

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 786 652**

51 Int. Cl.:

**C12Q 1/6806** (2008.01)

**C12Q 1/6869** (2008.01)

**C12N 15/10** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **10.02.2016 PCT/US2016/017391**

87 Fecha y número de publicación internacional: **18.08.2016 WO16130704**

96 Fecha de presentación y número de la solicitud europea: **10.02.2016 E 16711039 (4)**

97 Fecha y número de publicación de la concesión europea: **25.03.2020 EP 3256604**

54 Título: **Métodos y composiciones para analizar componentes celulares**

30 Prioridad:

**10.02.2015 US 201562114505 P**

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

**13.10.2020**

73 Titular/es:

**ILLUMINA, INC. (100.0%)  
5200 Illumina Way  
San Diego, CA 92122, US**

72 Inventor/es:

**GUNDERSON, KEVIN L.;  
STEEMERS, FRANK J.;  
FISHER, JEFFREY S. y  
RIGATTI, ROBERTO**

74 Agente/Representante:

**ELZABURU, S.L.P**

ES 2 786 652 T3

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

## DESCRIPCIÓN

Métodos y composiciones para analizar componentes celulares

**Campo de la divulgación**

5 Realizaciones de la presente solicitud se refieren a métodos y a una composición para analizar componentes celulares. En algunas realizaciones, la presente solicitud se refiere a métodos y a una composición para analizar componentes de una célula individual. En algunas realizaciones, la presente solicitud se refiere a métodos y a una composición para identificar un tipo de célula individual. En algunas realizaciones, los métodos y las composiciones se refieren a la secuenciación de ácidos nucleicos. Algunas realizaciones de los métodos y las composiciones proporcionados son útiles para derivar un estado compuesto de dicha célula individual.

**10 Antecedentes**

La detección de secuencias específicas de ácidos nucleicos presentes en una muestra biológica se ha utilizado, por ejemplo, como un método para identificar y clasificar microorganismos, diagnosticar enfermedades infecciosas, detectar y caracterizar anomalías genéticas, identificar cambios genéticos asociados con el cáncer, estudiar la susceptibilidad genética a la enfermedad y medir la respuesta a diversos tipos de tratamiento. Una técnica común para detectar secuencias específicas de ácidos nucleicos en una muestra biológica es la secuenciación de ácidos nucleicos.

15 La metodología de secuenciación de ácidos nucleicos ha evolucionado significativamente a partir de los métodos de degradación química utilizados por Maxam y Gilbert y los métodos de alargamiento de cadena utilizados por Sanger. Hoy en día, se utilizan varias metodologías de secuenciación que permiten el procesamiento paralelo de ácidos nucleicos, todo en una sola secuencia. Como tal, la información generada a partir de una sola secuencia puede ser enorme.

Amini y col., *Nat. Genet.* 2014 46(12): 1343-1349, describe un método de secuenciación del genoma resuelto por haplotipo mediante contigüidad preservando la transposición y la indexación combinatoria.

**Breve descripción de los dibujos**

25 La **Fig. 1** representa un esquema de una indexación combinatoria de cuatro niveles del elemento de preservación de contigüidad (CE) de ADN creado al incrustar contenidos de una célula individual en una matriz polimérica o unirlo a una perla. Los índices específicos para el compartimento se adjuntan en cada combinación combinatoria y paso de redistribución (niveles). En el ejemplo que se muestra, los cuatro niveles dan como resultado que cuatro índices se concatenan juntos (a través de rondas repetidas de ligadura, extensión de polimerasa, etiquetado, etc.) permitiendo una lectura de secuenciación fácil. Alternativamente, el elemento de preservación de contigüidad que comprende ADN puede crearse mediante una partición de ADN compartimentada (es decir, una dilución de ADN que submuestra la muestra de ADN original) que ha sido encapsulado en una matriz o inmovilizado en una perla. Este tipo de dilución es útil en aplicaciones de ajuste de fases y montaje.

35 La **Fig. 2** representa un método para preparar colecciones de ADN o ADNc de una célula individual utilizando un esquema de indexación combinatoria de dos niveles, en el que los índices de primer nivel se unen mediante etiqueta (índices específicos para el compartimento en transposones) y los índices de segundo nivel se unen por PCR (índices específicos para el compartimento en cebadores de PCR). El contenido del recipiente de células individuales (es decir, ADN genómico o ADNc) puede emplear una amplificación de genoma completo (WGA) opcional o una etapa de amplificación del transcriptoma completo.

40 La **Fig. 3** describe un método para hacer una colección de ADNc a partir del contenido de una célula individual en CE, tales como gotitas. En el ejemplo que se muestra, los índices se utilizan para marcar diferentes muestras.

La **Fig. 4** representa el contenido representativo de una célula individual que puede analizarse mediante el esquema de indexación combinatoria propuesto.

45 Las **Fig. 5A y 5B** representan realizaciones esquemáticas ilustrativas para crear elementos de preservación de contigüidad (CE) a partir de encapsular y lisar el contenido de una célula individual atrapada dentro de una CE, tal como en una perla polimérica. La célula está incrustada, por ejemplo, en un perla polimérica. Todos los componentes de una célula individual se mantienen cerca uno del otro en la perla. Posteriormente, uno o más componentes pueden amplificarse, modificarse (síntesis de ADNc) y posteriormente marcarse con índices o etiquetas. La **Fig. 5C** representa una realización esquemática a modo de ejemplo en la que la indexación de la muestra se puede lograr añadiendo secuencias de ADN codificantes (tales como un plásmido) en la fase de encapsulación, amplificación/ADNc o polimerización. Cada una de las muestras se prepara con un conjunto diferente de plásmidos codificadores o una combinación de plásmidos codificadores. Cada una de las CE indexada combinatoriamente producirá elementos de la colección codificadores de la muestra indexados combinatoriamente correspondientes. De esta manera, cada uno de los elementos de la colección se puede asignar a su CE de origen y muestra de origen.

La **Fig. 6** representa esquemas para encapsular contenidos de células individuales en CE, tales como perlas de matriz polimérica.

La **Fig. 7** representa un esquema ilustrativo de análisis de alto rendimiento de componentes celulares mediante captura directa de superficie. "A" muestra una colección de células. "B" muestra transposomas unidos a la superficie. En "C" las células fluyen sobre la superficie. En las células "D" se lisan y se permite que los componentes de la célula se difundan de forma controlada alrededor del sitio en el que se capturó la célula. En "E" los ácidos nucleicos son capturados (etiquetados) por los transposomas. Se capturan diferentes componentes celulares dependiendo de si la membrana celular o los núcleos se lisan. Mediante el uso de restos de captura específicos para los componentes (es decir, anticuerpos, receptores, ligandos), se pueden capturar diversos componentes celulares. El análisis de las moléculas capturadas puede llevarse a cabo directamente en la superficie de captura. Alternativamente, las moléculas capturadas pueden cosecharse y analizarse en una superficie diferente. En este caso, la primera superficie está compuesta de múltiples áreas (es decir, almohadillas) y cada una de las almohadillas está recubierta con oligos que comparten un código de barras idéntico para que las moléculas que se capturan en la misma almohadilla compartan el mismo código de barras de identificación.

