuenciación. Esta etapa da como resultado la indexación de las moléculas adaptadoras de fragmentos por PCR (Figura 1, bloque 19). Las secuencias universales 31, 32 y/o 35 presentes en los extremos de la molécula adaptadora de fragmentos pueden utilizarse para la unión de secuencias de anclaje universales que pueden servir como cebadores y extenderse en una reacción de amplificación. Por lo general, se utilizan dos cebadores diferentes. Un cebador se hibrida con secuencias universales en el extremo 3' de una cadena de la molécula adaptadora de fragmentos, y un segundo cebador se hibrida con secuencias universales en el extremo 3' de la otra cadena de la molécula adaptadora de fragmentos. Por lo tanto, la secuencia de anclaje de cada cebador puede ser diferente. Cada uno de los cebadores adecuados puede incluir secuencias universales adicionales, tal como una secuencia de captura universal y otra secuencia índice. Dado que cada cebador puede incluir un índice, esta etapa da como resultado la adición de una o dos secuencias índice, p. ej., un segundo y un tercer índice opcional. Las moléculas adaptadoras de fragmentos que tienen el segundo y el tercer índice opcional se denominan moléculas adaptadoras de fragmentos de doble índice. El segundo y tercer índice pueden ser el complemento inverso entre sí, o el segundo y tercer índice pueden tener secuencias que no sean complementos inversos entre sí. Estas secuencias de segundo y tercer índice opcionales son únicas para cada compartimento en que se colocaron los núcleos indexados distribuidos antes del tratamiento con bisulfito sódico. El resultado de esta amplificación por PCR es una pluralidad o biblioteca de moléculas adaptadoras de fragmentos que tienen una estructura similar o idéntica a la de la molécula adaptadora de fragmentos que se muestra en la Figura 2, bloque vii.

65 En otra realización, la modificación incluye someter los fragmentos de ácido nucleico tratados con bisulfito a condiciones que den como resultado el ligamiento de secuencias adicionales en ambos extremos de los fragmentos. En una realización, puede utilizarse ligamiento de extremos romos. En otra realización, los fragmentos se preparan

con nucleótidos salientes individuales, por ejemplo, mediante actividad de determinados tipos de ADN polimerasa tal como la Taq polimerasa o la exo minus Klenow polimerasa que tiene una actividad transferasa terminal no dependiente de molde que añade un solo desoxinucleótido, por ejemplo, desoxiadenosina (A) a los extremos 3' de los fragmentos de ácido nucleico tratados con bisulfito. Dichas enzimas pueden utilizarse para añadir un solo nucleótido 'A' en el extremo como 3' terminal de cada cadena de los fragmentos. Por lo tanto, se podría añadir una 'A' al extremo 3' de cada cadena de los fragmentos diana monocatenarios por reacción con Taq o Klenow exo minus polimerasa, mientras que las secuencias adicionales que se añadirán a cada extremo del fragmento pueden incluir un saliente 'T' compatible presente en el extremo 3' de cada región de ácido nucleico bicatenario que se va a añadir. Esta modificación de los extremos también impide el auto-ligamiento de los ácidos nucleicos, de modo que, en esta realización, existe una tendencia hacia la formación de fragmentos de ácido nucleico tratados con bisulfito flanqueados por las secuencias que se añaden.

La fragmentación de moléculas de ácido nucleico mediante los métodos descritos en la presente memoria da como resultado fragmentos con una mezcla heterogénea de extremos romos y salientes en 3' y 5'. Por lo tanto, es deseable reparar los extremos del fragmento utilizando métodos o kits (tal como el kit de reparación de extremos del terminador de ADN Lucigen) conocidos en la técnica para generar extremos que sean óptimos para la inserción, por ejemplo, en sitios romos de vectores de clonación. En una realización particular, los extremos de los fragmentos de la población de ácidos nucleicos son extremos romos. Más particularmente, los extremos de los fragmentos tienen extremos romos y están fosforilados. El resto fosfato puede introducirse mediante tratamiento enzimático, por ejemplo, utilizando polinucleótido cinasa.

En una realización, los fragmentos de ácido nucleico tratados con bisulfito se tratan ligando primero adaptadores universales idénticos (también denominados “adaptadores con emparejamiento incorrecto”, cuyas características generales se describen en Gormley y col., US-7,741,463, y en Bignell y col, US-8,053,192) a los extremos 5' y 3' de los fragmentos de ácido nucleico tratados con bisulfito para formar moléculas adaptadoras de fragmentos. En una realización, el adaptador universal incluye todas las secuencias necesarias para la secuenciación, incluida la inmovilización de las moléculas adaptadoras de fragmentos en una matriz. Dado que los ácidos nucleicos que se van a secuenciar proceden de células individuales, la amplificación adicional de las moléculas adaptadoras de fragmentos es útil para conseguir un número suficiente de moléculas adaptadoras de fragmentos para la secuenciación.

En otra realización, cuando el adaptador universal no incluye todas las secuencias necesarias para la secuenciación, se puede utilizar una etapa de PCR para modificar adicionalmente el adaptador universal presente en cada molécula adaptadora de fragmentos antes de la inmovilización y secuenciación. Por ejemplo, se lleva a cabo una reacción inicial de extensión del cebador utilizando una secuencia de anclaje universal complementaria a una secuencia universal presente en la molécula adaptadora de fragmentos, en que se forman productos de extensión complementarios a ambas cadenas de cada molécula adaptadora de fragmentos individual. Normalmente, la PCR añade secuencias universales adicionales, tal como una secuencia de captura universal y otra secuencia índice. Dado que cada cebador puede incluir un índice, esta etapa da como resultado la adición de una o dos secuencias índice, p. ej., un segundo y un tercer índice opcional, y la indexación de las moléculas adaptadoras de fragmentos mediante ligamiento del adaptador (Figura 1, bloque 19). Las moléculas adaptadoras de fragmentos resultantes se denominan moléculas adaptadoras de fragmentos de doble índice.

Después de añadir los adaptadores universales, ya sea mediante un método de ligamiento de una sola etapa de un adaptador universal que incluya todas las secuencias necesarias para la secuenciación, o mediante un método de ligamiento de dos etapas de un adaptador universal y posterior amplificación por PCR para modificar adicionalmente el adaptador universal, la molécula adaptadora de fragmentos final incluirá una secuencia de captura universal, una segunda secuencia índice y una tercera secuencia índice opcional. Estos índices son análogos al segundo y tercer índice descritos en la producción de adaptadores de fragmentos de doble índice por amplificación lineal. El segundo y tercer índice pueden ser el complemento inverso entre sí, o el segundo y tercer índice pueden tener secuencias que no sean complementos inversos entre sí. Estas secuencias de segundo y tercer índice opcional son únicas para cada compartimento en que se colocaron los núcleos indexados distribuidos antes del tratamiento con bisulfito sódico. El resultado de añadir adaptadores universales a cada extremo es una pluralidad o una biblioteca de moléculas adaptadoras de fragmentos que tienen una estructura similar o idéntica a la de la molécula adaptadora de fragmentos 40 mostrada en la Figura 4. La molécula adaptadora de fragmentos 40 incluye una secuencia de captura 41 y 48, también denominada adaptador de cubeta de lectura 3' (p. ej., P5) y adaptador de cubeta de lectura 5' (p. ej., P7'), respectivamente, y un índice 42 y 47, tal como i5 e i7. La molécula adaptadora de fragmentos 40 también incluye nucleótidos que se originan a partir de la cadena transferida del complejo transposómico 43, que incluye un índice de transposasa 44 y una secuencia universal 45 que puede utilizarse para la amplificación y/o secuenciación. La molécula adaptadora de fragmentos también incluye los nucleótidos que se originan a partir del ADN genómico de un núcleo 46.

Las moléculas adaptadoras de fragmentos de doble índice resultantes proporcionan en conjunto una biblioteca de ácidos nucleicos que se pueden inmovilizar y posteriormente secuenciar. El término biblioteca se refiere a la colección de fragmentos de células individuales que contienen secuencias universales conocidas en sus extremos 3' y 5'.

Después de que los fragmentos de ácido nucleico tratados con bisulfito se modifiquen para incluir nucleótidos adicionales, las moléculas adaptadoras de fragmentos de doble índice pueden someterse a condiciones que se seleccionen con respecto a un intervalo de tamaño predeterminado, tal como de 150 a 400 nucleótidos de longitud, tal como de 150 hasta 300 nucleótidos. Las moléculas adaptadoras de fragmentos de doble índice resultantes se agrupan y, opcionalmente, pueden someterse a un proceso de limpieza para mejorar la pureza de las moléculas de ADN mediante la eliminación de al menos una porción de los adaptadores o cebadores universales no incorporados. Puede utilizarse cualquier proceso de limpieza adecuado, tal como electroforesis, cromatografía de exclusión por tamaño, o similar. En algunas realizaciones, se pueden emplear perlas paramagnéticas de inmovilización reversible en fase sólida para separar las moléculas de ADN deseadas de adaptadores o cebadores universales no unidos, y para seleccionar ácidos nucleicos en función del tamaño. Las perlas paramagnéticas de inmovilización reversible en fase sólida están disponibles comercialmente en Beckman Coulter (Agencourt AMPure XP), Thermofisher (MagJet), Omega Biotek (Mag-Bind), Promega Beads (Promega) y Kapa Biosystems (Kapa Pure Beads).

La pluralidad de moléculas adaptadoras de fragmentos puede prepararse para la secuenciación. Después de agrupar las moléculas adaptadoras de fragmentos, estas se inmovilizan y amplifican antes de la secuenciación (Figura 1, bloque 20). En la técnica se conocen métodos para unir moléculas adaptadoras de fragmentos de una o más fuentes a un sustrato. Asimismo, los métodos para amplificar moléculas adaptadoras de fragmentos inmovilizadas incluyen, pero no se limitan a, amplificación de puente y exclusión cinética. Los métodos de inmovilización y amplificación antes de la secuenciación se describen, por ejemplo, en Bignell y col. (US-8,053,192), Gunderson y col. (WO2016/130704), Shen y col. (US-8,895,249) y Pipenburg y col. (US 9.309.502).

Para la secuenciación, en la preparación puede inmovilizarse una muestra agrupada. La secuenciación puede realizarse como una matriz de moléculas individuales, o puede amplificarse antes de la secuenciación. La amplificación puede llevarse a cabo utilizando uno o más cebadores inmovilizados. El cebador o cebadores inmovilizado(s) puede(n) ser un césped sobre una superficie plana o sobre un grupo de perlas. El conjunto de perlas puede aislarse en una emulsión con una perla individual en cada "compartimento" de la emulsión. A una concentración de solo un molde por "compartimento", únicamente se amplifica un molde individual en cada perla.

La expresión "amplificación en fase sólida" como se utiliza en la presente memoria se refiere a cualquier reacción de amplificación de ácido nucleico llevada a cabo en, o en asociación con, un soporte sólido, de tal manera que todos o una porción de los productos amplificados se inmovilizan sobre el soporte sólido a medida que se forman. En particular, el término abarca la reacción en cadena de la polimerasa en fase sólida (PCR en fase sólida) y la amplificación isotérmica en fase sólida, que son reacciones análogas a la amplificación en fase de solución estándar, salvo que uno o ambos de los cebadores de amplificación directo e inverso se inmovilizan en el soporte sólido. La PCR de fase sólida cubre sistemas tales como emulsiones, en donde un cebador está anclado a una perla y el otro está en solución libre, y la formación de colonias en matrices de gel de fase sólida donde un cebador está anclado a la superficie, y uno está en solución libre.

En algunas realizaciones, el soporte sólido comprende una superficie con patrones. Una "superficie con patrones" se refiere a una disposición de diferentes regiones en o sobre una capa expuesta de un soporte sólido. Por ejemplo, una o más de las regiones pueden ser casillas donde haya uno o más cebadores de amplificación. Las casillas pueden estar separadas por regiones intersticiales donde no hay cebadores de amplificación. En algunas realizaciones, el patrón puede ser un formato de casillas x-y que estén en filas y columnas. En algunas realizaciones, el patrón puede ser una disposición repetitiva de casillas y/o regiones intersticiales. En algunas realizaciones, el patrón puede ser una disposición aleatoria de casillas y/o regiones intersticiales. Se describen superficies con patrones ilustrativas que pueden utilizarse en los métodos y composiciones expuestos en la presente memoria en las patentes US-8,778,848, 8,778,849 y 9,079,148, y la publicación US-2014/0243224.

En algunas realizaciones, el soporte sólido incluye una matriz de pocillos o surcos en una superficie. Este puede fabricarse, como se sabe generalmente en la técnica, utilizando una variedad de técnicas, incluyendo, aunque no de forma limitativa, fotolitografía, técnicas de estampado, técnicas de moldeo y técnicas de micrograbado. Como apreciarán los expertos en la materia, la técnica utilizada dependerá de la composición y forma del sustrato de la matriz.

Las características de una superficie con patrones pueden ser pocillos en una matriz de pocillos (p. ej., micropocillos o nanopocillos) sobre vidrio, silicio, plástico u otros soportes sólidos adecuados con gel unido covalentemente en patrones, tal como poli(N-(5-azidoacetamidilpentil)acrilamida-co-acrilamida) (PAZAM, véase, por ejemplo, las publicaciones US-2013/184796, WO 2016/066586 y WO 2015/002813). El proceso crea almohadillas de gel utilizadas para la secuenciación, que pueden ser estables durante las tandas de secuenciación con un gran número de ciclos. La unión covalente del polímero a los pocillos es útil para mantener el gel en las casillas estructuradas a lo largo de la vida útil del sustrato estructurado durante diversos usos. Sin embargo, en muchas realizaciones, el gel no tiene por qué estar unido de manera covalente a los pocillos. Por ejemplo, en algunas condiciones, acrilamida exenta de silano (SFA, véase, por ejemplo, la patente US-8,563,477), que no está unida covalentemente a ninguna parte del sustrato estructurado, se puede utilizar como material de gel.

En realizaciones particulares, puede fabricarse un sustrato estructurado diseñando un material de soporte sólido con pocillos (p. ej., micropocillos o nanopocillos), recubriendo el soporte diseñado con un material de gel (p. ej., PAZAM, SFA o variantes de los mismos modificadas químicamente, tal como la versión azidolizada de SFA (azido-SFA)) y puliendo el soporte recubierto con gel, por ejemplo, mediante pulido químico o mecánico, conservando de este modo el gel en los pocillos pero eliminando o inactivando sustancialmente todo el gel de las regiones intersticiales en la superficie del sustrato estructurado entre los pocillos. Los ácidos nucleicos cebadores pueden unirse al material de gel. A continuación, una solución de moléculas adaptadoras de fragmentos se puede poner en contacto con el sustrato pulido, de manera que las moléculas adaptadoras de fragmentos individuales sembrarán pocillos individuales mediante interacciones con cebadores unidos al material de gel; sin embargo, los ácidos nucleicos diana no ocuparán las regiones intersticiales debido a la ausencia o inactividad del material de gel. La amplificación de las moléculas adaptadoras de fragmentos quedará confinada a los pocillos, ya que la ausencia o inactividad del gel en las regiones intersticiales impide la migración hacia el exterior de la colonia de ácido nucleico en crecimiento. El proceso puede manufacturarse cómodamente, siendo escalable y utilizando métodos de micro o nanofabricación convencionales.

Aunque la exposición comprende métodos de amplificación de “fase sólida” en los que solo se inmoviliza un cebador de amplificación (el otro cebador generalmente está presente en solución libre), se prefiere que el soporte sólido se proporcione tanto con los cebadores directos como con los cebadores inversos inmovilizados. En la práctica, habrá una “pluralidad” de cebadores directos idénticos y/o una “pluralidad” de cebadores inversos idénticos inmovilizados sobre el soporte sólido, ya que el proceso de amplificación requiere un exceso de cebadores para sustentar la amplificación. Por consiguiente, en la presente memoria, las referencias a cebadores directos e inversos deben interpretarse como que abarcan una “pluralidad” de tales cebadores salvo que el contexto indique lo contrario.