La **Fig. 8** representa un esquema ilustrativo de análisis de ácidos nucleicos usando elementos de preservación de contigüidad en perlas.

La **Fig. 9A-D** representa una estrategia de modelado ilustrativa.

La **Fig. 10** muestra un método para crear partículas que son útiles para crear elementos de contigüidad.

#### Descripción detallada

La invención se define por las reivindicaciones adjuntas.

Algún aspecto de la presente descripción se refiere a métodos y composiciones relacionados con la evaluación de componentes de una célula individual preservada o incrustada o contenida dentro de elementos de preservación de contigüidad (CE).

En un aspecto dado a conocer en el presente documento, se describen métodos para analizar una pluralidad de tipos de analitos de una célula individual. En algunas realizaciones, se proporciona una pluralidad de elementos de preservación de contigüidad (CE), cada CE comprende una célula individual. Las células se lisan dentro de la CE de modo que la pluralidad de analitos dentro de la célula individual se libera dentro de la CE. En algunas realizaciones, se proporciona una pluralidad de tipos de restos informadores de manera que cada tipo de resto informador es específico para cada tipo de analito. En algunas realizaciones, el resto informador identifica una célula individual. La pluralidad de analitos se modifica de modo que cada tipo de analito comprenda un resto informador específico para el tipo de analito. En algunas realizaciones, se combinan el CE que comprende los analitos que comprenden dichos restos informadores. En algunas realizaciones, la CE combinada que comprende los analitos que comprenden dichos restos informadores está compartimentada. En algunas realizaciones, se proporcionan restos informadores adicionales y se combinan con los analitos que comprenden analitos, de modo que los analitos comprenden dos o más restos informadores diferentes. Los analitos que comprenden los restos informadores se analizan de manera que se detecta la identidad del analito y el resto informador identifica la fuente del analito a partir de una célula individual.

En algunas realizaciones, la pluralidad ilustrativa de analitos incluye, pero no se limita a ADN, ARN, ADNc, proteínas, lípidos, carbohidratos, organelas celulares (por ejemplo, núcleo, aparato de Golgi, ribosomas, mitocondrias, retículo endoplasmático, cloroplasto, membrana celular, etc.), metabolitos celulares, secciones de tejido, células, célula individual, contenido de células o de una célula individual, ácido nucleico aislado de células o de una célula individual, o ácido nucleico aislado de células o de una célula individual y modificado adicionalmente, o ADN libre de células (por ejemplo, de líquido placentario o plasma). En algunas realizaciones, la pluralidad de analitos incluye ADN genómico y ARNm. En algunas realizaciones, el ARNm tiene cola poli A. En algunas realizaciones, el ADN genómico y el ARNm se inmovilizan en un soporte sólido dentro del CE simultáneamente. En algunas realizaciones, la inmovilización del ADN genómico es secuencial a la inmovilización del ARNm en el soporte sólido. En algunas realizaciones, el ADN genómico se combina con complejos de transposomas y los extremos del transposón se inmovilizan en un soporte sólido y el ARNm se inmoviliza en el sólido por hibridación de sondas oligo (dT) inmovilizadas en un soporte sólido. En algunas realizaciones, el ADN genómico se combina con complejos de transposomas y, opcionalmente, los extremos del transposón se hibridan con secuencias complementarias inmovilizadas en un soporte sólido de modo que el ARNm se inmoviliza en el sólido por hibridación de sondas oligo (dT) inmovilizadas en un soporte sólido. También se pueden usar otros métodos para inmovilizar el ARNm. En algunas realizaciones, el soporte sólido es una perla. En algunas realizaciones, el soporte sólido es una superficie de celda de flujo. En algunas realizaciones, la superficie sólida es la pared de un recipiente de reacción.

En algunas realizaciones, los métodos incluyen secuenciación de ácidos nucleicos conservados o incrustados o contenidos dentro de CE. En particular, realizaciones de los métodos y composiciones proporcionados en el presente documento se refieren a la preparación de moldes de ácido nucleico y a la obtención de datos de secuencia a partir de los mismos. Los métodos y las composiciones proporcionados en este documento están

relacionados con los métodos y composiciones proporcionados en la publicación de la solicitud de Patente de EE.UU. No. 2012/0208705, la publicación de la solicitud de Patente de EE.UU. No. 2012/0208724 y la publicación de la solicitud de patente No. WO 2012/061832. Algunas realizaciones se refieren a la preparación de ADN dentro de CE para obtener información de ajuste de fases y ensamblaje de secuencia de un ácido nucleico diana, y obtener información de secuencia de ensamblaje de secuencia y secuencia de dichos moldes. Realizaciones particulares proporcionadas en el presente documento se refieren al uso de integrasas por ejemplo, transposasas, para mantener la proximidad física de extremos asociados de ácidos nucleicos fragmentados; y al uso de indexación combinatoria para crear colecciones individuales de cada una de las CE. Obtener información de haplotipo de CE incluye distinguir entre diferentes alelos (*p.ej.*, SNP, anomalías genéticas, etc.) en un ácido nucleico diana. Métodos de este tipo son útiles para caracterizar diferentes alelos en un ácido nucleico diana y para reducir la tasa de error en la información de secuencia.

En una realización, un ácido nucleico molde puede diluirse en CE tal como gotitas. Se puede emplear la amplificación opcional del genoma completo, y se puede obtener información de la secuencia a partir de una cantidad de ácido nucleico molde equivalente a aproximadamente un equivalente haploide del ácido nucleico diana.

En realizaciones adicionales, un ácido nucleico molde puede compartimentarse de manera tal que múltiples copias de un cromosoma pueden estar presentes en el mismo compartimento, como resultado de la indexación dual o múltiple proporcionada aquí, también se puede determinar un haplotipo. En otras palabras, se puede preparar un ácido nucleico molde usando compartimentos virtuales. En estas realizaciones, un ácido nucleico puede distribuirse entre varios primeros compartimentos, proporcionando un primer índice al ácido nucleico de cada uno de los compartimentos, combinando los ácidos nucleicos, distribuyendo el ácido nucleico entre varios segundos compartimentos y proporcionando un segundo índice al ácido nucleico de cada uno de los compartimentos. Ventajosamente, dicha indexación permite obtener información del haplotipo a concentraciones más altas de ácido nucleico en comparación con la mera dilución de un ácido nucleico en un único compartimento a una cantidad equivalente a un haplotipo del ácido nucleico.

Como se usa en este documento, el término "compartimento" pretende significar un área o volumen que separa o aísla algo de otras cosas. Compartimentos ilustrativos incluyen, pero no se limitan a viales, tubos, pocillos, gotitas, bolos, perlas, recipientes, características de superficie o áreas o volúmenes separados por fuerzas físicas tales como flujo de fluido, magnetismo, corriente eléctrica o similares.

Un método ilustrativo para hacer compartimentos se muestra en la Fig. 10. Una placa maestra de silicio que tiene postes puede usarse para imprimir pocillos en una lámina de hidrogel (los pocillos en el hidrogel son las imágenes inversas de los postes). Los pocillos resultantes en el hidrogel se pueden llenar con un material que forma partículas (por ejemplo, un gel o polímero) junto con un analito diana u otro reactivo. La lámina de hidrogel se puede disolver mediante una técnica que no disuelve las partículas. Luego, las partículas pueden recogerse y manipularse usando los métodos establecidos aquí.

En algunas realizaciones proporcionadas en el presente documento, las colecciones de moldes se preparan usando transposomas. En algunas de estas colecciones, el ácido nucleico diana puede estar fragmentado. Por consiguiente, algunas realizaciones proporcionadas en el presente documento se refieren a métodos para mantener información de la secuencia para la contigüidad física de fragmentos adyacentes. Métodos de este tipo incluyen el uso de integrasas para mantener la asociación de fragmentos de ácido nucleico molde adyacentes en el ácido nucleico diana. Ventajosamente, dicho uso de integrasas para mantener la proximidad física de los ácidos nucleicos fragmentados aumenta la probabilidad de que los ácidos nucleicos fragmentados de la misma molécula original, por ejemplo, el cromosoma, se presenten en el mismo compartimento.

Otras realizaciones proporcionadas en el presente documento se refieren a la obtención de información de la secuencia de cada una de las cadenas de un ácido nucleico que puede ser útil para reducir la tasa de error en la información de secuenciación. Métodos para preparar colecciones de ácidos nucleicos molde para obtener información de la secuencia de cada una de las cadenas de un ácido nucleico se pueden preparar de modo que cada una de las cadenas se pueda distinguir, y los productos de cada una de las cadenas también se puedan distinguir.

Algunos de los métodos proporcionados en el presente documento incluyen métodos de análisis de ácidos nucleicos. Métodos de este tipo incluyen preparar una colección de ácidos nucleicos molde de un ácido nucleico diana, obtener datos de la secuencia de la colección de ácidos nucleicos molde y ensamblar una representación de secuencia del ácido nucleico diana a partir de dichos datos de la secuencia.

Generalmente, los métodos y las composiciones proporcionados en este documento están relacionados con los métodos y las composiciones proporcionados en la publicación de la solicitud de Patente de EE.UU. No. 2012/0208705, la publicación de la solicitud de Patente de EE.UU. No. 2012/0208724 y la publicación de la solicitud de Patente Internacional No. WO 2012/061832.

Los métodos proporcionados en el presente documento se refieren al uso de transposomas útiles para insertar características en un ácido nucleico diana. Dichas características incluyen sitios de fragmentación, sitios de

cebadores, códigos de barras, etiquetas de afinidad, restos informadores, etc.

En un método útil con las realizaciones proporcionadas en el presente documento, se prepara una colección de ácidos nucleicos molde a partir de un CE que comprende ácido nucleico diana. La colección se prepara insertando o fijando una pluralidad de códigos de barras únicos en todo el ácido nucleico diana. En algunas realizaciones, cada uno de los códigos de barras incluye una primera secuencia de código de barras y una segunda secuencia de código de barras, que tiene un sitio de fragmentación dispuesto entre ellas. La primera secuencia de código de barras y la segunda secuencia de código de barras pueden identificarse o designarse para ser emparejadas entre sí. El emparejamiento puede ser informativo para que un primer código de barras se asocie con un segundo código de barras. Ventajosamente, las secuencias de códigos de barras emparejadas pueden usarse para ensamblar datos de secuenciación de la colección de ácidos nucleicos moldes. Por ejemplo, la identificación de un primer ácido nucleico molde que comprende una primera secuencia de código de barras y un segundo ácido nucleico molde que comprende una segunda secuencia de código de barras que está emparejada con la primera indica que el primer y el segundo ácido nucleico moldes representan secuencias adyacentes entre sí en una representación de secuencia del ácido nucleico diana. Métodos de este tipo se pueden usar para ensamblar una representación de secuencia de un ácido nucleico diana *de novo*, sin el requisito de un genoma de referencia.

En algunas realizaciones, se puede emplear un código de barras combinatorio múltiple de modo que el ácido nucleico diana de cada una de las células individuales comprenda un código de barras único (por ejemplo, una combinación única de códigos de barras) y se pueda identificar fácilmente a partir de un ácido nucleico diana diferente de una célula individual diferente. En algunas realizaciones, un CE puede comprender el ácido nucleico diana de una célula individual. En algunas realizaciones, el ácido nucleico diana dentro de un CE tendrá códigos de barras únicos identificables que son diferentes del ácido nucleico diana dentro de un CE diferente.

En algunas realizaciones, se puede emplear un esquema de marcaje combinatorio múltiple para los componentes dentro de una célula individual, además del ácido nucleico, por ejemplo, proteínas, organelas, lípidos o membranas celulares, de modo que los componentes dentro de una célula individuales puedan identificar a partir de componentes de una célula individual diferente. En algunas realizaciones, un CE puede comprender los componentes dentro de una célula individual. En algunas realizaciones, los componentes de una célula individual dentro de un CE tendrán etiqueta(s) identificable(s) única(s) que son diferentes de los componentes de una célula individual dentro de un CE diferente.

En algunas realizaciones, se pueden emplear múltiples esquemas de código de barras combinatorios para el ácido nucleico diana de una célula individual y se pueden emplear múltiples esquemas de marcaje combinatorio para los componentes dentro de una célula individual juntos. En algunas realizaciones, dicho código de barras combinatorio y combinatorio se pueden realizar dentro de un CE que comprende una célula individual. En algunas realizaciones, dicho código de barras combinatorio y marcaje combinatorio pueden realizarse para CE múltiples que comprenden células individuales en paralelo.

En algunas realizaciones, las proteínas conservadas, incrustadas, inmovilizadas o contenidas dentro de CE pueden secuenciarse. En algunas realizaciones, tales proteínas están marcadas de forma única. En algunas realizaciones, las proteínas conservadas, incrustadas, inmovilizadas o contenidas dentro de CE pueden identificarse mediante métodos conocidos en la técnica. En algunas realizaciones, la identificación y/o secuenciación de la proteína se puede llevar a cabo junto con la recopilación de información de la secuencia de los ácidos nucleicos.