Como apreciará el lector experto, cualquier reacción de amplificación dada requiere al menos un tipo de cebador directo y al menos un tipo de cebador inverso específico para el molde que se va a amplificar. Sin embargo, en determinadas realizaciones, los cebadores directo e inverso pueden incluir porciones de secuencia idéntica específicas de molde y pueden tener una secuencia y estructura nucleotídica completamente idéntica (incluida cualquier modificación no nucleotídica). En otras palabras, es posible llevar a cabo la amplificación en fase sólida utilizando solo un tipo de cebador, y dichos métodos de un solo cebador están incluidos en el ámbito de la invención. Otras realizaciones pueden utilizar cebadores directos e inversos que contienen secuencias específicas de molde idénticas, pero que difieren en algunos otros rasgos estructurales. Por ejemplo, un tipo de cebador puede contener una modificación no nucleotídica que no esté presente en el otro.

En todas las realizaciones de la exposición, los cebadores para la amplificación en fase sólida se inmovilizan preferiblemente mediante unión covalente de un solo punto al soporte sólido en, o cerca del, extremo 5' del cebador, dejando la porción específica del molde del cebador libre para emparejarse con su molde análogo y el grupo hidroxilo 3' libre para la extensión del cebador. Para ello puede utilizarse cualquier medio de unión covalente adecuado conocido en la técnica. La química de unión elegida dependerá de la naturaleza del soporte sólido y de cualquier derivatización o funcionalización que se le aplique. El propio cebador puede incluir un resto, que puede ser una modificación química no nucleotídica, para facilitar la unión. En una realización particular, el cebador puede incluir un nucleófilo que contenga azufre, tal como fosforotioato o tiosfosfato, en el extremo 5'. En el caso de hidrogeles de poli(acrilamida) de soporte sólido, este nucleófilo se unirá a un grupo de bromoacetamida presente en el hidrogel. Un medio más particular para unir cebadores y moldes a un soporte sólido es mediante la unión con 5' fosforotioato a un hidrogel que comprende acrilamida polimerizada y N-(5-bromoacetamidilpentil)acrilamida (BRAPA), como se describe totalmente en WO 05/065814.

Determinadas realizaciones de la exposición pueden utilizar soportes sólidos compuestos por un sustrato o matriz inerte (p. ej., portaobjetos de vidrio, perlas de polímero, etc.) que se han “funcionalizado”, por ejemplo, mediante la aplicación de una capa o recubrimiento de un material intermedio que comprende grupos reactivos que permiten la unión covalente a biomoléculas, tales como polinucleótidos. Como ejemplos de dichos soportes se incluyen, pero sin limitación, hidrogeles de poli(acrilamida) soportados sobre un sustrato inerte tal como vidrio. En dichas realizaciones, las biomoléculas (p. ej., polinucleótidos) pueden estar unidas directamente de manera covalente al material intermedio (p. ej., al hidrogel), pero el propio material intermedio puede estar unido de manera no covalente al sustrato o matriz (p. ej., al sustrato de vidrio). Por consiguiente, la expresión “unión covalente a un soporte sólido” debe interpretarse como que abarca este tipo de disposición.

Las muestras agrupadas pueden amplificarse en perlas, en donde cada perla contiene un cebador de amplificación directo e inverso. En una realización particular, la biblioteca de moléculas adaptadoras de fragmentos se usa para preparar matrices agrupadas de colonias de ácido nucleico, análogas a las descritas en la publicación US-2005/0100900, la patente US- 7,115,400, WO 00/18957 y WO 98/44151, mediante amplificación en fase sólida y, más particularmente, amplificación isotérmica en fase sólida. Los términos 'grupo' y 'colonia' se utilizan indistintamente en la presente memoria en referencia a un sitio separado en un soporte sólido que incluye una pluralidad de idénticas cadenas de ácido nucleico inmovilizadas y una pluralidad de idénticas cadenas de ácido nucleico complementarias inmovilizadas. La expresión “matriz agrupada” se refiere a una matriz formada a partir de tales agrupaciones o colonias. En este contexto, no debe entenderse que el término “matriz” exija una disposición ordenada de agrupaciones.

El término “fase sólida” o “superficie”, se utiliza en referencia a una matriz plana en donde los cebadores están unidos a una superficie plana, por ejemplo, portaobjetos de microscopio de vidrio, sílice o plástico o dispositivos de cubeta de lectura similares; perlas, en donde uno o dos cebadores están unidos a las perlas y las perlas se amplifican; o una matriz de perlas en una superficie después de haber amplificado las perlas.

Se pueden preparar matrices agrupadas utilizando un proceso de termociclado, como se describe en WO 98/44151, o un proceso donde la temperatura se mantiene constante, y los ciclos de extensión y desnaturalización se llevan a cabo utilizando cambios de reactivos. Dichos métodos de amplificación isotérmica se describen en la solicitud de patente WO 02/46456 y en la publicación US- 2008/0009420. Debido a las temperaturas más bajas útiles en el proceso isotérmico, esto es particularmente preferido.

Se apreciará que cualquiera de las metodologías de amplificación descritas en la presente memoria o generalmente conocidas en la técnica pueden utilizarse con cebadores universales o específicos de la diana para amplificar fragmentos de ADN inmovilizados. Como métodos de amplificación adecuados se incluyen, aunque no de forma limitativa, la reacción en cadena de la polimerasa (PCR), la amplificación por desplazamiento de cadena (SDA, por sus siglas en inglés), la amplificación mediada por transcripción (TMA, por sus siglas en inglés) y la amplificación basada en secuencias de ácidos nucleicos (NASBA, por sus siglas en inglés), como se describe en la patente US- 8,003,354. Los métodos de amplificación anteriores pueden emplearse para amplificar uno o más ácidos nucleicos de interés. Por ejemplo, la PCR, que incluye PCR múltiple, SDA, TMA, NASBA y similares, puede utilizarse para amplificar fragmentos de ADN inmovilizados. En algunas realizaciones, en la reacción de amplificación se incluyen cebadores dirigidos específicamente al polinucleótido de interés.

Otros métodos adecuados para la amplificación de polinucleótidos pueden incluir extensión y ligamiento de oligonucleótidos, amplificación por círculo rodante (RCA) (Lizardi y col., Nat. Genet. 19:225-232 (1998)) y ensayo de ligamiento de oligonucleótidos (OLA) (Véase de forma general las patentes US- 7.582.420, US-5.185.243, US-5.679.524 y US-5.573.907; EP 0 320 308 B1; EP 0 336 731 B1; EP 0 439 182 B1; WO 90/01069; WO 89/12696; y WO 89/09835). Se apreciará que estas metodologías de amplificación se pueden diseñar para amplificar fragmentos de ADN inmovilizados. Por ejemplo, en algunas realizaciones, el método de amplificación puede incluir amplificación por sonda de ligamiento o reacciones de ensayo de ligamiento de oligonucleótidos (OLA) que contienen cebadores dirigidos específicamente al ácido nucleico de interés. En algunas realizaciones, el método de amplificación puede incluir una reacción de extensión de cebador por ligamiento que contiene cebadores dirigidos específicamente al ácido nucleico de interés. Como ejemplo no limitativo de cebadores de extensión y ligamiento de cebadores que se pueden diseñar específicamente para amplificar un ácido nucleico de interés, la amplificación puede incluir cebadores utilizados para el ensayo GoldenGate (Illumina, Inc., San Diego, CA) como se ilustra en las patentes US- 7,582,420 y 7,611,869.

Los métodos de amplificación isotérmica ilustrativos que se pueden utilizar en un método de la presente exposición incluyen, aunque no de forma limitativa, la amplificación por desplazamiento múltiple (MDA) como se ilustra, por ejemplo, en Dean y col., Proc. Natl. Acad. Sci. EE. UU. 99:5261-66 (2002) o amplificación isotérmica de ácido nucleico por desplazamiento de cadena ilustrada, por ejemplo, en la patente US- US-6.214.587. Otros métodos no basados en PCR que se pueden utilizar en la presente exposición incluyen, por ejemplo, la amplificación por desplazamiento de cadena (SDA) que se describe, por ejemplo, en Walker y col., Molecular Methods for Virus Detection, Academic Press, Inc., 1995; las patentes US- 5,455,166 y 5,130,238, y en Walker y col., Nucl. Acids Res. 20:1691-96 (1992) o amplificación por desplazamiento de cadena hiperramificada que se describe, por ejemplo, en Lage y col., Genome Res. 13:294-307 (2003). Pueden utilizarse métodos de amplificación isotérmica con la polimerasa Phi 29 de desplazamiento de cadena o el fragmento grande de polimerasa de ADN Bst, *exo-5'->3'* para la amplificación aleatoria de cebador de ADN genómico. El uso de estas polimerasas aprovecha su alta procesividad y actividad de desplazamiento de cadenas. La alta procesividad permite que las polimerasas produzcan fragmentos que tienen una longitud de 10-20 kb. Como se ha expuesto anteriormente, pueden producirse fragmentos más pequeños en condiciones isotérmicas utilizando polimerasas que tienen baja procesividad y actividad de desplazamiento de cadena tal como la polimerasa Klenow. Se expone una descripción adicional en detalle de las reacciones, condiciones y componentes de amplificación en la exposición de la patente US-7.670.810.

Otro método de amplificación de polinucleótidos que es útil en la presente exposición es la PCR etiquetada que utiliza una población de cebadores de dos dominios que tienen una región 5' constante seguida por una región 3' aleatoria como se describe, por ejemplo, en Grothues y col. Nucleic Acids Res. 21(5):1321-2 (1993). Las primeras rondas de amplificación se llevan a cabo para permitir una multitud de iniciaciones en ADN desnaturalizado por calor basadas en hibridación individual de la región 3' sintetizada al azar. Debido a la naturaleza de la región 3', se contempla que los sitios de iniciación sean aleatorios en todo el genoma. Posteriormente, los cebadores no unidos pueden retirarse y puede producirse una replicación adicional utilizando cebadores complementarios a la región 5' constante.

En algunas realizaciones, la amplificación isotérmica puede realizarse utilizando amplificación por exclusión cinética (KEA, kinetic exclusion amplification), también denominada amplificación por exclusión (ExAmp). Puede obtenerse una biblioteca de ácidos nucleicos de la presente exposición utilizando un método que incluya una etapa de reacción de un reactivo de amplificación para producir una pluralidad de sitios de amplificación que incluyan, cada uno de ellos, una población sustancialmente clonal de amplicones de un ácido nucleico diana individual que se haya sembrado en

el sitio. En algunas realizaciones, la reacción de amplificación continúa hasta que se genera una cantidad suficiente de amplicones para llenar la capacidad del sitio de amplificación respectivo. De esta manera, llenar un sitio ya sembrado hasta su capacidad inhibe la llegada de los ácidos nucleicos diana y su amplificación en el sitio, produciendo así una población clonal de amplicones en el sitio. En algunas realizaciones, la clonalidad aparente puede conseguirse incluso si un sitio de amplificación no se llena hasta su capacidad antes de que un segundo ácido nucleico diana llegue al sitio. En algunas condiciones, la amplificación de un primer ácido nucleico diana puede continuar hasta un punto en que se produce una cantidad suficiente de copias para superar o saturar eficazmente la producción de copias de un segundo ácido nucleico diana que se transporta al sitio. Por ejemplo, en una realización en que se utiliza un proceso de amplificación en puente en una casilla circular con un diámetro menor de 500 nm, se ha determinado que después de 14 ciclos de amplificación exponencial para un primer ácido nucleico diana, la contaminación de un segundo ácido nucleico diana en el mismo sitio producirá una cantidad insuficiente de amplicones contaminantes que afecte negativamente al análisis de secuenciación por síntesis en una plataforma de secuenciación Illumina.

En algunas realizaciones, los sitios de amplificación en una matriz pueden ser, pero no necesariamente, completamente clonales. En cambio, para algunas aplicaciones, un sitio de amplificación individual puede estar predominantemente poblado con amplicones de una primera molécula adaptadora de fragmento y también puede tener un nivel bajo de amplicones contaminantes de un segundo ácido nucleico diana. Una matriz puede tener uno o más sitios de amplificación que tengan un nivel bajo de amplicones contaminantes siempre que el nivel de contaminación no tenga un impacto inaceptable en un uso posterior de la matriz. Por ejemplo, cuando la matriz va a utilizarse en una aplicación de detección, un nivel aceptable de contaminación sería un nivel que no altere la relación señal a ruido o la resolución de la técnica de detección de una forma inaceptable. Por consiguiente, en general, la clonalidad aparente será relevante para un uso o aplicación particular de una matriz creada mediante los métodos expuestos en la presente memoria. Niveles de contaminación ilustrativos que pueden ser aceptables en un sitio de amplificación individual para aplicaciones particulares incluyen, aunque no de forma limitativa, como máximo 0,1 %, 0,5 %, 1 %, 5 %, 10 % o 25 % de amplicones contaminantes. Una matriz puede incluir uno o más sitios de amplificación que tengan estos niveles ilustrativos de amplicones contaminantes. Por ejemplo, hasta un 5 %, 10 %, 25 %, 50 %, 75 % o incluso 100 % de los sitios de amplificación en una matriz pueden tener algunos amplicones contaminantes. Se entenderá que en una matriz u otro conjunto de sitios, al menos 50 %, 75 %, 80 %, 85 %, 90 %, 95 % o 99 % o más de los sitios pueden ser clonales o aparentemente clonales.

En algunas realizaciones, puede producirse exclusión cinética cuando un proceso se produce a una velocidad lo suficientemente rápida como para excluir eficazmente que se produzca otro acontecimiento o proceso. Tomemos por ejemplo la creación de una matriz de ácido nucleico donde los sitios de la matriz se siembran aleatoriamente con moléculas adaptadoras de fragmento de una solución y se generan copias de ellas en un proceso de amplificación para llenar cada uno de los sitios sembrados hasta su capacidad. Según los métodos de exclusión cinética de la presente exposición, los procesos de siembra y amplificación pueden llevarse a cabo simultáneamente en condiciones en que la velocidad de amplificación supere la velocidad de siembra. Como tal, la velocidad relativamente rápida a la que se realizan copias en un sitio que se ha sembrado con un primer ácido nucleico diana excluirá eficazmente la siembra de un segundo ácido nucleico en el sitio para la amplificación. Pueden llevarse a cabo métodos de amplificación de exclusión cinética como se describe en detalle en la exposición de la publicación de solicitud de patente US-2013/0338042.

La exclusión cinética puede aprovechar una velocidad relativamente lenta para iniciar la amplificación (p. ej., una velocidad lenta para hacer una primera copia de una molécula adaptadora de fragmento) frente a una velocidad relativamente rápida para hacer copias posteriores de dicha molécula (o de la primera copia de dicha molécula). En el ejemplo del párrafo anterior, la exclusión cinética se produce debido a la velocidad relativamente lenta de siembra de la molécula adaptadora de fragmento (por ejemplo, difusión o transporte relativamente lentos) frente a la velocidad relativamente rápida a la que se produce la amplificación para llenar el sitio con copias de la semilla adaptadora de fragmento. En otra realización ilustrativa, la exclusión cinética puede producirse debido a un retraso en la formación de una primera copia de una molécula adaptadora de fragmento que ha sembrado un sitio (p. ej., activación retardada o lenta) frente a la velocidad relativamente rápida a la que se realizan copias posteriores para llenar el sitio. En este ejemplo, puede haberse sembrado un sitio individual con varias moléculas adaptadoras de fragmentos distintas (p. ej., antes de la amplificación, puede haber varias moléculas adaptadoras de fragmentos en cada sitio). Sin embargo, la primera formación de copias para cualquier molécula adaptadora de fragmento dada puede activarse de forma aleatoria, de modo que la velocidad promedio de la primera formación de copias sea relativamente lenta en comparación con la velocidad a la que se generan copias posteriores. En este caso, aunque un sitio individual pueda haberse sembrado con varias moléculas adaptadoras de fragmentos distintas, la exclusión cinética permitirá que solo se amplifique una de estas moléculas. Más específicamente, una vez activada una primera molécula adaptadora de fragmento para la amplificación, el sitio se llenará rápidamente hasta su capacidad con sus copias, impidiendo de este modo que se realicen copias de una segunda molécula adaptadora de fragmento en el sitio.

Un reactivo de amplificación puede incluir componentes adicionales que faciliten la formación de amplicones y, en algunos casos, aumenten la velocidad de formación de los amplicones. Un ejemplo es una recombinasa. La recombinasa puede facilitar la formación de amplicones al permitir la invasión/extensión repetida. Más específicamente, la recombinasa puede facilitar la invasión de una molécula adaptadora de fragmentos por la polimerasa y la extensión de un cebador por la polimerasa utilizando como molde la molécula adaptadora de

fragmentos para la formación de amplicones. Este proceso puede repetirse como una reacción en cadena donde los amplicones producidos a partir de cada ronda de invasión/extensión sirven como moldes en una ronda posterior. El proceso puede producirse más rápidamente que la PCR estándar, ya que no se requiere un ciclo de desnaturalización (por ejemplo, mediante calentamiento o desnaturalización química). Como tal, la amplificación facilitada por recombinasa puede llevarse a cabo isotérmicamente. De forma general, es deseable incluir ATP u otros nucleótidos (o, en algunos casos, análogos no hidrolizables de los mismos) en un reactivo de amplificación facilitada por recombinasa para facilitar la amplificación. Una mezcla de recombinasa y proteína de unión monocatenaria (SSB, single stranded binding) es especialmente útil ya que la SSB puede facilitar aún más la amplificación. Formulaciones ilustrativas para la amplificación facilitada por recombinasa incluyen las vendidas comercialmente como kits TwistAmp por TwistDx (Cambridge, Reino Unido). Componentes útiles del reactivo de amplificación facilitado por recombinasa y las condiciones de reacción se exponen en US-5,223,414 y US-7,399,590.

Otro ejemplo de un componente que puede incluirse en un reactivo de amplificación para facilitar la formación de amplicones y en algunos casos para aumentar la velocidad de formación de amplicones es una helicasa. La helicasa puede facilitar la formación de amplicones al permitir una reacción en cadena de la formación de amplicones. El proceso puede producirse más rápidamente que la PCR estándar, ya que no se requiere un ciclo de desnaturalización (por ejemplo, mediante calentamiento o desnaturalización química). Como tal, la amplificación facilitada por helicasa puede llevarse a cabo isotérmicamente. Una mezcla de helicasa y proteína de unión monocatenaria (SSB) es especialmente útil ya que la SSB puede facilitar aún más la amplificación. Formulaciones ilustrativas para la amplificación facilitada por helicasa incluyen las comercializadas como kits de IsoAmp de Biohelix (Beverly, MA). Además, se describen ejemplos de formulaciones útiles que incluyen una proteína helicasa en US-7,399,590 y US-7,829,284.

Otro ejemplo de componente que puede incluirse en un reactivo de amplificación para facilitar la formación de amplicones y, en algunos casos, aumentar la velocidad de formación de amplicones, es una proteína de unión al origen.

Después de la unión de las moléculas adaptadoras de fragmentos a una superficie, se determina la secuencia de las moléculas adaptadoras de fragmentos inmovilizadas y amplificadas. La secuenciación se puede llevar a cabo utilizando cualquier técnica de secuenciación adecuada, y los métodos para determinar la secuencia de las moléculas adaptadoras de fragmentos inmovilizadas y amplificadas, que incluyen la resíntesis de cadena, son conocidos en la técnica y se describen, por ejemplo, en Bignell y col. (US-8,053,192), Gunderson y col. (WO2016/130704), Shen y col. (US-8,895,249) y Pipenburg y col. (US 9.309.502).

Los métodos descritos en la presente memoria pueden utilizarse junto con una variedad de técnicas de secuenciación de ácidos nucleicos. Las técnicas particularmente aplicables son aquellas en donde los ácidos nucleicos se fijan en ubicaciones fijas en una matriz de modo que sus posiciones relativas no cambien, y en donde se obtienen imágenes de la matriz de manera repetida. Son particularmente aplicables las realizaciones en las que se obtienen imágenes en diferentes canales de color, por ejemplo, coincidiendo con diferentes marcadores utilizados para distinguir un tipo de base de nucleótido de otro. En algunas realizaciones, el proceso para determinar la secuencia de nucleótidos de una molécula adaptadora de fragmentos puede ser un proceso automatizado. Las realizaciones preferidas incluyen técnicas de secuenciación por síntesis ("SBS").

Las técnicas de SBS generalmente implican la extensión enzimática de una cadena de ácido nucleico en formación mediante la adición iterativa de nucleótidos frente a una cadena molde. En los métodos tradicionales de SBS, en cada suministro, puede proporcionarse un monómero de un solo nucleótido a un nucleótido diana en presencia de una polimerasa. Sin embargo, en los métodos descritos en la presente memoria, en un suministro puede proporcionarse más de un tipo de monómero nucleotídico a un ácido nucleico diana en presencia de una polimerasa.

En una realización, un monómero nucleotídico incluye ácidos nucleicos bloqueados (LNA) o ácidos nucleicos con puente (BNA). Cuando las moléculas adaptadoras de fragmentos se producen utilizando una o más secuencias de nucleótidos empobrecidas en citosina, tal como sucede cuando las secuencias de nucleótidos empobrecidas en citosina están presentes en una cadena transferida de un complejo transposómico, la temperatura de fusión de un monómero de nucleótido que se hibrida a una región empobrecida en citosina está alterada. El uso de los LNA o BNA en un monómero nucleotídico aumenta la fuerza de hibridación entre un monómero nucleotídico y una secuencia cebadora de secuenciación presente en una molécula adaptadora de fragmentos inmovilizada.

La SBS puede utilizar monómeros de nucleótidos que tienen un resto terminador o aquellos que carecen de restos terminadores. Los métodos que utilizan monómeros de nucleótidos que carecen de terminadores incluyen, por ejemplo, pirosecuenciación y secuenciación utilizando nucleótidos marcados con γ -fosfato, como se expone con más detalle a continuación. En los métodos en los cuales se utilizan monómeros de nucleótidos que carecen de terminadores, el número de nucleótidos añadidos en cada ciclo es generalmente variable y depende de la secuencia molde y del modo de suministro de los nucleótidos. Para las técnicas de SBS que utilizan monómeros de nucleótidos que tienen un resto terminador, el terminador puede ser efectivamente irreversible en las condiciones de secuenciación utilizadas como es el caso de la secuenciación tradicional de Sanger que utiliza didesoxinucleótidos, o el terminador puede ser reversible como es el caso de los métodos de secuenciación desarrollados por Solexa (ahora Illumina, Inc.).

Las técnicas de SBS pueden utilizar monómeros de nucleótidos que tengan un resto marcador o que carezcan de este. Por consiguiente, pueden detectarse acontecimientos de incorporación basándose en una característica del marcador, tal como fluorescencia del marcador; una característica del monómero nucleotídico, tal como la carga o el peso molecular; un subproducto de incorporación del nucleótido, tal como la liberación de pirofosfato; o similares. En realizaciones en que hay dos o más nucleótidos diferentes en un reactivo de secuenciación, los diferentes nucleótidos pueden diferenciarse entre sí, o como alternativa, los dos o más marcadores diferentes pueden no diferenciarse con las técnicas de detección utilizadas. Por ejemplo, los diferentes nucleótidos presentes en un reactivo de secuenciación, pueden tener diferentes marcadores y se pueden distinguir usando ópticas apropiadas como se ilustra mediante los métodos de secuenciación desarrollados por Solexa (ahora Illumina, Inc.).

Las realizaciones preferidas incluyen técnicas de pirosecuenciación. La pirosecuenciación detecta la liberación de pirofosfato inorgánico (PPi) a medida que se incorporan nucleótidos particulares en la cadena en formación (Ronaghi, M., Karamohamed, S., Pettersson, B., Uhlen, M. y Nyren, P. (1996) "Real-time DNA sequencing using detection of pyrophosphate release". *Analytical Biochemistry* 242(1), 84-9; Ronaghi, M. (2001) "Pyrosequencing sheds light on DNA sequencing". *Genome Res.* 11(1), 3-11; Ronaghi, M., Uhlen, M. y Nyren, P. (1998) "A sequencing method based on real-time pyrophosphate". *Science* 281(5375), 363; las patentes US- 6,210,891; 6,258,568 y 6,274,320). En la pirosecuenciación, el PPi (pirofosfato inorgánico) liberado puede detectarse convirtiéndolo inmediatamente en adenosín trifosfato (ATP) por la ATP sulfurilasa, y el nivel de ATP generado se detecta mediante fotones producidos por luciferasa. Los ácidos nucleicos que se van a secuenciar se pueden fijar a las casillas de una matriz y se pueden obtener imágenes de la matriz para capturar las señales quimioluminiscentes que se producen debido a la incorporación de nucleótidos en las casillas de la matriz. Puede obtenerse una imagen después de que la matriz se trate con un tipo de nucleótido particular (por ejemplo, A, T, C o G). Las imágenes obtenidas después de la adición de cada tipo de nucleótido diferirán con respecto a las casillas de la matriz que se detecten. Estas diferencias en la imagen reflejan el contenido de secuencia diferente de las casillas en la matriz. Sin embargo, las ubicaciones relativas de cada casilla permanecerán sin cambios en las imágenes. Las imágenes pueden almacenarse, procesarse y analizarse utilizando los métodos expuestos en la presente memoria. Por ejemplo, las imágenes obtenidas después del tratamiento de la matriz con cada tipo de nucleótido diferente se pueden manejar de la misma manera que se ilustra en la presente memoria para imágenes obtenidas de diferentes canales de detección para métodos de secuenciación basados en terminadores reversibles.

En otro tipo ilustrativo de SBS, la secuenciación del ciclo se logra mediante la adición escalonada de nucleótidos terminadores reversibles que contienen, por ejemplo, un marcador de colorante escindible o fotoblanqueable como se describe, por ejemplo, en la patente WO 04/018497 y en la patente US- 7,057,026. Este enfoque está siendo comercializado por Solexa (ahora Illumina Inc.), y también se describe en las patentes WO 91/06678 y WO 07/123.744. La disponibilidad de terminadores marcados con fluorescencia, en los cuales se puede invertir tanto la terminación como escindir el marcador fluorescente, facilita la secuenciación eficaz de la terminación cíclica reversible (CRT). Las polimerasas también pueden diseñarse conjuntamente para incorporar y extenderse eficazmente a partir de estos nucleótidos modificados.

Preferiblemente, en realizaciones de secuenciación basadas en terminadores reversibles, los marcadores no inhiben sustancialmente la extensión en las condiciones de reacción de SBS. Sin embargo, los marcadores de detección pueden eliminarse, por ejemplo, mediante escisión o degradación. Las imágenes pueden capturarse después de la incorporación de marcadores en las casillas de ácidos nucleicos de la matriz. En realizaciones particulares, cada ciclo implica el suministro simultáneo de cuatro tipos de nucleótidos diferentes a la matriz y cada tipo de nucleótido tiene un marcador espectralmente distinto. A continuación, se pueden obtener cuatro imágenes, cada una de las cuales utiliza un canal de detección que es selectivo para uno de los cuatro marcadores diferentes. Como alternativa, se pueden añadir secuencialmente diferentes tipos de nucleótidos y se puede obtener una imagen de la matriz entre cada etapa de adición. En dichas realizaciones, cada imagen mostrará casillas de ácido nucleico que han incorporado nucleótidos de un tipo particular. En las diferentes imágenes habrá o no diferentes casillas debido al diferente contenido de secuencia de cada casilla. Sin embargo, la posición relativa de las casillas permanecerá sin cambios en las imágenes. Las imágenes obtenidas a partir de dichos métodos de SBS con terminador reversible pueden almacenarse, procesarse y analizarse como se expone en la presente memoria. Después de la etapa de captura de imágenes, los marcadores se pueden eliminar y los restos terminadores reversibles se pueden eliminar para los ciclos posteriores de adición y detección de nucleótidos. La eliminación de los marcadores después de que se hayan detectado en un ciclo particular y antes de un ciclo posterior, puede proporcionar la ventaja de reducir la señal de fondo y la diafonía entre ciclos. A continuación se exponen ejemplos de marcadores útiles y de métodos de eliminación.

En realizaciones particulares, algunos o todos los monómeros de nucleótidos pueden incluir terminadores reversibles. En dichas realizaciones, los terminadores reversibles/fluoróforos escindibles pueden incluir fluoróforos ligados al resto de ribosa mediante un enlace éster 3' (Metzker, *Genome Res.* 15:1767-1776 (2005)). Otros enfoques han separado la química del terminador de la escisión del marcador de fluorescencia (Ruparel y col., *Proc Natl Acad Sci USA* 102: 5932-7 (2005)). Ruparel y col., describieron el desarrollo de terminadores reversibles que utilizaban un pequeño grupo alilo 3' para bloquear la extensión, pero podían desbloquearse fácilmente mediante un tratamiento corto con un catalizador de paladio. El fluoróforo se unió a la base mediante un enlazador fotoescindible que podía escindirse fácilmente mediante una exposición de 30 segundos a luz UV de longitud de onda larga. Por lo tanto, como un

enlazador escindible, puede utilizarse la reducción de disulfuro o la fotoescisión. Otro enfoque de la terminación reversible es el uso de la terminación natural que se produce después de la colocación de un colorante voluminoso en un dNTP. La presencia de un colorante voluminoso cargado en el dNTP puede actuar como un terminador eficaz a través de un impedimento estérico y/o electrostático. La presencia de un acontecimiento de incorporación impide otras incorporaciones a menos que se elimine el colorante. La escisión del colorante elimina el fluoróforo e invierte eficazmente la terminación. También se describen ejemplos de nucleótidos modificados en la patentes US- 7,427,673 y 7,057,026.

Sistemas y métodos de SBS ilustrativos adicionales que se pueden utilizar con los métodos y sistemas descritos en la presente memoria se describen en las publicaciones US-2007/0166705, 2006/0188901, 2006/0240439, 2006/0281109, 2012/0270305, y 2013/0260372, la patente US-7,057,026, la publicación PCT WO 05/065814, la solicitud de patente US-2005/0100900 y las publicaciones PCT WO 06/064199 y WO 07/010,251.

Algunas realizaciones pueden utilizar la detección de cuatro nucleótidos diferentes usando menos de cuatro marcadores diferentes. Por ejemplo, puede realizarse SBS utilizando los métodos y los sistemas descritos en los materiales incorporados de la publicación US- 2013/0079232. Como primer ejemplo, puede detectarse un par de tipos de nucleótidos en la misma longitud de onda, pero distinguirse en función de la diferencia de intensidad de un miembro del par en comparación con el otro, o en función de un cambio en un miembro del par (por ejemplo, mediante modificación química, modificación fotoquímica o modificación física) que hace que aparezca o desaparezca una señal aparente en comparación con la señal detectada para el otro miembro del par. Como un segundo ejemplo, se pueden detectar tres de cuatro tipos de nucleótidos diferentes en condiciones particulares, mientras que un cuarto tipo de nucleótidos carece de un marcador que sea detectable en esas condiciones, o se detecta mínimamente en esas condiciones (p. ej., detección mínima debido a la fluorescencia de fondo, etc.). La incorporación de los primeros tres tipos de nucleótidos en un ácido nucleico se puede determinar basándose en la presencia de sus respectivas señales y la incorporación del cuarto tipo de nucleótidos en el ácido nucleico se puede determinar basándose en la ausencia o detección mínima de cualquier señal. Como un tercer ejemplo, un tipo de nucleótidos puede incluir un(os) marcador(es) que se detecta(n) en dos canales diferentes, mientras que otros tipos de nucleótidos se detectan en no más de uno de los canales. Las tres configuraciones ilustrativas mencionadas anteriormente no se consideran mutuamente excluyentes y pueden usarse en diversas combinaciones. Una realización ilustrativa que combina los tres ejemplos es un método de SBS basado en fluorescencia que usa un primer tipo de nucleótido que se detecta en un primer canal (por ejemplo, dATP que tiene un marcador que se detecta en el primer canal cuando se excita con una primera longitud de onda de excitación), un segundo tipo de nucleótido que se detecta en un segundo canal (por ejemplo, dCTP que tiene un marcador que se detecta en el segundo canal cuando se excita con una segunda longitud de onda de excitación), un tercer tipo de nucleótido que se detecta tanto en el primer como en el segundo canal (por ejemplo, dTTP que tiene al menos un marcador que se detecta en ambos canales cuando se excita con la primera y/o segunda longitud de onda de excitación) y un cuarto tipo de nucleótido que carece de un marcador que no se detecta, o se detecta mínimamente, en ninguno de los canales (por ejemplo, dGTP sin marcador).