Como se usa en el presente documento, la expresión "ácido nucleico" y/o el término "oligonucleótido" y/o equivalentes gramaticales de los mismos pueden referirse a al menos dos monómeros de nucleótidos unidos entre sí. Un ácido nucleico puede contener generalmente enlaces fosfodiéster; sin embargo, en algunas realizaciones, los análogos de ácido nucleico pueden tener otros tipos de cadenas principales, que comprenden, por ejemplo, fosforamida (Beaucage, *et al.*, *Tetrahedron*, 49:1925 (1993); Letsinger, *J. Org. Chem.*, 35:3800 (1970); Sprinzl, *et al.*, *Eur. J. Biochem.*, 81:579 (1977); Letsinger *et al.*, *Nucl. Acids Res.*, 14:3487 (1986); Sawai *et al.*, *Chem. Lett.*, 805 (1984); Letsinger *et al.*, *J. Am. Chem. Soc.*, 110:4470 (1988); y Pauwels, *et al.*, *Chemica Scripta*, 26: 141 (1986)), fosforotioato (Mag, *et al.*, *Nucleic Acids Res.*, 19:1437 (1991); y Pat. de EE.UU. N° 5.644.048), fosforoditioato (Briu *et al.*, *J. Am. Chem. Soc.*, 111:2321 (1989), enlaces de O-metilfosforoamidita (véase Eckstein, *Oligonucleotides and Analogues: A Practical Approach*, Oxford University Press), y cadenas principales y enlaces de ácido nucleico peptídico (véase Egholm, *J. Am. Chem. Soc.*, 114:1895 (1992); Meier *et al.*, *Chem. Int. Ed. Engl.*, 31:1008 (1992); Nielsen, *Nature*, 365: 566 (1993); Carlsson, *et al.*, *Nature*, 380: 207 (1996)).

Otros ácidos nucleicos análogos incluyen aquellos con cadenas principales positivas (Denpcy, *et al.*, *Proc. Natl. Acad. Sci. USA*, 92:6097 (1995)); cadenas principales no iónicas (patentes de EE.UU. N°s. 5.386.023; 5.637.684; 5.602.240; 5.216.141; y 4.469.863; Kiedrowski, *et al.*, *Angew. Chem Intl. Ed. Inglés*, 30:423 (1991); Letsinger *et al.*, *J. Am. Chem. Soc.*, 110:4470 (1988); Letsinger, *et al.*, *Nucleosides & Nucleotides*, 13:1597 (1994); Capítulos 2 y 3, *ASC Symposium Series 580*, "Carbohydrate Modifications in Antisense Research", Ed. Y. S. Sanghui y P. Dan Cook; Mesmaeker, *et al.*, *Bioorganic & Medicinal Chem. Lett.*, 4:395 (1994); Jeffs, *et al.*, *J. Biomolecular NMR*, 34:17 (1994); *Tetrahedron Lett.*, 37:743 (1996)) y no ribosa (Patente de EE.UU. N° 5.235.033 y N° 5.034.506 y Capítulos 6 y 7, *ASC Symposium Series 580*, "Carbohydrate Modifications in Antisense Research", Ed. Y. S. Sanghui y P. Dan Cook). Los ácidos nucleicos también pueden contener uno o más azúcares carbocíclicos (véase Jenkins *et al.*, *Chem.*

Soc. Rev., (1995) págs. 169 176).

Se pueden hacer modificaciones de la cadena principal de ribosa-fosfato para facilitar la adición de restos adicionales tales como marcadores, o para aumentar la estabilidad de estas moléculas bajo ciertas condiciones. Además, se pueden hacer mezclas de ácidos nucleicos y análogos que se producen de forma natural. Alternativamente, se pueden hacer mezclas de diferentes análogos de ácidos nucleicos y mezclas de ácidos nucleicos y análogos que se producen de forma natural. Los ácidos nucleicos pueden ser monocatenarios o bicatenarios, según se especifique, o pueden contener porciones de secuencias tanto bicatenarias como monocatenarias. El ácido nucleico puede ser ADN, por ejemplo, genómico o ADNc, ARN o un híbrido, de células individuales, de múltiples células o de múltiples especies, tal como con muestras metagenómicas, tales como muestras ambientales, además de muestras mixtas, por ejemplo, muestras de tejido mixto o muestras mixtas para diferentes individuos de la misma especie, muestras de enfermedades, tales como ácidos nucleicos relacionados con el cáncer y similares. Un ácido nucleico puede contener cualquier combinación de desoxirribo- y ribonucleótidos, y cualquier combinación de bases, incluidos uracilo, adenina, timina, citosina, guanina, inosina, xantantina, hipoxantantina, isocitosina, isoguanina y análogos de bases tales como nitropirrol (incluido 3-nitropirrol) y nitroindol (incluido 5-nitroindol), etc.

En algunas realizaciones, un ácido nucleico puede incluir al menos una base promiscua. Bases promiscuas pueden aparearse con más de un tipo diferente de base. En algunas realizaciones, una base promiscua puede aparearse con al menos dos tipos diferentes de bases y no más de tres tipos diferentes de bases. Un ejemplo de una base promiscua incluye inosina que puede aparearse con adenina, timina o citosina. Otros ejemplos incluyen hipoxantina, 5-nitroindol, 5-nitroindol acílico, 4-nitropirazol, 4-nitroimidazol y 3-nitropirrol (Loakes *et al.*, *Nucleic Acid Res.* 22:4039 (1994); Van Aerschot *et al.*, *Nucleic Acid Res.* 23: 4363 (1995); Nichols *et al.*, *Nature* 369:492 (1994); Bergstrom *et al.*, *Nucleic Acid Res.* 25: 1935 (1997); Loakes *et al.*, *Nucleic Acid Res.* 23:2361 (1995); Loakes *et al.*, *J. Mol. Biol.* 270:426 (1997); y Fotin *et al.*, *Nucleic Acid Res.* 26:1515 (1998)). También se pueden usar bases promiscuas que pueden aparearse con al menos tres, cuatro o más tipos de bases.