Además, como se describe en los materiales incorporados de la publicación US- 2013/0079232, los datos de secuenciación pueden obtenerse usando un solo canal. En los llamados enfoques de secuenciación de un colorante, el primer tipo de nucleótido se marca pero el marcador se elimina después de que se genere la primera imagen, y el segundo tipo de nucleótido se marca solo después de que se genere una primera imagen. El tercer tipo de nucleótido conserva su marcador tanto en la primera como en la segunda imagen, y el cuarto tipo de nucleótido permanece sin marcar en ambas imágenes.

Algunas realizaciones pueden utilizar secuenciación mediante técnicas de ligamiento. Tales técnicas utilizan ADN ligasa para incorporar oligonucleótidos e identificar la incorporación de tales oligonucleótidos. Los oligonucleótidos tienen normalmente diferentes marcadores que se correlacionan con la identidad de un nucleótido particular en una secuencia con la que hibridan los oligonucleótidos. Al igual que con otros métodos de SBS, se pueden obtener imágenes después del tratamiento de una matriz de casillas de ácidos nucleicos con los reactivos de secuenciación marcados. Cada imagen mostrará casillas de ácido nucleico que hayan incorporado marcadores de un tipo particular. En las diferentes imágenes habrá o no diferentes casillas debido al diferente contenido de secuencia de cada casilla, pero la posición relativa de las casillas permanecerá sin cambios en las imágenes. Las imágenes obtenidas a partir de métodos de secuenciación basados en ligamiento se pueden almacenar, procesar y analizar como se expone en la presente memoria. Sistemas y métodos de SBS ilustrativos que pueden utilizarse con los métodos y sistemas descritos en la presente memoria se describen en las patentes US- 6,969,488, 6,172,218 y 6,306,597.

Algunas realizaciones pueden utilizar la secuenciación de nanoporos (Deamer, D. W. y Akeson, M. "Nanopores and nucleic acids: prospects for ultrarapid sequencing". Trends Biotechnol. 18, 147-151 (2000); Deamer, D. y D. Branton, "Characterization of nucleic acids by nanopore analysis", Acc. Chem. Res. 35:817-825 (2002); Li, J., M. Gershow, D. Stein, E. Brandin y J. A. Golovchenko, "DNA molecules and configurations in a solid-state nanopore microscope" Nat. Mater. 2:611-615 (2003)). En dichas realizaciones, la molécula adaptadora de fragmento pasa a través de un nanoporo. El nanoporo puede ser un poro sintético o una proteína de membrana biológica, tal como la α -hemolisina. Cuando la molécula adaptadora de fragmento pasa a través del nanoporo, cada par de bases puede identificarse midiendo las fluctuaciones en la conductancia eléctrica del poro. (Patente US- 7,001,792; Soni, G. V. & Meller, "A.

Progress toward ultrafast DNA sequencing using solid-state nanopores". Clin. Chem. 53, 1996-2001 (2007); Healy, K. "Nanopore-based single-molecule DNA analysis". Nanomed. 2, 459-481 (2007); Cockroft, S. L., Chu, J., Amorin, M. & Ghadiri, M. R. "A single-molecule nanopore device detects DNA polymerase activity with single-nucleotide resolution". J. Am. Chem. Soc. 130, 818-820 [2008]). Los datos obtenidos de la secuenciación de nanoporos se pueden almacenar, procesar y analizar como se expone en la presente memoria. En particular, los datos se pueden tratar como una imagen según el tratamiento ilustrativo de imágenes ópticas y otras imágenes que se expone en la presente memoria.

Algunas realizaciones pueden utilizar métodos que suponen la monitorización en tiempo real de la actividad de la ADN polimerasa. Las incorporaciones de nucleótidos pueden detectarse mediante interacciones de transferencia de energía por resonancia de fluorescencia (FRET) entre una polimerasa portadora de fluoróforo y nucleótidos marcados con γ -fosfato como se describe, por ejemplo, en las patentes US- 7,329,492 y 7,211,414, o las incorporaciones de nucleótidos se pueden detectar con guías de onda en modo cero como se describe, por ejemplo, en la patente US- 7,315,019 y utilizando análogos de nucleótidos fluorescentes y polimerasas diseñadas como se describe, por ejemplo, en la patente US- 7,405,281 y la publicación US- 2008/0108082. La iluminación se puede restringir a un volumen a escala de zeptolitros alrededor de una polimerasa anclada a la superficie, de modo que se pueda observar la incorporación de nucleótidos marcados con fluorescencia con un fondo bajo (Levene, M. J. y col. "Zero-mode waveguides for single-molecule analysis at high concentrations". Science 299, 682-686 (2003); Lundquist, P. M. y col. "Parallel confocal detection of single molecules in real time". Opt. Lett. 33, 1026-1028 (2008); Korlach, J. y col. "Selective aluminum passivation for targeted immobilization of single DNA polymerase molecules in zero-mode waveguide nano structures". Proc. Natl. Acad. Sci. EE. UU. 105, 1176-1181 [2008]). Las imágenes obtenidas de dichos métodos se pueden almacenar, procesar y analizar tal como se expone en la presente memoria.

Algunas realizaciones de SBS (secuenciación por síntesis) incluyen la detección de un protón liberado tras la incorporación de un nucleótido en un producto de extensión. Por ejemplo, en la secuenciación basada en la detección de protones liberados se puede utilizar un detector eléctrico y técnicas asociadas que son comercializadas por Ion Torrent (Guilford, CT, una filial de Life Technologies) o métodos y sistemas de secuenciación descritos en las publicaciones US- 2009/0026082; 2009/0127589; 2010/0137143; y 2010/0282617. Los métodos expuestos en la presente memoria para amplificar ácidos nucleicos diana utilizando exclusión cinética pueden aplicarse fácilmente a sustratos utilizados para detectar protones. Más específicamente, los métodos expuestos en la presente memoria pueden utilizarse para producir poblaciones clonales de amplicones que se utilizan para detectar protones.

Los métodos de SBS anteriores se pueden llevar a cabo ventajosamente en formatos multiplexados de modo que se manipulen simultáneamente múltiples moléculas adaptadoras de fragmentos diferentes. En realizaciones particulares, se pueden tratar diferentes moléculas adaptadoras de fragmentos en un recipiente de reacción habitual o en una superficie de un sustrato particular. Esto permite el suministro conveniente de reactivos de secuenciación, la eliminación de reactivos que no han reaccionado y la detección de acontecimientos de incorporación de manera multiplexada. En realizaciones en que se utilizan ácidos nucleicos diana unidos a la superficie, las moléculas adaptadoras de fragmentos pueden estar en un formato de matriz. En un formato de matriz, las moléculas adaptadoras de fragmentos pueden unirse normalmente a una superficie de una manera espacialmente diferenciable. Las moléculas adaptadoras de fragmentos pueden unirse mediante unión covalente directa, unión a una perla o a otra partícula o unión a una polimerasa o a otra molécula que esté unida a la superficie. La matriz puede incluir una sola copia de una molécula adaptadora de fragmento en cada sitio (también denominado casilla) o puede haber múltiples copias que tengan la misma secuencia en cada sitio o casilla. Pueden producirse múltiples copias mediante métodos de amplificación tales como amplificación en puente o PCR en emulsión, como se describe con más detalle a continuación.

En los métodos expuestos en la presente memoria pueden utilizarse matrices que tengan casillas en cualquiera de una variedad de densidades, incluyendo, por ejemplo, al menos aproximadamente 10 casillas/cm², 100 casillas/cm², 500 casillas/cm², 1000 casillas/cm², 5000 casillas/cm², 10.000 casillas/cm², 50.000 casillas/cm², 100.000 casillas/cm², 1.000.000 casillas/cm², 5.000.000 casillas/cm² o un valor superior.

Una ventaja de los métodos expuestos en la presente memoria es que proporcionan una detección rápida y eficaz de una pluralidad de cm² en paralelo. Por consiguiente, la presente exposición proporciona sistemas integrados capaces de preparar y detectar ácidos nucleicos usando técnicas conocidas en la materia tales como las ilustradas anteriormente. Por lo tanto, un sistema integrado de la presente exposición puede incluir componentes fluidicos capaces de suministrar reactivos de amplificación y/o reactivos de secuenciación a uno o más fragmentos de ADN inmovilizados, incluyendo el sistema componentes tales como bombas, válvulas, depósitos, líneas fluidicas y similares. En un sistema integrado para la detección de ácidos nucleicos diana puede configurarse y/o utilizarse una cubeta de lectura. Se describen cubetas de lectura ilustrativas, por ejemplo, en la publicación US- 2010/0111768 y US con n.º de serie 13/273,666. Como se ilustra en las cubetas de lectura, puede utilizarse uno o más de los componentes fluidicos de un sistema integrado para un método de amplificación y para un método de detección. Tomando como ejemplo una realización de secuenciación de ácidos nucleicos, puede utilizarse uno o más de los componentes fluidicos de un sistema integrado para un método de amplificación expuesto en la presente memoria y para el suministro de reactivos de secuenciación en un método de secuenciación tales como en los ilustrados anteriormente. Como alternativa, un sistema integrado puede incluir sistemas fluidicos distintos para llevar a cabo métodos de amplificación y para llevar a cabo métodos de detección. Los ejemplos de sistemas de secuenciación integrados que

son capaces de crear ácidos nucleicos amplificados y también de determinar la secuencia de los ácidos nucleicos incluyen, sin limitarse a, la plataforma MiSeq™ (Illumina, Inc., San Diego, CA) y los dispositivos descritos en US con n.º de serie 13/273,666.

5 Durante la práctica de los métodos descritos en la presente memoria pueden resultar diversas composiciones. Por ejemplo, puede resultar una molécula adaptadora de fragmentos de doble índice, que incluya una molécula adaptadora de fragmentos de doble índice que tenga la estructura mostrada en la Figura 2, bloque vii o en la Figura 4, y composiciones que incluyan una molécula adaptadora de fragmentos de doble índice. Puede resultar una biblioteca de secuenciación de moléculas adaptadoras de fragmentos de doble índice, incluyendo moléculas adaptadoras de fragmentos de doble índice que tengan una estructura mostrada en la Figura 2, bloque vii o en la Figura 4, y composiciones que incluyan una biblioteca de secuenciación. Dicha biblioteca de secuenciación puede unirse a una matriz.

15 La presente invención se define mediante los siguientes ejemplos. Debe entenderse que los ejemplos, materiales, cantidades y procedimientos particulares deben interpretarse ampliamente según el alcance de la invención tal como se expone en la presente memoria.

Ejemplos

20 Reactivos utilizados en los ejemplos

- Solución salina tamponada con fosfato (PBS, Thermo Fisher, Cat. 10010023)
- Tripsina al 0,25 % (Thermo Fisher, Cat. 15050057)
- 25 · Tris (Fisher, Cat. T1503)
- HCl (Fisher, Cat. A144)
- 30 · NaCl (Fisher, Cat. M-11624)
- MgCl₂ (Sigma, Cat. M8226)
- Igepal® CA-630 (Sigma, I8896)
- 35 · Inhibidores de proteasa (Roche, Cat. 11873580001)
- PCR-Clean ddH₂O
- 40 · Ácido 3,5-diyodosalicílico de litio (Sigma, Cat. D3635) - solo método LAND
- Formaldehído (Sigma, Cat. F8775) - solo método xSDS
- Glicina (Sigma, Cat. G8898) - solo método xSDS
- 45 · NEBuffer 2.1 (NEB, Cat. B7202) - solo método xSDS
- SDS (Sigma, Cat. L3771) - solo método xSDS
- 50 · Triton™ X-100 (Sigma, Cat. 9002-93-1) - solo método xSDS
- DAPI (Thermo Fisher, Cat. D1306)
- Tampón TD de kit Nextera® (Illumina, Cat. FC-121-1031)
- 55 · 96 transposomas empobrecidos en citosina indexados (ensamblados utilizando métodos publicados, las secuencias se muestran en la Tabla 1)
- Cebador aleatorio de 9 nucleótidos (Tabla 2)
- 60 · Mezcla de dNTP 10 mM (NEB, Cat. N0447)
- Klenow (3'→5' Exo-) polimerasa (Enzymatics, Cat. P7010-LC-L)
- 65 · Etanol de prueba 200

ES 3 014 783 T3

- Cebadores de PCR indexados i5 e i7 (Tabla 3)
 - Kapa HiFi™ HotStart Mezcla preparada
 - 5 · SYBR® verde (FMC BioProducts, Cat. 50513)
 - QIAquick® kit de purificación de PCR (Qiagen, Cat. 28104)
 - ADNbc de alta sensibilidad Qubit® (Thermo Fisher, Cat. Q32851)
 - 10 · Kit Bioanalizador de Alta Sensibilidad (Agilent, Cat. 5067-4626)
 - Kit de secuenciación NextSeq (ciclo alto o medio de 150)
 - 15 · ADN Lambda no metilado (Promega, Cat. D1521)
 - Kit de secuenciación HiSeq® 2500 (Illumina)
 - Kit de secuenciación HiSeq® X (Illumina)
 - 20 · Kit MagPrep de metilación de ADN EZ-96 (Zymo Research, Cat D5040)
 - Cebadores de secuenciación de LNA personalizados (Tabla 4)
 - 25 · Polietilenglicol (PEG)
 - Perlas SPRI
- Equipo utilizado en los ejemplos
- 30 · Filtro celular de 35 µM (BD Biosciences, Cat. 352235)
 - Gradilla magnética compatible con placas de 96 pocillos
 - 35 · Clasificador de células Sony SH800 (Sony Biotechnology, Cat. SH800) u otro instrumento FACS capaz de clasificar núcleos individuales basándose en DAPI
 - Termociclador CFX Connect RT (Bio-Rad, Cat. 1855200) u otro termociclador en tiempo real
 - 40 · Termomezclador
 - Fluorómetro Qubit® 2.0 (Thermo Fisher, Cat. Q32866)
 - Bioanalizador 2100 (Agilent, Cat. G2939A)
 - 45 · NextSeq® 500 (Illumina, Cat. SY-415-1001-1)
 - HiSeq® 2500 (Illumina)
 - 50 · HiSeq® X (Illumina)

Oligonucleótidos utilizados en los ejemplos

Tabla 1: Oligonucleótidos cargados con transposasa sciMET (5'-3')

Complemento inverso:	(5fos) CTGTCTCTTATACACATCT			
	Nombre	i5_bsPCR	Índice	i5_Tn5
sciMET_Tn5_1	GGTGTAGTGGGTTTGG	GTTAAGAGGAA	TGGTAGAGAGGGTG	AGATGTGTATAAGAGACAG
sciMET_Tn5_2	GGTGTAGTGGGTTTGG	AGTAGGAAGAT	TGGTAGAGAGGGTG	AGATGTGTATAAGAGACAG
sciMET_Tn5_3	GGTGTAGTGGGTTTGG	GAATTAGGTGT	TGGTAGAGAGGGTG	AGATGTGTATAAGAGACAG
sciMET_Tn5_4	GGTGTAGTGGGTTTGG	GGAGATTAATG	TGGTAGAGAGGGTG	AGATGTGTATAAGAGACAG
sciMET_Tn5_5	GGTGTAGTGGGTTTGG	TATTGTGGAAT	TGGTAGAGAGGGTG	AGATGTGTATAAGAGACAG

ES 3 014 783 T3

Complemento inverso:	(5fos) CTGTCTCTTATACACATCT			
Nombre	i5_bsPCR	Índice		i5_Tn5
sciMET_Tn5_6	GGTGTAGTGGGTTTGG	ATATAGATGAT	TGGTAGAGAGGGTG	AGATGTGTATAAGAGACAG
sciMET_Tn5_7	GGTGTAGTGGGTTTGG	GTAAGAGGAAT	TGGTAGAGAGGGTG	AGATGTGTATAAGAGACAG
sciMET_Tn5_8	GGTGTAGTGGGTTTGG	GAGAGTTATTG	TGGTAGAGAGGGTG	AGATGTGTATAAGAGACAG
sciMET_Tn5_9	GGTGTAGTGGGTTTGG	AGTTAGTGTGA	TGGTAGAGAGGGTG	AGATGTGTATAAGAGACAG
sciMET_Tn5_10	GGTGTAGTGGGTTTGG	GATATAGAATT	TGGTAGAGAGGGTG	AGATGTGTATAAGAGACAG
sciMET_Tn5_11	GGTGTAGTGGGTTTGG	AAGGAAGTGAA	TGGTAGAGAGGGTG	AGATGTGTATAAGAGACAG
sciMET_Tn5_12	GGTGTAGTGGGTTTGG	AATAAGGAAGG	TGGTAGAGAGGGTG	AGATGTGTATAAGAGACAG
sciMET_Tn5_13	GGTGTAGTGGGTTTGG	GTATGGATATA	TGGTAGAGAGGGTG	AGATGTGTATAAGAGACAG
sciMET_Tn5_14	GGTGTAGTGGGTTTGG	TTAGATAATGA	TGGTAGAGAGGGTG	AGATGTGTATAAGAGACAG
sciMET_Tn5_15	GGTGTAGTGGGTTTGG	GGTGTGTAAAT	TGGTAGAGAGGGTG	AGATGTGTATAAGAGACAG
sciMET_Tn5_16	GGTGTAGTGGGTTTGG	GAAGTGGAGAG	TGGTAGAGAGGGTG	AGATGTGTATAAGAGACAG
sciMET_Tn5_17	GGTGTAGTGGGTTTGG	TTGAGTGGTAG	TGGTAGAGAGGGTG	AGATGTGTATAAGAGACAG
sciMET_Tn5_18	GGTGTAGTGGGTTTGG	GATAATGGTGA	TGGTAGAGAGGGTG	AGATGTGTATAAGAGACAG
sciMET_Tn5_19	GGTGTAGTGGGTTTGG	GTGTTAATGGA	TGGTAGAGAGGGTG	AGATGTGTATAAGAGACAG
sciMET_Tn5_20	GGTGTAGTGGGTTTGG	TAGGAATGGTG	TGGTAGAGAGGGTG	AGATGTGTATAAGAGACAG
sciMET_Tn5_21	GGTGTAGTGGGTTTGG	ATGTATGGATA	TGGTAGAGAGGGTG	AGATGTGTATAAGAGACAG
sciMET_Tn522	GGTGTAGTGGGTTTGG	TGATTGTTGGT	TGGTAGAGAGGGTG	AGATGTGTATAAGAGACAG
sciMET_Tn5_23	GGTGTAGTGGGTTTGG	AAGAGAATTAT	TGGTAGAGAGGGTG	AGATGTGTATAAGAGACAG
sciMET_Tn5_24	GGTGTAGTGGGTTTGG	AATGGTTGGTA	TGGTAGAGAGGGTG	AGATGTGTATAAGAGACAG
sciMET_Tn5_25	GGTGTAGTGGGTTTGG	GGTTAATTGAG	TGGTAGAGAGGGTG	AGATGTGTATAAGAGACAG
sciMET_Tn5_26	GGTGTAGTGGGTTTGG	GTATAATAGTT	TGGTAGAGAGGGTG	AGATGTGTATAAGAGACAG
sciMET_Tn5_27	GGTGTAGTGGGTTTGG	TTAGTTGAATT	TGGTAGAGAGGGTG	AGATGTGTATAAGAGACAG
sciMET_Tn5_28	GGTGTAGTGGGTTTGG	TTGGTGAAGGT	TGGTAGAGAGGGTG	AGATGTGTATAAGAGACAG
sciMET_Tn5_29	GGTGTAGTGGGTTTGG	TTAATATTGAA	TGGTAGAGAGGGTG	AGATGTGTATAAGAGACAG
sciMET_Tn5_30	GGTGTAGTGGGTTTGG	GTTAGAATTGG	TGGTAGAGAGGGTG	AGATGTGTATAAGAGACAG
sciMET_Tn5_31	GGTGTAGTGGGTTTGG	GTTATTAATTA	TGGTAGAGAGGGTG	AGATGTGTATAAGAGACAG
sciMET_Tn5_32	GGTGTAGTGGGTTTGG	GATTGGTAAGA	TGGTAGAGAGGGTG	AGATGTGTATAAGAGACAG
sciMET_Tn5_33	GGTGTAGTGGGTTTGG	TGAAGTATTGT	TGGTAGAGAGGGTG	AGATGTGTATAAGAGACAG
sciMET_Tn5_34	GGTGTAGTGGGTTTGG	GATGGATTATG	TGGTAGAGAGGGTG	AGATGTGTATAAGAGACAG
sciMET_Tn5_35	GGTGTAGTGGGTTTGG	ATTAGTATATT	TGGTAGAGAGGGTG	AGATGTGTATAAGAGACAG
sciMET_Tn5_36	GGTGTAGTGGGTTTGG	GTAGGTGTGGT	TGGTAGAGAGGGTG	AGATGTGTATAAGAGACAG
sciMET_Tn5_37	GGTGTAGTGGGTTTGG	AGTTGAATGTA	TGGTAGAGAGGGTG	AGATGTGTATAAGAGACAG
sciMET_Tn5_38	GGTGTAGTGGGTTTGG	ATTGTGAGATA	TGGTAGAGAGGGTG	AGATGTGTATAAGAGACAG
sciMET_Tn5_39	GGTGTAGTGGGTTTGG	TTGTGGTGAAT	TGGTAGAGAGGGTG	AGATGTGTATAAGAGACAG
sciMET_Tn540	GGTGTAGTGGGTTTGG	TTAAGTTGGTT	TGGTAGAGAGGGTG	AGATGTGTATAAGAGACAG
sciMET_Tn5_41	GGTGTAGTGGGTTTGG	TATAATAATAT	TGGTAGAGAGGGTG	AGATGTGTATAAGAGACAG
sciMET_Tn5_42	GGTGTAGTGGGTTTGG	AAGGTATGAGT	TGGTAGAGAGGGTG	AGATGTGTATAAGAGACAG
sciMET_Tn5_43	GGTGTAGTGGGTTTGG	AGGATTATAAG	TGGTAGAGAGGGTG	AGATGTGTATAAGAGACAG
sciMET_Tn5_44	GGTGTAGTGGGTTTGG	AGAGTTAGGTT	TGGTAGAGAGGGTG	AGATGTGTATAAGAGACAG
sciMET_Tn5_45	GGTGTAGTGGGTTTGG	ATGGATAGTAT	TGGTAGAGAGGGTG	AGATGTGTATAAGAGACAG
sciMET_Tn5_46	GGTGTAGTGGGTTTGG	ATATTATGTTG	TGGTAGAGAGGGTG	AGATGTGTATAAGAGACAG
sciMET_Tn5_47	GGTGTAGTGGGTTTGG	GGTGGAGATAG	TGGTAGAGAGGGTG	AGATGTGTATAAGAGACAG
sciMET_Tn5_48	GGTGTAGTGGGTTTGG	TGGTGGTAGTG	TGGTAGAGAGGGTG	AGATGTGTATAAGAGACAG

ES 3 014 783 T3

Complemento inverso:	(5fos) CTGTCTCTTATACACATCT			
Nombre	i5_bsPCR	Índice	i5_Tn5	
sciMET_Tn5_49	GGTGTAGTGGGTTTGG	AGGTGAGAAAGT	TGGTAGAGAGGGTG	AGATGTGTATAAGAGACAG
sciMET_Tn5_50	GGTGTAGTGGGTTTGG	TAGGAGGTTGT	TGGTAGAGAGGGTG	AGATGTGTATAAGAGACAG
sciMET_Tn5_51	GGTGTAGTGGGTTTGG	TGTATAGGTAT	TGGTAGAGAGGGTG	AGATGTGTATAAGAGACAG
sciMET_Tn5_52	GGTGTAGTGGGTTTGG	TGTTATGTAGA	TGGTAGAGAGGGTG	AGATGTGTATAAGAGACAG
sciMET_Tn5_53	GGTGTAGTGGGTTTGG	TGGAAGGTATG	TGGTAGAGAGGGTG	AGATGTGTATAAGAGACAG
sciMET_Tn5_54	GGTGTAGTGGGTTTGG	AATGTAAGGAG	TGGTAGAGAGGGTG	AGATGTGTATAAGAGACAG
sciMET_Tn5_55	GGTGTAGTGGGTTTGG	GTTATGTTAAG	TGGTAGAGAGGGTG	AGATGTGTATAAGAGACAG
sciMET_Tn5_56	GGTGTAGTGGGTTTGG	TGTTATAGGTG	TGGTAGAGAGGGTG	AGATGTGTATAAGAGACAG
sciMET_Tn5_57	GGTGTAGTGGGTTTGG	AAGGAGAATTG	TGGTAGAGAGGGTG	AGATGTGTATAAGAGACAG
sciMET_Tn5_58	GGTGTAGTGGGTTTGG	AGAGGTGGAAG	TGGTAGAGAGGGTG	AGATGTGTATAAGAGACAG
sciMET_Tn5_59	GGTGTAGTGGGTTTGG	GATTAGGTGTA	TGGTAGAGAGGGTG	AGATGTGTATAAGAGACAG
sciMET_Tn5_60	GGTGTAGTGGGTTTGG	ATTATATAAGA	TGGTAGAGAGGGTG	AGATGTGTATAAGAGACAG
sciMET_Tn5_61	GGTGTAGTGGGTTTGG	GAGAATATGGT	TGGTAGAGAGGGTG	AGATGTGTATAAGAGACAG
sciMET_Tn5_62	GGTGTAGTGGGTTTGG	GGATTGAGAGG	TGGTAGAGAGGGTG	AGATGTGTATAAGAGACAG
sciMET_Tn5_63	GGTGTAGTGGGTTTGG	ATTATGGTGGT	TGGTAGAGAGGGTG	AGATGTGTATAAGAGACAG
sciMET_Tn5_64	GGTGTAGTGGGTTTGG	GAAGGAAGTTA	TGGTAGAGAGGGTG	AGATGTGTATAAGAGACAG
sciMET_Tn5_65	GGTGTAGTGGGTTTGG	GAATATGTAAG	TGGTAGAGAGGGTG	AGATGTGTATAAGAGACAG
sciMET_Tn5_66	GGTGTAGTGGGTTTGG	TAGTTAATATT	TGGTAGAGAGGGTG	AGATGTGTATAAGAGACAG
sciMET_Tn5_67	GGTGTAGTGGGTTTGG	TGAATGAATAG	TGGTAGAGAGGGTG	AGATGTGTATAAGAGACAG
sciMET_Tn5_68	GGTGTAGTGGGTTTGG	AGGATGGATTA	TGGTAGAGAGGGTG	AGATGTGTATAAGAGACAG
sciMET_Tn5_69	GGTGTAGTGGGTTTGG	AAGTGTATAGA	TGGTAGAGAGGGTG	AGATGTGTATAAGAGACAG
sciMET_Tn5_70	GGTGTAGTGGGTTTGG	GAGGTTGAAGA	TGGTAGAGAGGGTG	AGATGTGTATAAGAGACAG
sciMET_Tn5_71	GGTGTAGTGGGTTTGG	TGTGTAATAGG	TGGTAGAGAGGGTG	AGATGTGTATAAGAGACAG
sciMET_Tn5_72	GGTGTAGTGGGTTTGG	TTGATTAGAGA	TGGTAGAGAGGGTG	AGATGTGTATAAGAGACAG
sciMET_Tn5_73	GGTGTAGTGGGTTTGG	TATGTGTGTGG	TGGTAGAGAGGGTG	AGATGTGTATAAGAGACAG
sciMET_Tn5_74	GGTGTAGTGGGTTTGG	GAGATGAGAAT	TGGTAGAGAGGGTG	AGATGTGTATAAGAGACAG
sciMET_Tn5_75	GGTGTAGTGGGTTTGG	TGGTGAAGTGA	TGGTAGAGAGGGTG	AGATGTGTATAAGAGACAG
sciMET_Tn5_76	GGTGTAGTGGGTTTGG	GTGGTAGGATG	TGGTAGAGAGGGTG	AGATGTGTATAAGAGACAG
sciMET_Tn5_77	GGTGTAGTGGGTTTGG	TGTAGGTGATA	TGGTAGAGAGGGTG	AGATGTGTATAAGAGACAG
sciMET_Tn5_78	GGTGTAGTGGGTTTGG	GTAAGGTGTGA	TGGTAGAGAGGGTG	AGATGTGTATAAGAGACAG
sciMET_Tn5_79	GGTGTAGTGGGTTTGG	AGAAGAGAGTG	TGGTAGAGAGGGTG	AGATGTGTATAAGAGACAG
sciMET_Tn5_80	GGTGTAGTGGGTTTGG	GGATGTTGTAT	TGGTAGAGAGGGTG	AGATGTGTATAAGAGACAG
sciMET_Tn5_81	GGTGTAGTGGGTTTGG	AAGTTATATAA	TGGTAGAGAGGGTG	AGATGTGTATAAGAGACAG
sciMET_Tn5_82	GGTGTAGTGGGTTTGG	TGGAATTAAGT	TGGTAGAGAGGGTG	AGATGTGTATAAGAGACAG
sciMET_Tn5_83	GGTGTAGTGGGTTTGG	TAATGAGAGGA	TGGTAGAGAGGGTG	AGATGTGTATAAGAGACAG
sciMET_Tn5_84	GGTGTAGTGGGTTTGG	ATAATTGATGG	TGGTAGAGAGGGTG	AGATGTGTATAAGAGACAG
sciMET_Tn5_85	GGTGTAGTGGGTTTGG	TGTGAAGAGTA	TGGTAGAGAGGGTG	AGATGTGTATAAGAGACAG
sciMET_Tn5_86	GGTGTAGTGGGTTTGG	GATGAATATGT	TGGTAGAGAGGGTG	AGATGTGTATAAGAGACAG
sciMET_Tn5_87	GGTGTAGTGGGTTTGG	TGAGGATAGAT	TGGTAGAGAGGGTG	AGATGTGTATAAGAGACAG
sciMET_Tn5_88	GGTGTAGTGGGTTTGG	ATTAATTAGAG	TGGTAGAGAGGGTG	AGATGTGTATAAGAGACAG
sciMET_Tn5_89	GGTGTAGTGGGTTTGG	GGAGAGATGGA	TGGTAGAGAGGGTG	AGATGTGTATAAGAGACAG
sciMET_Tn5_90	GGTGTAGTGGGTTTGG	TAATTGAGGAA	TGGTAGAGAGGGTG	AGATGTGTATAAGAGACAG
sciMET_Tn5_91	GGTGTAGTGGGTTTGG	TTGGAATTAAT	TGGTAGAGAGGGTG	AGATGTGTATAAGAGACAG

ES 3 014 783 T3

Complemento inverso:		(5fos) CTGTCTCTTATACACATCT		
Nombre	i5_bsPCR	Índice	i5_Tn5	
sciMET_Tn5_92	GGTGTAGTGGGTTTGG	AATGTTATTGT	TGGTAGAGAGGGTG	AGATGTGTATAAGAGACAG
sciMET_Tn5_93	GGTGTAGTGGGTTTGG	GTAGTTATTAG	TGGTAGAGAGGGTG	AGATGTGTATAAGAGACAG
sciMET_Tn5_94	GGTGTAGTGGGTTTGG	TATATTGTGAG	TGGTAGAGAGGGTG	AGATGTGTATAAGAGACAG
sciMET_Tn5_95	GGTGTAGTGGGTTTGG	GTGTAGGATAG	TGGTAGAGAGGGTG	AGATGTGTATAAGAGACAG
sciMET_Tn5_96	GGTGTAGTGGGTTTGG	AGAGAAGTTGG	TGGTAGAGAGGGTG	AGATGTGTATAAGAGACAG