Como se usa en el presente documento, la expresión "análogo de nucleótido" y/o equivalentes gramaticales del mismo pueden referirse a análogos sintéticos que tienen porciones de base de nucleótidos modificadas, porciones de pentosa modificadas y/o porciones de fosfato modificadas y, en el caso de polinucleótidos, enlaces internucleotídicos modificados, tal como se describe generalmente en otra parte (p.ej., Scheit, *Nucleotide Analogs*, John Wiley, Nueva York, 1980; Englisch, *Angew. Chem Int. Ed. Engl.* 30: 613-29 1991; Agarwal, *Protocols for Polynucleotides and Analogs*, Humana Press, 1994; y S. Verma y F. Eckstein, *Ann. Rev. Biochem.* 67:99-134, 1998) En general, las porciones de fosfato modificadas comprenden análogos de fosfato en los que el átomo de fósforo está en el estado de oxidación +5 y uno o más de los átomos de oxígeno se reemplazan por un resto sin oxígeno, p.ej., azufre. Análogos de fosfato ilustrativos incluyen, pero no se limitan a, fosforotioato, fosforoditioato, fosforoselenoato, fosforodiselenoato, fosforoanilotoato, fosforanilidato, fosforamidato, boronofosfatos, incluidos los contraiones asociados, p. ej.,  $H^+$ ,  $NH_4^+$ ,  $Na^+$ , si este tipo de contraiones están presentes. Las porciones de base de nucleótidos modificadas ilustrativas incluyen, pero sin limitación, 5-metilcitosina (5 mC); análogos de C-5-propinilo, que incluyen pero no se limitan a, C-5 propinil-C y C-5 propinil-U; 2,6-diaminopurina, también conocida como 2-amino adenina o 2-amino-dA); hipoxantina, pseudouridina, 2-tiopirimidina, isocitosina (isoC), 5-metil isoC e isoguanina (isoG; véase, por ejemplo, la patente de EE.UU. N<sup>o</sup>. 5.432.272). Porciones de pentosa modificadas ilustrativas incluyen, pero no se limitan a análogos de ácido nucleico bloqueado (LNA) que incluyen, sin limitación, Bz-A-LNA, 5-Me-Bz-C-LNA, dmf-G-LNA y T-LNA (véase, por ejemplo, *he Glen Report*, 16(2):5, 2003; Koshkin *et al.*, *Tetrahedron* 54:3607-30, 1998), y modificaciones 2' o 3' donde la posición 2' o 3' es hidrógeno, hidroxilo, alcoxi (p.ej., metoxi, etoxi, aliloxi, isopropoxi, butoxi, isobutoxi y fenoxi), azido, amino, alquilamino, fluoro, cloro o bromo. Enlaces internucleotídicos modificados incluyen análogos de fosfato, análogos que tienen enlaces entre subunidades aquirales y no cargados (p. ej., Sterchak, E. P. *et al.*, *Organic Chem.*, 52: 4202, 1987) y polímeros a base de morfolino sin carga que tienen enlaces entre subunidades aquirales (véase, p. ej., la patente de EE.UU. N<sup>o</sup> 5.034.506). Algunos análogos de enlaces internucleotídicos incluyen heterociclos unidos por morfolidato, acetal y poliamida. En una clase de análogos de nucleótidos, conocidos como ácidos nucleicos peptídicos, incluidos los ácidos nucleicos peptídicos pseudocomplementarios ("PNA"), un enlace convencional de azúcar e internucleótido se ha reemplazado con un polímero de cadena principal de 2-aminoetilglicina amida (véase, p. ej., Nielsen *et al.*, *Science*, 254:1497-1500, 1991; Egholm *et al.*, *J. Am. Chem Soc.*, 114:1895-1897 1992; Demidov *et al.*, *Proc. Natl. Acad. Sci.* 99:5953-58 2002; *Peptide Nucleic Acids Protocols and Applications*, Nielsen, ed., Horizon Bioscience, 2004).

Como se usa en el presente documento, la expresión "lectura de secuenciación" y/o equivalentes gramaticales de la misma puede referirse a un proceso repetitivo de etapas físicas o químicas que se lleva a cabo para obtener señales indicativas del orden de los monómeros en un polímero. Las señales pueden ser indicativas de un orden de monómeros en resolución de monómero individual o resolución más baja. En realizaciones particulares, las etapas pueden iniciarse en una diana de ácido nucleico y pueden llevarse a cabo para obtener señales indicativas del orden de bases en la diana de ácido nucleico. El proceso puede llevarse a cabo hasta su finalización típica, que habitualmente se define por el punto en el que señales del proceso ya no pueden distinguir las bases de la diana con un nivel razonable de certeza. Si se desea, la finalización puede producirse antes, por ejemplo, una vez que se ha obtenido la cantidad deseada de información de la secuencia. Una lectura de secuenciación puede llevarse a cabo en una molécula de ácido nucleico diana individual o simultáneamente en una población de moléculas de ácido

nucleico diana que tienen la misma secuencia, o simultáneamente en una población de ácidos nucleicos diana que tienen diferentes secuencias. En algunas realizaciones, una lectura de secuenciación finaliza cuando ya no se obtienen señales de una o más moléculas de ácido nucleico diana a partir de las cuales se inició la adquisición de la señal. Por ejemplo, se puede iniciar una lectura de secuenciación para una o más moléculas de ácido nucleico diana que están presentes en un sustrato en fase sólida y se puede terminar después de la separación de una o más moléculas de ácido nucleico diana del sustrato. La secuenciación puede terminarse dejando de detectar los ácidos nucleicos diana que estaban presentes en el sustrato cuando se inició la secuenciación.

Como se usa en el presente documento, la expresión "representación de secuenciación" y/o equivalentes gramaticales de la misma puede referirse a información que significa el orden y el tipo de unidades monoméricas en el polímero. Por ejemplo, la información puede indicar el orden y el tipo de nucleótidos en un ácido nucleico. La información puede estar en cualquiera de una variedad de formatos que incluyen, por ejemplo, una representación, imagen, medio electrónico, series de símbolos, series de números, series de letras, series de colores, etc. La información puede estar en una resolución de un monómero individual o a menor resolución. Un polímero ilustrativo es un ácido nucleico, tal como ADN o ARN, que tiene unidades de nucleótidos. Una serie de letras "A", "T", "G" y "C" es una representación de secuencia bien conocida para el ADN que se puede correlacionar, a una resolución de un nucleótido individual, con la secuencia real de una molécula de ADN. Otros polímeros ilustrativos son proteínas que tienen unidades de aminoácidos y polisacáridos que tienen unidades de sacáridos.

Como se usa en el presente documento, la expresión "al menos una porción" y/o equivalentes gramaticales de la misma pueden referirse a cualquier fracción de una cantidad total. Por ejemplo, "al menos una porción" puede referirse a al menos aproximadamente 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 99%, 99,9% o 100% de una cantidad total.

Como se usa en el presente documento, el término "detectar" y/o equivalentes gramaticales del mismo puede referirse a identificar la presencia o existencia de un analito, identificar los componentes individuales de un analito, por ejemplo, información de la secuencia y/o cuantificar la cantidad de dicho analito.