Tabla 2: Cebador aleatorio de 9 nucleótidos sciMET (5'-3')

Nombre	Secuencia
sciMET_N9_IPE2	GGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNN

5 Tabla 3: Cebadores de PCR sciMET (5'-3')

Nombre	Secuencia
sciMET_i7_1	CAAGCAGAAGACGGCATAACGAGATcaagatgccgGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
sciMET_i7_2	CAAGCAGAAGACGGCATAACGAGATaacgtctagtGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
sciMET_i7_3	CAAGCAGAAGACGGCATAACGAGATaggatatactGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
sciMET_i7_4	CAAGCAGAAGACGGCATAACGAGATtcataggacGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
sciMET_i7_5	CAAGCAGAAGACGGCATAACGAGATggaggcctccGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
sciMET_i7_6	CAAGCAGAAGACGGCATAACGAGATtcaatataaGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
sciMET_i7_7	CAAGCAGAAGACGGCATAACGAGATacgtcatataGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
sciMET_i7_8	CAAGCAGAAGACGGCATAACGAGATtgaccagggaGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
sciMET_i7_9	CAAGCAGAAGACGGCATAACGAGATcggttgcgGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
sciMET_i7_10	CAAGCAGAAGACGGCATAACGAGATcaaggaggctGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
sciMET_i7_11	CAAGCAGAAGACGGCATAACGAGATttacgatgaaGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
sciMET_i7_12	CAAGCAGAAGACGGCATAACGAGATttgctggcatGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
sciMET_i7_13	CAAGCAGAAGACGGCATAACGAGATaatactctcGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
sciMET_i7_14	CAAGCAGAAGACGGCATAACGAGATccaactaacGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
sciMET_i7_15	CAAGCAGAAGACGGCATAACGAGATtatctcaatGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
sciMET_i7_16	CAAGCAGAAGACGGCATAACGAGATgccgtcgcGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
sciMET_i7_17	CAAGCAGAAGACGGCATAACGAGATccgctgctcGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
sciMET_i7_18	CAAGCAGAAGACGGCATAACGAGATtgaccgaatcGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
sciMET_i7_19	CAAGCAGAAGACGGCATAACGAGATgtctccagagGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
sciMET_i7_20	CAAGCAGAAGACGGCATAACGAGATaatgetagteGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
sciMET_i7_21	CAAGCAGAAGACGGCATAACGAGATgacgacctgcGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
sciMET_i7_22	CAAGCAGAAGACGGCATAACGAGATagagccagccGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
sciMET_i7_23	CAAGCAGAAGACGGCATAACGAGATeeaggegeaGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
sciMET_i7_24	CAAGCAGAAGACGGCATAACGAGATcaggatggaGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
sciMET_i7_1	AATGATACGGCGACCACCGAGATCTACACgatatcatcgaGGTGTAGTGGGTTTGG
sciMET_i7_2	AATGATACGGCGACCACCGAGATCTACACcccgattatGGTGTAGTGGGTTTGG
sciMET_i7_3	AATGATACGGCGACCACCGAGATCTACACattcaggtagGGTGTAGTGGGTTTGG
sciMET_i7_4	AATGATACGGCGACCACCGAGATCTACACatggaattggGGTGTAGTGGGTTTGG
sciMET_i7_5	AATGATACGGCGACCACCGAGATCTACACgacgaagcgtGGTGTAGTGGGTTTGG
sciMET_i7_6	AATGATACGGCGACCACCGAGATCTACACettgeagtagGGTGTAGTGGGTTTGG
sciMET_i7_7	AATGATACGGCGACCACCGAGATCTACACettgtaatgGGTGTAGTGGGTTTGG

ES 3 014 783 T3

Nombre	Secuencia
sciMET_i7_8	AATGATACGGCGACCAACCGAGATCTACACcaagtgaccGGTGTAGTGGGTTTGG

Tabla 4: Cebadores de secuenciación sciMET (LNA, 5'-3')

Nombre	Secuencia
sciMET_Read1	TGGTAGAGAGGGTG AGATGTGTATAAGAGATAG
sciMET_lindex1	CTATCTCTTATACACATCT CACCCTCTCTACCA

5 Ejemplo 1

Preparación de ADN Lambda de control no metilado

10 Se combinaron cien nanogramos de ADN Lambda no metilado, 5 µl de tampón TD 2X, 5 µl de tampón NIB (Tris-HCl 10 mM pH 7,4, NaCl 10 mM, MgCl₂ 3 mM, Igepal® al 0,1 %, 1 inhibidor de proteasas) y 4 µl de transposoma empobrecido en citosina indexado de forma única a 500 nM. La mezcla se incubó durante 20 minutos a 55 °C y después se purificó utilizando la columna de purificación de PCR QIAquick® y se eluyó en 30 µl de EB.

15 La concentración de ADN se cuantificó con un fluorómetro Qubit 2.0 para ADNbc de alta sensibilidad utilizando 2 ul de la mezcla. La concentración se diluyó a 17,95 pg/ul, lo que simula la masa genómica de aproximadamente 5 células humanas.

Ejemplo 2

20 Preparación de una mezcla de perlas SPRI PEG al 18 %

25 Perlas Sera-Mag (1 ml) se dividieron en alícuotas en un tubo de 1,5 ml de unión baja y después se colocaron en un soporte magnético hasta que se limpió el sobrenadante. Las perlas se lavaron con una solución de 500 ul de Tris-HCl 10 mM, pH 8,0, y la solución se eliminó después de que se limpiase el sobrenadante, y esta etapa de lavado se repitió durante un total de cuatro lavados. Las perlas se resuspendieron en la siguiente mezcla: PEG 8000 al 18 % (en masa), NaCl 1 M, Tris-HCl 10 mM, pH 8,0, EDTA 1 mM, Tween-20 al 0,05 %; se incubaron a temperatura ambiente con agitación suave durante al menos una hora y después se almacenaron perlas de SPRI PEG al 18 % a 4 °C. Se permitió que las perlas alcanzaran la temperatura ambiente antes de su uso.

30 Ejemplo 3

Preparación de núcleos utilizando ácido 3,5-diyodosalicílico de litio (LAND) o SDS (xSDS)

35 A. Método LAND de preparación de núcleos y agotamiento de nucleosomas

Si las células estaban en un cultivo celular en suspensión, el cultivo se trituró suavemente hasta romper los grumos de células, las células se sedimentaron centrifugando a 500xg durante 5 minutos a 4 °C y se lavaron con 500 µl de PBS enfriado con hielo.

40 Si las células estaban en un cultivo celular adherente, se aspiraba el medio y se lavaban las células con 10 ml de PBS a 37 °C, y después se añadía suficiente tripsina al 0,25 % a 37 °C para cubrir la monocapa. Después de incubar a 37 °C durante 5 minutos o hasta que el 90 % de las células ya no se adhirieran a la superficie, se añadió medio a 37 °C en una proporción de 1:1 para desactivar la tripsina. Las células se sedimentaron centrifugando a 500xg durante 5 minutos a 4 °C y después se lavaron con 500 µl de PBS enfriado con hielo.

45 Las células del cultivo celular en suspensión o del cultivo celular adherente se sedimentaron centrifugando a 500xg durante 5 minutos y después se resuspendieron en 200 µl de LIS 12,5 mM en tampón NIB (2,5 µl de LIS 1 M + 197,5 µl de tampón NIB). Después de incubar en hielo durante 5 minutos, se añadieron 800 µl de tampón NIB. Las células se pasaron suavemente a través de un filtro celular de 35 µm y se añadieron 5 µl de DAPI (5 mg/ml).

50 B. Método xSDS de preparación de núcleos y empobrecimiento de nucleosomas

55 Si las células estaban en un cultivo celular en suspensión, el medio se trituró suavemente hasta romper los grumos de células. A 10 ml de células en medio se añadieron 406 µl de formaldehído al 37 % y se incubaron a temperatura ambiente durante 10 minutos con agitación suave. A las células se les añadieron 800 microlitros de glicina 2,5 M y se incubaron en hielo durante 5 minutos, y después se centrifugaron a 550xg durante 8 minutos a 4 °C. Después de lavar con 10 ml de PBS enfriado con hielo, las células se resuspendieron en 5 ml de NIB enfriado con hielo (Tris-HCl 10 mM pH 7,4, NaCl 10 mM, MgCl₂ 3 mM, Igepal® al 0,1 %, 1 inhibidor de proteasas) y se incubaron en hielo durante 20 minutos mezclando suavemente.

Si las células estaban en un cultivo celular adherente, se aspiraba el medio y se lavaban las células con 10 ml de PBS a 37 °C, y después se añadía suficiente tripsina al 0,25 % a 37 °C para cubrir la monocapa. Después de incubar a 37 °C durante 5 minutos o hasta que el 90 % de las células ya no se adhiriera a la superficie, se añadió medio a 37 °C en una proporción de 1:1 para desactivar la tripsina, y el volumen se llevó a 10 ml con medio. Las células se resuspendieron en 10 ml de medio, se añadieron 406 µl de formaldehído al 37 % y se incubaron a temperatura ambiente durante 10 minutos con agitación suave. A las células se añadieron 800 microlitros de glicina 2,5 M y se incubaron en hielo durante 5 minutos. Las células se centrifugaron a 550xg durante 8 minutos a 4 ° y se lavaron con 10 ml de PBS enfriado con hielo. Después de resuspender las células en 5 ml de NIB enfriado con hielo, se incubaron en hielo durante 20 minutos con una mezcla suave.

Las células o los núcleos del cultivo celular en suspensión o del cultivo celular adherente, se sedimentaron centrifugando a 500xg durante 5 minutos y se lavaron con 900 µl de 1x NEBuffer 2.1. Después de sedimentar a 500 xg durante 5 minutos, el sedimento se resuspendió en 800 µl 1x NEBuffer 2.1 con 12 µl de SDS al 20 % y se incubó a 42 °C con intensa agitación durante 30 minutos, y después con 200 µl de Triton™ X-100 al 10 % y se incubó a 42 °C con agitación intensa durante 30 minutos. Las células se pasaron suavemente a través de un filtro celular de 35 µM y se añadieron 5 µl de DAPI (5 mg/ml).

Ejemplo 4

Clasificación y tagmentación de núcleos

Se preparó una placa de tagmentación con 10 µl de tampón TD 1x (para 1 placa: 500 µl de tampón NIB + 500 µl de tampón TD) y en cada pocillo de la placa de tagmentación se clasificaron 2500 núcleos individuales. En esta etapa, el número de núcleos por pocillo puede variar ligeramente siempre que el número de núcleos por pocillo sea constante en toda la placa. También es posible multiplexar diferentes muestras en diferentes pocillos de la placa ya que se conservará el índice de transposasa. Las células se seleccionaron según la Figura 2. Después de centrifugar la placa a 500 x g durante 5 min, se añadieron a cada pocillo 4 µl de transposoma empobrecido en citosina indexado de forma única a 500 nM. Después del sellado, la placa se incubó a 55 °C durante 15 minutos con agitación suave. A continuación, la placa se colocó en hielo. Todos los pocillos se agruparon y se pasaron a través de un filtro celular de 35 µM. Se añadieron cinco microlitros de DAPI (5 mg/ml).

Ejemplo 5

Segunda clasificación de núcleos

Se preparó una mezcla maestra para cada pocillo con 5 µl de reactivo de digestión Zymo (2,5 µl de tampón de digestión M, 2,25 µl de H₂O y 0,25 µl de proteinasa K). Se clasificaron 10 o 22 núcleos individuales en cada pocillo utilizando los ajustes de clasificación más estrictos. Diez núcleos individuales se clasificaron en los pocillos que se utilizarán como enriquecimiento de control no metilado, y en los demás pocillos se clasificaron 22 células. A continuación, la placa se centrifugó a 600 x g durante 5 min a 4 °C.

Ejemplo 6

Digestión y conversión de bisulfito

Se utilizaron aproximadamente ~35 pg (2 ul) de ADN Lambda de control no metilado pretratado con un transposoma empobrecido en C para enriquecer los pocillos con 10 núcleos individuales. La placa se incubó durante 20 minutos a 50 °C para digerir los núcleos y se añadieron 32,5 ul de reactivo de conversión Zymo CT recién preparado siguiendo el protocolo del fabricante. Los pocillos se mezclaron por trituración y la placa se centrifugó a 600 x g durante 2 min a 4 °C. La placa se colocó en un termociclador durante las siguientes etapas antes de continuar: 98 °C durante 8 minutos, 64 °C durante 3,5 horas, después se mantuvo a 4 °C durante menos de 20 horas. Se añadieron perlas Zymo MagBinding (5 ul) a cada pocillo y 150 ul de tampón de unión M a cada pocillo. Después de mezclar los pocillos mediante trituración, la placa se incubó a temperatura ambiente durante 5 minutos. La placa se colocó en una gradilla magnética compatible con placas de 96 pocillos hasta que el sobrenadante se limpió.

Se eliminó el sobrenadante y los pocillos se lavaron con etanol reciente al 80 % (en volumen), i) retirando la placa de la gradilla magnética, ii) añadiendo 100 ul de etanol al 80 % a cada pocillo, pasando por encima del sedimento de perlas, y iii) colocando la placa de nuevo en la gradilla magnética y después eliminando el sobrenadante una vez limpio.

La desulfonación se realizó añadiendo 50 ul de tampón de desulfonación M a cada pocillo, resuspendiendo las perlas completamente por trituración, incubando a temperatura ambiente durante 15 minutos y colocando la placa en la gradilla magnética y eliminando el sobrenadante una vez limpio.

ES 3 014 783 T3

Se eliminó el sobrenadante y los pocillos se lavan con etanol reciente al 80 % (en volumen), i) retirando la placa de la gradilla magnética, ii) añadiendo 100 ul de etanol al 80 % a cada pocillo, pasando por encima del sedimento de perlas, y iii) colocando la placa de nuevo en la gradilla magnética y después eliminando el sobrenadante una vez limpio.

- 5 Los sedimentos de perlas se dejaron secar durante ~10 minutos hasta que los sedimentos comenzaron a agrietarse visiblemente.

La elución se realizó añadiendo 25 ul de tampón de elución Zymo M a cada pocillo, triturando para disociar completamente el sedimento y calentando la placa a 55 °C durante 4 minutos.

10

Ejemplo 7

Amplificación lineal

- 15 La elución completa se trasladó a una placa preparada con la siguiente mezcla de reacción por pocillo: 16 ul de H₂O PCR-clean, 5 ul de 10X NEBuffer 2.1, 2 ul de mezcla de dNTP 10 mM y 2 ul de cebador aleatorio de 9 nucleótidos 10 uM.

- 20 La amplificación lineal se realizó de la siguiente manera: i) convertir el ADN en monocatenario incubando a 95 °C durante 45 segundos, después enfriar rápidamente en hielo y mantener en hielo, ii) añadir 10 U de (3'->5' exo-) polimerasa Klenow a cada pocillo una vez que se haya enfriado totalmente y iii) incubar la placa a 4 °C durante 5 minutos, después aumentar la temperatura a una velocidad de +1 °C/15 segundos a 37 °C, después mantener a 37 °C durante 90 minutos.

- 25 Las etapas i-iii se repitieron tres veces más durante un total de cuatro rondas de amplificación lineal. Para cada amplificación, a la reacción de cada pocillo se añadió la siguiente mezcla: 1 ul de cebador aleatorio de 9 nucleótidos de 10 uM, 1 ul de mezcla de dNTP 10 mM y 1,25 ul de 4X NEBuffer 2.1. En comparación con menos rondas, cuatro rondas de amplificación lineal normalmente aumentan significativamente la tasa de alineación de lectura y la complejidad de la biblioteca.

30

- Los pocillos se limpiaron de la siguiente manera utilizando la mezcla de perlas de SPRI PEG al 18 % preparada a 1,1X (concentración por volumen en comparación con el volumen de reacción del pocillo). La placa se incubó durante 5 minutos a temperatura ambiente, se colocó en la gradilla magnética y se eliminó el sobrenadante una vez limpio. Los sedimentos de perlas se lavaron con 50 ul de etanol al 80 %. Se eliminó cualquier líquido restante y se dejó secar el sedimento de perlas hasta que comenzó a agrietarse. El ADN se eluyó en 21 µl de Tris-Cl 10 mM (pH 8,5).

35

Ejemplo 8

Reacción de indexación por PCR

40

- La elución completa se trasladó a una placa preparada con la siguiente mezcla de reacción por pocillo: 2 ul de cebador de PCR de índice i7 10 uM, 2 ul de cebador de PCR de índice i5 10 uM, 25 ul de 2X KAPA HiFi™ HotStart ReadyMix y 0,5 ul de 100X SYBR® Green I. La amplificación por PCR se realizó en un termociclador en tiempo real con los siguientes ciclos: 95 °C durante 2 minutos (94 °C durante 80 segundos, 65 °C durante 30 segundos, 72 °C durante 30 segundos), y la reacción se detuvo una vez que la mayoría de los pocillos mostraron una inflexión de la fluorescencia SYBR® Green medida. Se observaron mesetas de inflexión entre 16 y 21 ciclos de PCR en las preparaciones de las bibliotecas.

45

Ejemplo 9

50

Limpieza y cuantificación de las bibliotecas

- Los pocillos de las bibliotecas se limpiaron utilizando la mezcla de perlas SPRI PEG al 18 % a 0,8X (concentración por volumen en comparación con el volumen de reacción del pocillo) de la siguiente manera. La placa se incubó durante 5 minutos a temperatura ambiente, se colocó en la gradilla magnética y el sobrenadante se eliminó una vez limpio. Los sedimentos de perlas se lavaron con 50 ul de etanol al 80 %. Se eliminó cualquier líquido restante y se dejó secar el sedimento de perlas hasta que comenzó a agrietarse. El ADN se eluyó en 25 ul de Tris-Cl 10 mM (pH 8,5).

55

- Las bibliotecas se agruparon utilizando 5 ul de cada pocillo, y se utilizaron 2 ul para cuantificar la concentración de ADN con un fluorómetro Qubit® 2.0 para ADNbc de alta sensibilidad siguiendo protocolo del fabricante. La lectura del Qubit® se utilizó para diluir la biblioteca a ~4 ng/ul, y se procesó 1 ul en un Bioanalizador 2100 de alta sensibilidad, siguiendo el protocolo del fabricante. A continuación, para la secuenciación de Illumina, la biblioteca se cuantificó en el intervalo de 200 pb - 1 kpb para diluir el grupo a 1 nM.

60

Ejemplo 10

Secuenciación

- 5 Siguiendo las instrucciones del fabricante, se configuró un NextSeq® 500 para un experimento de una muestra de 1 nM, excepto por los siguientes cambios. El grupo de bibliotecas se cargó a una concentración de 0,9 pM y a un volumen total de 1,5 ml y se depositó en la posición 10 del cartucho; los cebadores personalizados se configuraron diluyendo 9 µl de cebador 1 de secuenciación de reserva 100 µM en un total de 1,5 ml de tampón HT1 en la posición 7 del cartucho y 18 µl de cada cebador de secuenciación de índice personalizado a concentraciones de reserva de 100 µM hasta un total de 3 ml de tampón HT1 en la posición 9 del cartucho; el NextSeq® 500 funcionaba en modo independiente; la receta química personalizada de SC1seq (Amini y col. 2014, Nat. Genet. 46, 1343-1349) se seleccionó; se seleccionó doble índice; se introdujo el número apropiado de ciclos de lectura (se recomiendan 150); 10 ciclos para el índice 1 y 20 ciclos para el índice 2; se seleccionó la casilla de verificación personalizada de todas las lecturas e índices.
- 10
- 15 La descripción detallada y los ejemplos anteriores se han facilitado únicamente para facilitar la comprensión. De ello no se desprende ninguna limitación innecesaria.
- 20 Salvo que se indique lo contrario, se debe entender que todos los números que expresan cantidades de componentes, pesos moleculares, etc. utilizados en la memoria descriptiva y las reivindicaciones están modificados en todos los casos por el término “aproximadamente”. En consecuencia, salvo que se indique lo contrario, los parámetros numéricos expuestos en la memoria descriptiva y en las reivindicaciones son aproximaciones que pueden variar dependiendo de las propiedades deseadas que se busca obtener mediante la presente invención.
- 25 A pesar de que los intervalos y parámetros numéricos que exponen el amplio alcance de la invención son aproximaciones, los valores numéricos establecidos en los ejemplos específicos se informan con la mayor precisión posible. Sin embargo, todos los valores numéricos contienen inherentemente un intervalo que es el resultado necesariamente de la desviación estándar encontrada en sus respectivas mediciones de prueba.
- 30 Todos los encabezados son por comodidad de uso del lector y no deben usarse para limitar el significado del texto que sigue al encabezado, salvo que así se especifique.

REIVINDICACIONES

1. Un método para preparar una biblioteca de secuenciación para determinar el estado de metilación de los ácidos nucleicos de una pluralidad de células individuales, comprendiendo el método:

- (a) proporcionar núcleos aislados de una pluralidad de células;
- (b) distribuir subconjuntos de los núcleos hacia una primera pluralidad de compartimentos que comprende un complejo transposómico, en donde el complejo transposómico de cada compartimento comprende una primera secuencia índice que es diferente de las primeras secuencias índice en los otros compartimentos;
- (c) fragmentar ácidos nucleicos en los subconjuntos de núcleos hacia una pluralidad de fragmentos de ácido nucleico e incorporar las primeras secuencias índice hacia al menos una cadena de los fragmentos de ácido nucleico para generar núcleos indexados;
- (d) combinar los núcleos indexados para generar núcleos indexados agrupados;
- (e) distribuir subconjuntos de los núcleos indexados agrupados hacia una segunda pluralidad de compartimentos y someter los núcleos indexados a tratamiento para identificar nucleótidos metilados y generar fragmentos de ácido nucleico tratado, opcionalmente en donde el tratamiento comprende bisulfito o hidrazina;
- (f) amplificar los fragmentos de ácido nucleico tratados en cada compartimento mediante amplificación lineal con una pluralidad de cebadores que comprenden una secuencia de nucleótidos universal en el extremo 5' y una secuencia de nucleótidos aleatoria en el extremo 3' para generar moléculas adaptadoras de fragmentos amplificados;
- (g) incorporar una segunda secuencia índice hacia las moléculas adaptadoras de fragmentos amplificados para generar moléculas adaptadoras de fragmentos de doble índice, en donde la segunda secuencia índice de cada compartimento es diferente de las segundas secuencias índice en los otros compartimentos; y
- (h) combinar las moléculas adaptadoras de fragmentos de doble índice, produciendo de esta manera una biblioteca de secuenciación para determinar el estado de metilación de los ácidos nucleicos de la pluralidad de células individuales, opcionalmente en donde la distribución en las etapas (b) y (e) se realiza por clasificación de núcleos activados por fluorescencia.

2. El método de la reivindicación 1, en donde el tratamiento identifica secuencias de dinucleótidos CpG metiladas, o en donde la amplificación lineal de los fragmentos de ácido nucleico tratados comprende de 1 a 10 ciclos, o en donde la secuencia de nucleótidos universal en el extremo 5' de los cebadores en la etapa (f) comprende una segunda secuencia cebadora de secuenciación, o en donde la secuencia de nucleótidos aleatoria en el extremo 3' de los cebadores en la etapa (f) consiste en 9 nucleótidos aleatorios.

3. Un método para preparar una biblioteca de secuenciación para determinar el estado de metilación de los ácidos nucleicos de una pluralidad de células individuales, comprendiendo el método:

- (a) proporcionar núcleos aislados de una pluralidad de células;
- (b) distribuir subconjuntos de los núcleos hacia una primera pluralidad de compartimentos que comprende un complejo transposómico, en donde el complejo transposómico de cada compartimento comprende una primera secuencia índice que es diferente de las primeras secuencias índice en los otros compartimentos;
- (c) fragmentar ácidos nucleicos en los subconjuntos de núcleos hacia una pluralidad de fragmentos de ácido nucleico e incorporar las primeras secuencias índice hacia al menos una cadena de los fragmentos de ácido nucleico para generar núcleos indexados;
- (d) combinar los núcleos indexados para generar núcleos indexados agrupados;
- (e) distribuir subconjuntos de los núcleos indexados agrupados hacia una segunda pluralidad de compartimentos y someter los núcleos indexados a tratamiento para identificar nucleótidos metilados y generar fragmentos de ácido nucleico tratado, opcionalmente en donde el tratamiento comprende bisulfito o hidrazina;
- (f) ligar los fragmentos de ácido nucleico tratados en cada compartimento a un adaptador universal para generar moléculas adaptadoras de fragmentos ligadas;
- (g) incorporar una segunda secuencia índice en las moléculas adaptadoras de fragmentos ligadas para generar moléculas adaptadoras de fragmentos de doble índice, en donde la segunda secuencia índice en cada compartimento es diferente de las segundas secuencias índice en los otros compartimentos; y
- (h) combinar las moléculas adaptadoras de fragmentos de doble índice, produciendo de esta manera una biblioteca de secuenciación para determinar el estado de metilación de los ácidos nucleicos de la pluralidad de células individuales, opcionalmente en donde la distribución en las etapas (b) y (e) se realiza por clasificación de núcleos activados por fluorescencia.

ES 3 014 783 T3

4. El método de la reivindicación 1 o la reivindicación 3, en donde los subconjuntos de los núcleos comprenden aproximadamente el mismo número de núcleos, opcionalmente en donde los subconjuntos de los núcleos comprenden de 1 a aproximadamente 2000 núcleos.
5. El método de la reivindicación 1 o la reivindicación 3, en donde la primera pluralidad de compartimentos es una placa multipocillo, opcionalmente en donde la placa multipocillo es una placa de 96 pocillos o una placa de 384 pocillos, o en donde la segunda pluralidad de compartimentos es una placa multipocillo, opcionalmente en donde la placa multipocillo es una placa de 96 pocillos o una placa de 384 pocillos.
6. El método de la reivindicación 1 o la reivindicación 3, en donde los subconjuntos de los núcleos indexados agrupados comprenden aproximadamente el mismo número de núcleos, opcionalmente en donde los subconjuntos de los núcleos indexados agrupados comprenden de 1 a aproximadamente 25 núcleos, o en donde los subconjuntos de los núcleos indexados agrupados incluyen al menos 10 veces menos núcleos que los subconjuntos de los núcleos, opcionalmente en donde los subconjuntos de los núcleos indexados agrupados incluyen al menos 100 veces menos núcleos que los subconjuntos de los núcleos.
7. El método de la reivindicación 1 o la reivindicación 3, en donde cada uno de los complejos transposómicos comprende transposasas y transposones, comprendiendo cada uno de los transposones una cadena transferida, opcionalmente en donde la cadena transferida no comprende ningún residuo de citosina, opcionalmente además en donde la cadena transferida comprende la primera secuencia índice, opcionalmente además en donde la cadena transferida comprende además una primera secuencia universal y una primera secuencia cebadora de secuenciación.
8. El método de la reivindicación 3, en donde el tratamiento identifica secuencias de dinucleótidos CpG metiladas.
9. El método de la reivindicación 3, que comprende además añadir uno o más nucleótidos a los extremos 3' de los fragmentos de ácido nucleico tratados para crear un saliente 3' antes del ligamiento del adaptador universal, en donde opcionalmente se realiza la adición de uno o más nucleótidos utilizando una transferasa terminal, opcionalmente además en donde el adaptador universal comprende un saliente que es complementario inverso al saliente 3' en los fragmentos de ácido nucleico tratados.
10. El método de la reivindicación 1 o la reivindicación 3, en donde la incorporación de la segunda secuencia índice en la etapa (g) comprende poner en contacto las moléculas adaptadoras de fragmentos de doble índice de cada compartimento con un primer cebador universal y un segundo cebador universal, comprendiendo cada uno una secuencia índice, y realizar una reacción de amplificación exponencial, opcionalmente en donde la secuencia índice del primer cebador universal es el complemento inverso de la secuencia índice del segundo cebador universal, o en donde la secuencia índice del primer cebador universal es diferente del complemento inverso de la secuencia índice del segundo cebador universal.
11. El método de la reivindicación 1 o la reivindicación 10, en donde el primer cebador universal comprende además una primera secuencia de captura y una primera secuencia de anclaje complementaria a una secuencia universal en el extremo 3' de las moléculas adaptadoras de fragmentos de doble índice, opcionalmente en donde la primera secuencia de captura comprende la secuencia cebadora P5, o en donde el segundo cebador universal comprende además una segunda secuencia de captura y una segunda secuencia de anclaje complementaria a una secuencia universal en el extremo 5' de las moléculas adaptadoras de fragmentos de doble índice, opcionalmente en donde la segunda secuencia de captura comprende el complemento inverso de la secuencia cebadora P7.
12. El método de la reivindicación 1 o la reivindicación 10, en donde la reacción de amplificación exponencial comprende una reacción en cadena de la polimerasa (PCR), opcionalmente en donde la PCR comprende de 15 a 30 ciclos.
13. El método de la reivindicación 1 o la reivindicación 3, que comprende además un enriquecimiento de ácidos nucleicos diana utilizando una pluralidad de oligonucleótidos de captura que tienen especificidad por los ácidos nucleicos diana, opcionalmente en donde los oligonucleótidos de captura se inmovilizan en una superficie de un sustrato sólido, o en donde los oligonucleótidos de captura comprenden un primer miembro de un par de unión universal, y en donde un segundo miembro del par de unión está inmovilizado sobre una superficie de un sustrato sólido.
14. El método de la reivindicación 1 o la reivindicación 3, que comprende además la selección de las moléculas adaptadoras de fragmentos de doble índice que están dentro de un intervalo de tamaño predeterminado, o que comprende además la secuenciación de las moléculas adaptadoras de fragmentos de doble índice para determinar el estado de metilación de los ácidos nucleicos de la pluralidad de células individuales.
15. El método de la reivindicación 14 como dependiente de la reivindicación 1, que comprende además:

- proporcionar una superficie que comprenda una pluralidad de sitios de amplificación, en donde los sitios de amplificación comprenden al menos dos poblaciones de ácidos nucleicos monocatenarios unidos que tienen un extremo 3' libre, y
- 5 poner en contacto la superficie que comprende los sitios de amplificación con la biblioteca de secuenciación en condiciones adecuadas para producir una pluralidad de sitios de amplificación que comprende cada uno una población clonal de amplicones de una molécula adaptadora de fragmentos de doble índice individual, opcionalmente en donde el número de las moléculas adaptadoras de fragmentos de doble índice supera el número de sitios de amplificación, en donde
- 10 las moléculas adaptadoras de fragmentos de doble índice tienen acceso fluido a los sitios de amplificación, y en donde cada uno de los sitios de amplificación comprende una capacidad para varias moléculas adaptadoras de fragmentos de doble índice en la biblioteca de secuenciación, o en donde la puesta en contacto comprende simultáneamente (i) transportar las moléculas adaptadoras de fragmentos de doble índice a los sitios de amplificación a una velocidad de transporte promedio, y (ii) amplificar las moléculas adaptadoras de fragmentos de doble índice que están en los sitios de amplificación a una velocidad de amplificación promedio, en donde la velocidad de amplificación promedio supera la velocidad de transporte promedio.
- 15
16. El método de la reivindicación 15 como dependiente de la reivindicación 3, que comprende además:
- 20 proporcionar una superficie que comprenda una pluralidad de sitios de amplificación, en donde los sitios de amplificación comprenden al menos dos poblaciones de ácidos nucleicos monocatenarios unidos que tienen un extremo 3' libre, y
- 25 poner en contacto la superficie que comprende los sitios de amplificación con la biblioteca de secuenciación en condiciones adecuadas para producir una pluralidad de sitios de amplificación que cada uno comprende una población clonal de amplicones de una molécula adaptadora de fragmentos de doble índice individual, opcionalmente en donde el número de moléculas adaptadoras de fragmentos de doble índice supera el número de sitios de amplificación, en donde las moléculas adaptadoras de fragmentos de doble índice tienen acceso fluido a los sitios de amplificación y en donde cada uno de los sitios de amplificación comprende una capacidad para varias moléculas adaptadoras de fragmentos de doble índice en la biblioteca de secuenciación, o en donde el contacto comprende simultáneamente (i) transportar las moléculas adaptadoras de fragmentos de doble índice a los sitios de amplificación a una velocidad de transporte promedio y (ii) amplificar las moléculas adaptadoras de fragmentos de doble índice que se encuentran en los sitios de amplificación a una velocidad de amplificación promedio, en donde la velocidad de amplificación promedio supera la velocidad de transporte promedio.
- 30
- 35
17. Una composición que comprende subconjuntos de núcleos que comprenden fragmentos de ácido nucleico que comprenden una primera secuencia índice y tratados para identificar nucleótidos metilados, en donde la primera secuencia índice es diferente para cada uno de los subconjuntos de núcleos.
- 40
18. La composición de la reivindicación 17, en donde los fragmentos de ácido nucleico comprenden una modificación química de secuencias CpG dependiente del estado de metilación.
- 45
19. La composición de la reivindicación 18, en donde los fragmentos de ácido nucleico comprenden además una secuencia universal y/o una primera secuencia cebadora de secuenciación.
20. La composición de una cualquiera de las reivindicaciones 17-19, en donde la primera secuencia índice, la secuencia universal o la primera secuencia cebadora de secuenciación está empobrecida en citosina.
- 50
21. La composición de la reivindicación 17, que comprende además complejos transposómicos.
22. La biblioteca de secuenciación producida por el método de la reivindicación 1 o la reivindicación 3.