#### Sitios de fragmentación

En algunas realizaciones que comprenden transposomas en bucle, el enlazador puede comprender un sitio de fragmentación. Se puede usar un sitio de fragmentación para escindir la asociación física, pero no la asociación informativa entre una primera secuencia de código de barras y una segunda secuencia de código de barras. La escisión puede ser por medios bioquímicos, químicos u otros. En algunas realizaciones, un sitio de fragmentación puede incluir una secuencia de nucleótidos o nucleótidos que puede fragmentarse por diversos medios. Por ejemplo, un sitio de fragmentación puede comprender un sitio de endonucleasa de restricción; al menos un ribonucleótido escindible con una RNasa; análogos de nucleótidos escindibles en presencia de cierto agente químico; un enlace diol escindible por tratamiento con peryodato; un grupo disulfuro escindible con un agente reductor químico; un resto escindible que puede estar sujeto a escisión fotoquímica; y un péptido escindible por una enzima peptidasa u otro medio adecuado. Véase *p. ej.*, la solicitud de patente de EE.UU. N<sup>o</sup> de publicación 2012/0208705, solicitud de patente de EE.UU. N<sup>o</sup> de publicación 2012/0208724 y la solicitud de patente internacional N<sup>o</sup> de publicación WO 2012/061832.

#### Sitios de cebadores

En algunas realizaciones, los restos informadores pueden comprender sitios de cebadores que pueden hibridarse con un cebador. En algunas realizaciones, un resto informador puede incluir al menos un primer sitio cebador útil para amplificación, secuenciación y similares.

En algunas realizaciones, una secuencia de transposón puede incluir un "adaptador de secuenciación" o "sitio de adaptador de secuenciación", es decir, una región que comprende uno o más sitios que pueden hibridarse con un cebador. En algunas realizaciones, una secuencia de transposón puede incluir al menos un primer sitio de cebador útil para la amplificación, secuenciación y similares. En algunas realizaciones que comprenden transposomas en bucle, un enlazador puede incluir un adaptador de la secuenciación. En más realizaciones que comprenden transposomas en bucle, un enlazador comprende al menos un primer sitio de cebador y un segundo sitio de cebador. La orientación de los sitios de cebador en tales realizaciones puede ser tal que un cebador que hibrida con el primer sitio de cebador y un cebador que hibrida con el segundo sitio de cebador estén en la misma orientación, o en diferentes orientaciones.

En algunas realizaciones, un enlazador puede incluir un primer sitio de cebador, un segundo sitio de cebador que tiene un sitio no amplificable dispuesto entre ellos. El sitio no amplificable es útil para bloquear la extensión de una cadena polinucleotídica entre el primer y el segundo sitio de cebador, en donde la cadena polinucleotídica se hibrida con uno de los sitios de cebador. El sitio no amplificable también puede ser útil para prevenir concatámeros. Ejemplos de sitios no amplificables incluyen un análogo de nucleótido, resto químico no nucleótido, aminoácido, péptido y polipéptido. En algunas realizaciones, un sitio no amplificable comprende un análogo de nucleótido que no se aparea significativamente con A, C, G o T. Algunas realizaciones incluyen un enlazador que comprende un primer sitio de cebador, un segundo sitio de cebador que tiene un sitio de fragmentación dispuesto entre ellos. Otras

realizaciones pueden usar un diseño de adaptador bifurcado o en forma de Y útil para la secuencia direccional, como se describe en la patente de EE.UU. N° 7.741.463.

Secuencias ilustrativas de sitios de unión de cebadores incluyen, pero no se limitan a, AATGATACGGCGACCACCGAGATCTACAC (secuencia P5) y CAAGCAGAAGACGGCATACGAGAT (secuencia P7).

## 5 Restos informadores

Como se usa en este documento, la expresión "resto informador" y equivalentes gramaticales pueden referirse a cualquier etiqueta, marcador, índice, código de barras o grupo identificable que permita determinar la composición, identidad y/o la fuente de un analito que se investiga.

10 El experto en la materia apreciará que se pueden usar muchas especies diferentes de restos informadores con los métodos y las composiciones descritos en este documento, ya sea individualmente o en combinación con uno o más restos informadores diferentes. En algunas realizaciones, se puede usar más de un resto informador diferente para analizar simultáneamente más de un analito. En algunas realizaciones, se puede usar una pluralidad de restos informadores diferentes simultáneamente para identificar de forma única una célula individual o componentes de una célula individual.

15 En ciertas realizaciones, un resto informador puede emitir una señal. Ejemplos de una señal incluyen, pero no se limitan a señales fluorescentes, quimioluminiscentes, bioluminiscentes, fosforescentes, radiactivas, calorimétricas, de actividad iónica, electrónicas o electroquimioluminiscentes. Se enumeran ejemplos de restos informadores, por ejemplo, solicitud de patente de EE.UU. N° de publicación 2012/0208705, solicitud de patente de EE.UU. N° de publicación 2012/0208724 y solicitud de patente internacional N° de publicación WO 2012/061832.

20 En algunas realizaciones, el resto informador puede ser un adaptador. En algunas realizaciones de las composiciones y los métodos descritos en el presente documento, una secuencia de transposón puede incluir un resto informador. En algunas realizaciones que comprenden transposomas en bucle, un enlazador o adaptador puede comprender un resto informador.

25 En algunas realizaciones, un resto informador puede no emitir una señal. En algunas realizaciones, un resto informador puede ser un fragmento de ácido nucleico tal como un código de barras, índice molecular único, un plásmido. En algunas realizaciones, un resto informador puede comprender un anticuerpo que se une específicamente a una proteína. En algunas realizaciones, el anticuerpo puede comprender un marcador detectable. En algunas realizaciones, el informador puede incluir un anticuerpo o reactivo de afinidad marcado con una etiqueta de ácido nucleico. La etiqueta de ácido nucleico puede ser detectable, por ejemplo, mediante un ensayo de ligadura de proximidad (PLA) o un ensayo de extensión de proximidad (PEA).

30 En algunas realizaciones se puede usar un conjunto de restos informadores. En algunas realizaciones, el conjunto de restos informadores puede comprender una mezcla de un subconjunto de restos informadores, en el que cada uno de los subconjuntos de restos informadores es específico para un tipo diferente de analito, por ejemplo, proteínas, ácidos nucleicos, lípidos, carbohidratos. En algunas realizaciones, el conjunto de restos informadores puede comprender una mezcla de un subconjunto de restos informadores, en el que cada uno de los subconjuntos de los restos informadores es diferente entre sí, pero son específicos para un mismo tipo de analito.

## Códigos de barras

35 En general, un código de barras puede incluir una o más secuencias de nucleótidos que pueden usarse para identificar uno o más analitos particulares, tales como ácidos nucleicos, proteínas, metabolitos u otros analitos recogidos en el presente documento o conocidos en la técnica. El código de barras puede ser una secuencia artificial, o puede ser una secuencia que se produce de forma natural generada durante la transposición, tal como secuencias de ADN genómico flanqueantes idénticas (códigos g) al final de fragmentos de ADN anteriormente yuxtapuestos. Un código de barras puede comprender al menos aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 o más nucleótidos consecutivos. En algunas realizaciones, un código de barras comprende al menos aproximadamente 10, 20, 30, 40, 50, 60, 70, 80, 90, 100 o más nucleótidos consecutivos. En algunas realizaciones, al menos una parte de los códigos de barras en una población de ácidos nucleicos que comprende códigos de barras es diferente. En algunas realizaciones, al menos aproximadamente 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 99% de los códigos de barras son diferentes. En más de estas realizaciones, todos los códigos de barras son diferentes. La diversidad de diferentes códigos de barras en una población de ácidos nucleicos que comprende códigos de barras se puede generar aleatoriamente o no aleatoriamente.

40 En algunas realizaciones, una secuencia de transposón comprende al menos un código de barras. En algunas realizaciones, estos transposomas comprenden dos secuencias de transposón no contiguas, la primera secuencia de transposón comprende un primer código de barras, y la segunda secuencia de transposón comprende un segundo código de barras. En algunas realizaciones, tal como en los transposomas en bucle, una secuencia de transposón comprende un código de barras que comprende una primera secuencia de código de barras y una segunda secuencia de código de barras. En algunas de las realizaciones anteriores, la primera secuencia de código

de barras puede identificarse o designarse para emparejarse con la segunda secuencia de código de barras. Por ejemplo, se puede saber que una primera secuencia de código de barras conocida se aparea con una segunda secuencia de código de barras conocida usando una tabla de referencia que comprende una pluralidad de secuencias de códigos de barras primera y segunda que se sabe que están apareadas entre sí.

- 5 En otro ejemplo, la primera secuencia de código de barras puede comprender la misma secuencia que la segunda secuencia de código de barras. En otro ejemplo, la primera secuencia de código de barras puede comprender el complemento inverso de la segunda secuencia de código de barras. En algunas realizaciones, la primera secuencia de código de barras y la segunda secuencia de código de barras son diferentes. Las secuencias de códigos de barras primera y segunda pueden comprender un bi-código.
- 10 En algunas realizaciones de composiciones y métodos descritos en este documento, los códigos de barras se usan en la preparación de ácidos nucleicos molde. Como se entenderá, el gran número de códigos de barras disponibles permite que cada una de las moléculas de ácido nucleico molde comprenda una identificación única. La identificación única de cada una de las moléculas en una mezcla de ácidos nucleicos molde puede usarse en varias aplicaciones. Por ejemplo, las moléculas identificadas de forma única se pueden emplear para identificar moléculas
- 15 de ácido nucleico individuales, en muestras que tienen múltiples cromosomas, en genomas, en células, en tipos de células, en estados de enfermedad celular y en especies, por ejemplo, en la secuenciación de haplotipos, en la discriminación de alelos parentales, en la secuenciación metagenómica y en la secuenciación de muestras de un genoma. Secuencias de códigos de barras ilustrativas incluyen, pero no se limitan a TATAGCCT, ATAGAGGC, CCTATCCT, GGCTCTGA, AGGCGAAG, TAATCTTA, CAGGACGT y GTA CTGAC.

## 20 Enlazadores

- Algunas realizaciones que comprenden transposomas en bucle incluyen secuencias de transposones que comprenden una primera secuencia de código de barras y una segunda secuencia de código de barras que tiene un enlazador dispuesto entre ellas. En otras realizaciones, el enlazador puede estar ausente, o puede ser la cadena principal de azúcar-fosfato que conecta un nucleótido con otro. El enlazador puede comprender, por ejemplo, uno o
- 25 más de un nucleótido, un ácido nucleico, un resto químico no nucleótido, un análogo de nucleótido, aminoácido, péptido, polipéptido o proteína. En realizaciones preferidas, un enlazador comprende un ácido nucleico. El enlazador puede comprender al menos aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 o más nucleótidos. En algunas realizaciones, un enlazador *r* puede comprender al menos aproximadamente 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500 o más nucleótidos.
- 30 En algunas realizaciones, un enlazador puede ser amplificable, por ejemplo, por PCR, amplificación de círculo rodante, amplificación de desplazamiento de cadena y similares. En otras realizaciones, un enlazador puede comprender restos no amplificables. Ejemplos de enlazadores no amplificables incluyen enlazadores químicos orgánicos, tales como alquilo, propilo, PEG; bases no naturales, tales como IsoC, isoG; o cualquier grupo que no se amplifique en esquemas de amplificación basados en ADN. Por ejemplo, los transposones que contienen pares de
- 35 isoC, isoG pueden amplificarse con mezclas de dNTP que carecen de isoG e isoC complementarios, asegurando que no se produzca amplificación a través de los transposones insertados.

En algunas realizaciones, el enlazador comprende un ácido nucleico monocatenario. En algunas realizaciones, el enlazador acopla secuencias de transposón en una orientación 5'-3', una orientación 5'-5' o una orientación 3'-3'.

## Etiquetas de afinidad

- 40 En algunas realizaciones, una secuencia de transposón puede incluir una etiqueta de afinidad. En algunas realizaciones que comprenden transposomas en bucle, un enlazador puede comprender una etiqueta de afinidad. Las etiquetas de afinidad pueden ser útiles para una diversidad de aplicaciones, por ejemplo, la separación masiva de ácidos nucleicos diana hibridados con etiquetas de hibridación. Aplicaciones adicionales incluyen, pero no se limitan a, el uso de etiquetas de afinidad para purificar complejos de transposasa/transposón y ADN diana insertado
- 45 con transposón, ARN diana o proteínas diana, por ejemplo. Tal como se usa en el presente documento, la expresión "etiqueta de afinidad" y equivalentes gramaticales pueden referirse a un componente de un complejo multi-componente, en el que los componentes del complejo multi-componente interactúan específicamente o se unen entre sí. Por ejemplo, una etiqueta de afinidad puede incluir biotina o poli-His que pueden unirse a estreptavidina o níquel, respectivamente. Se enumeran otros ejemplos de complejos de etiquetas de afinidad de múltiples
- 50 componentes, por ejemplo, solicitud de patente de EE.UU. N° de publicación 2012/0208705, solicitud de patente de EE.UU. N° de publicación 2012/0208724 y solicitud de patente internacional N° de publicación WO 2012/061832..