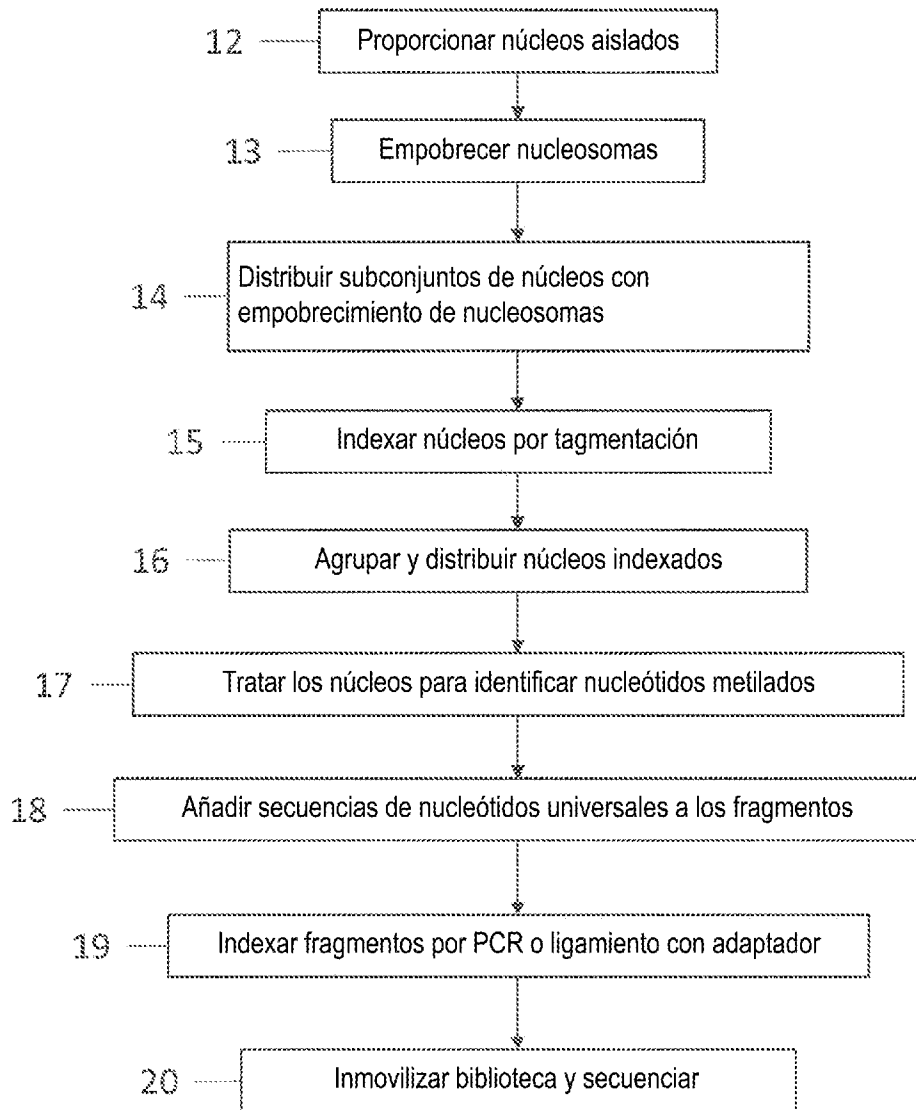


Figura 1

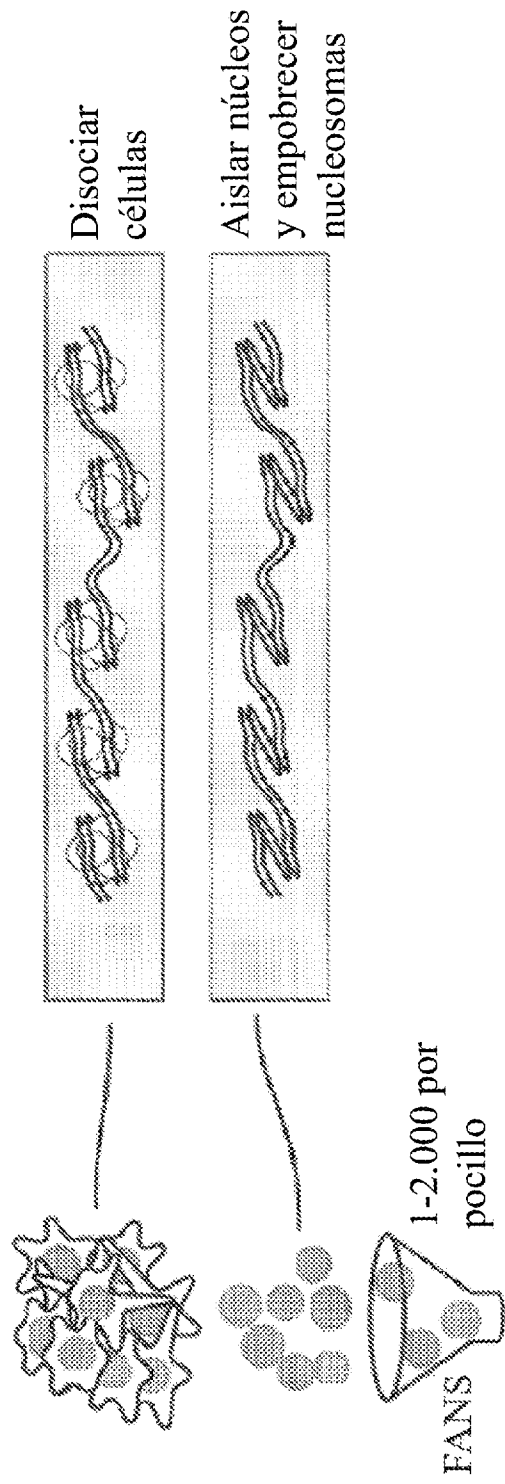


Figura 2A

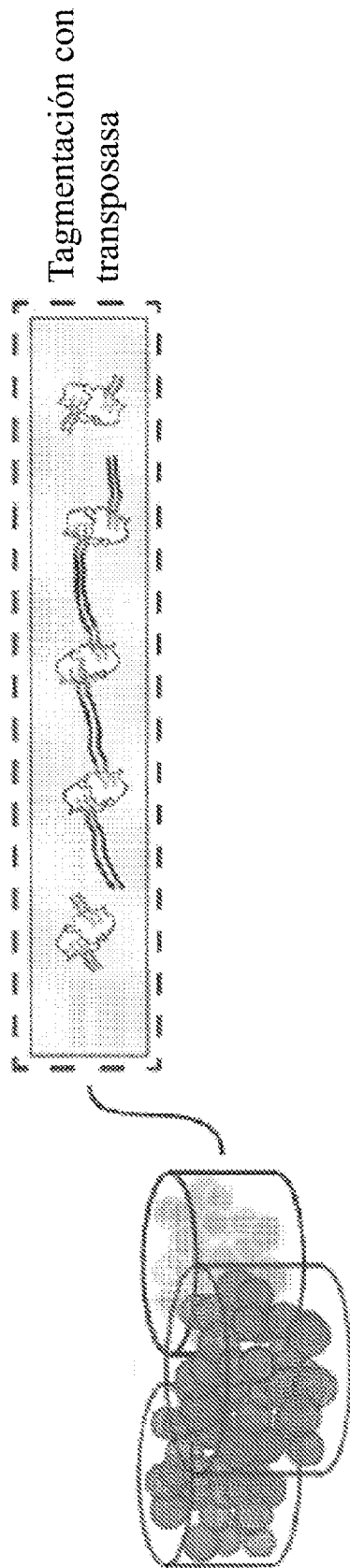


Figura 2B

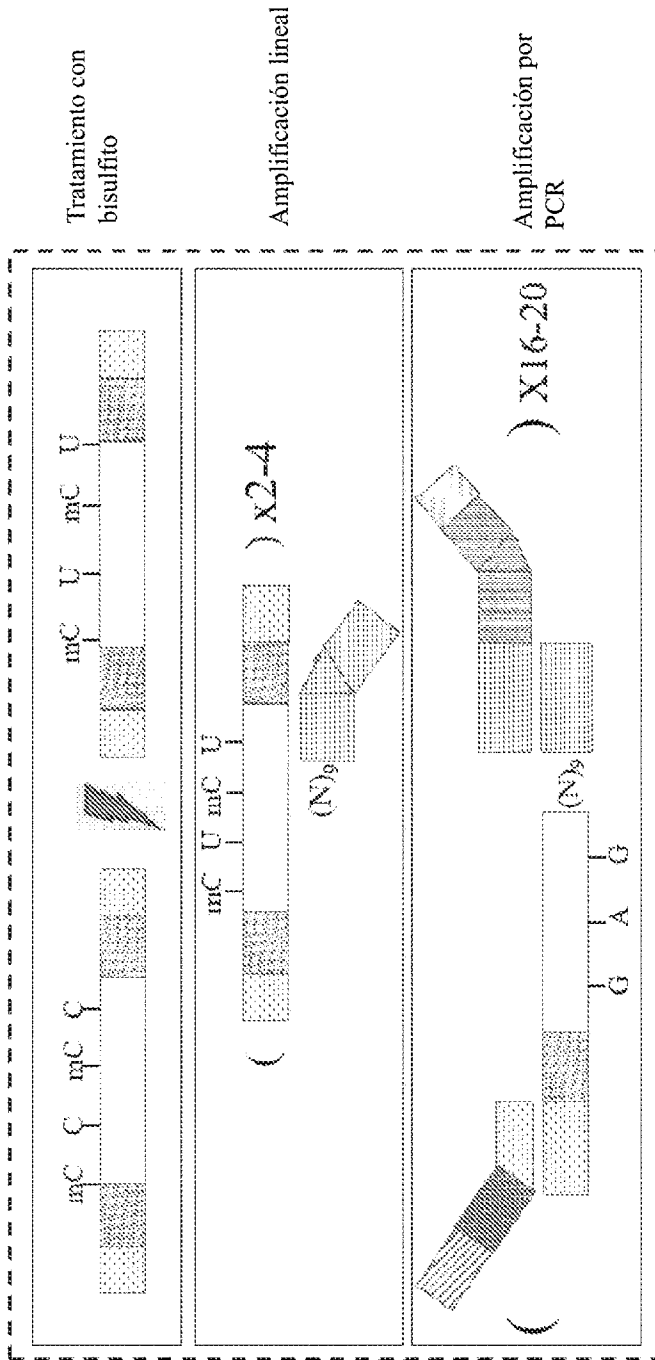


Figura 2C

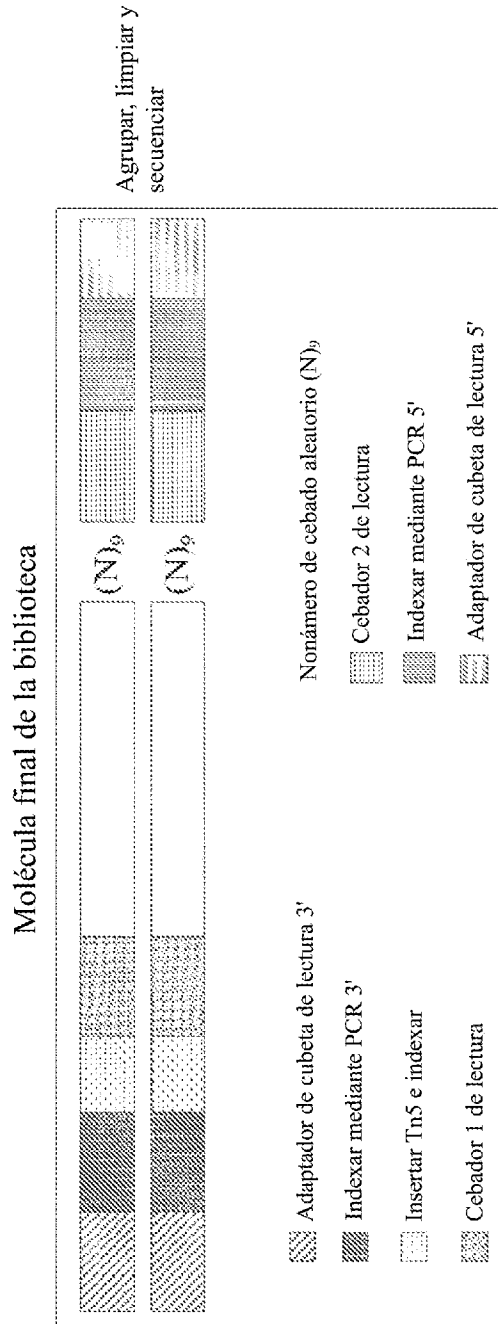


Figura 2D

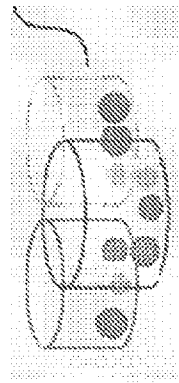


Figura 3

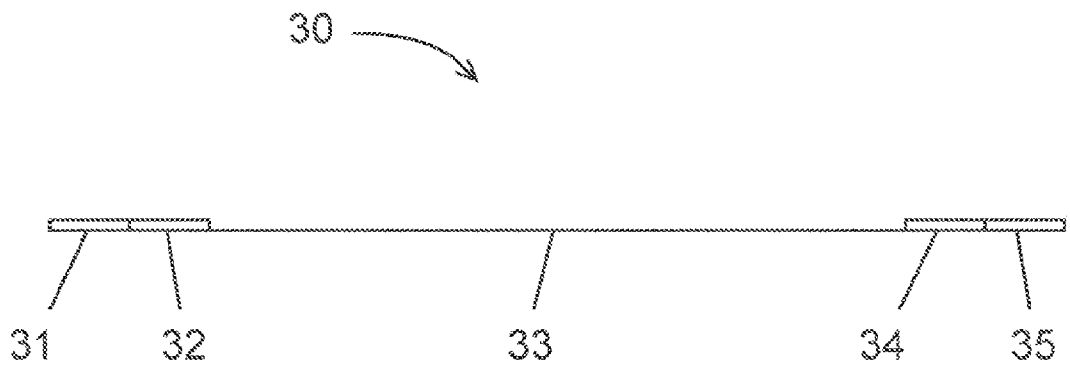


Figura 4