## Soporte Sólido

- Un soporte sólido puede ser bidimensional o tridimensional y puede comprender una superficie plana (*p. ej.*, un portaobjetos de vidrio) o puede tener forma. Un soporte sólido puede incluir vidrio (*p. ej.*, vidrio de poros controlados
- 55 (CPG)), cuarzo, plástico (tal como poliestireno (poliestireno de baja reticulación y alta reticulación), policarbonato, polipropileno y poli(metacrilato de metilo)), copolímero acrílico, poliamida, silicio, metal (*p. ej.*, oro derivado de alcanotiolato), celulosa, nailon, látex, dextrano, matriz de gel (*p. ej.*, gel de sílice), poliácroleína o materiales compuestos.

Soportes sólidos tridimensionales adecuados incluyen, por ejemplo, esferas, micropartículas, perlas, nanopartículas, matrices poliméricas tales como agarosa, poli(acrilamida), alginato, membranas, portaobjetos, placas, chips micromecanizados, tubos (*p.ej.*, tubos capilares), micropocillos, dispositivos microfluídicos, canales, filtros, celdas de flujo, estructuras adecuadas para inmovilizar un ácido nucleico, proteínas o células. Un soporte sólido puede incluir conjuntos o matrices planas capaces de tener regiones que incluyen poblaciones de ácidos nucleicos molde o cebadores. Ejemplos incluyen CPG y portaobjetos de poliestireno derivatizados de nucleósidos; portaobjetos magnéticos derivatizados; poliestireno injertado con polietilenglicol y similares.

En algunas realizaciones, el soporte sólido comprende microesferas o perlas. Por "microesferas" o "perlas" o "partículas" o equivalentes gramaticales en el presente documento se entiende pequeñas partículas discretas. Composiciones de perlas adecuadas incluyen, pero no se limitan a, plásticos, materiales cerámicos, vidrio, poliestireno, metilmetacrilato, polímeros acrílicos, materiales paramagnéticos, sol de toria, grafito de carbono, dióxido de titanio, látex o dextranos reticulados tales como Sepharose, celulosa, nailon, micelas reticuladas y teflón, así como puede usarse cualquier otro material descrito en el presente documento para soportes sólidos. La "Microsphere Detection Guide" de Bangs Laboratories, Fishers Ind., es una guía útil. En ciertas realizaciones, las microesferas son microesferas magnéticas o perlas. En algunas realizaciones, las perlas pueden estar codificadas por colores. Por ejemplo, se pueden usar las microesferas MicroPlex® de Luminex, Austin, TX.

Las perlas no necesitan ser esféricas; se pueden utilizar partículas irregulares. Alternativa o adicionalmente, las perlas pueden ser porosas. Los tamaños de las perlas varían desde nanómetros, es decir, 100 nm, hasta milímetros, es decir, 1 mm, prefiriéndose perlas de aproximadamente 0,2 micras a aproximadamente 200 micras, y siendo particularmente preferidas de aproximadamente 0,5 a aproximadamente 5 micras, aunque en algunas realizaciones se pueden utilizar perlas más pequeñas o más grandes. En algunas realizaciones, las perlas pueden ser de aproximadamente 1, 1,5, 2, 2,5, 2,8, 3, 3,5, 4, 4,5, 5, 5,5, 6, 6,5, 7, 7,5, 8, 8,5, 9, 9,5, 10, 10,5, 15 o 20 µm de diámetro.

En algunas realizaciones, las perlas pueden comprender anticuerpos u otras sondas de afinidad (véase *Immobilized Biomolecules in Analysis. A Practical Approach*. Cass T, Ligler F S, eds. Oxford University Press, Nueva York, 1998, págs. 1-14, para protocolos de unión típicos). En algunas realizaciones, los anticuerpos pueden ser monoclonales y en otras realizaciones, los anticuerpos pueden ser policlonales. En algunas realizaciones, los anticuerpos pueden ser específicos para un epítipo de la superficie celular. En algunas realizaciones, los anticuerpos pueden ser específicos para una proteína dentro de la célula.

En algunas realizaciones, el molde de ácido nucleico proporcionado en el presente documento puede unirse a un soporte sólido. Se pueden usar diversos métodos bien conocidos en la técnica para unir, anclar o inmovilizar ácidos nucleicos a la superficie del soporte sólido.

#### Analitos

Los analitos son biomoléculas cuya función, composición, identidad y/o su fuente se investigan. Analitos ilustrativos incluyen, pero no se limitan a ADN, ARN, ADNc, proteínas, lípidos, carbohidratos, organelas celulares (*p. ej.*, núcleos, aparatos de Golgi, ribosomas, mitocondrias, retículo endoplasmático, cloroplasto, membrana celular, etc.), metabolitos celulares, secciones de tejido, células, células individuales, contenido de células o de una célula individual, ácido nucleico aislado de células o de una célula individual, o ácido nucleico aislado de células o de una célula individual y ADN modificado o libre de células (*p. ej.*, de líquido placentario o plasma).

#### Ácidos nucleicos diana

Un ácido nucleico diana puede incluir cualquier ácido nucleico de interés. En una realización, el ácido nucleico diana puede incluir cualquier ácido nucleico de interés contenido, atrapado, incrustado o inmovilizado dentro de CE, tal como una matriz, gotita, emulsión, soporte sólido o compartimento que mantiene la contigüidad de los ácidos nucleicos dentro pero que permite la accesibilidad a líquidos y reactivos enzimáticos. Los ácidos nucleicos diana pueden incluir ADN, ADNc, productos de WGA, ARN, ácido nucleico peptídico, morfolino y ácido nucleico, ácido nucleico bloqueado, ácido nucleico de glicol, ácido nucleico de treosa, muestras mixtas de ácidos nucleicos, ADN de poliploidía (es decir, ADN de plantas), mezclas de los mismos e híbridos de los mismos. En una realización preferida, se usan fragmentos de ADN genómico o copias amplificadas de los mismos como el ácido nucleico diana. En otra realización preferida, se usa ADNc, ADN mitocondrial o ADN de cloroplasto.

Un ácido nucleico diana puede comprender cualquier secuencia de nucleótidos. En algunas realizaciones, el ácido nucleico diana comprende secuencias de homopolímeros. Un ácido nucleico diana también puede incluir secuencias repetidas. Las secuencias repetidas pueden tener cualquiera de una diversidad de longitudes que incluyen, por ejemplo, 2, 5, 10, 20, 30, 40, 50, 100, 250, 500 o 1000 nucleótidos o más. Las secuencias repetidas pueden repetirse, contiguas o no contiguas, cualquiera de una diversidad de veces, incluyendo, por ejemplo, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15 o 20 veces o más.

Algunas realizaciones descritas en el presente documento pueden utilizar un ácido nucleico diana individual. Otras realizaciones pueden utilizar una pluralidad de ácidos nucleicos diana. En tales realizaciones, una pluralidad de ácidos nucleicos diana puede incluir una pluralidad de los mismos ácidos nucleicos diana, una pluralidad de ácidos

nucleicos diana diferentes en que algunos ácidos nucleicos diana son iguales, o una pluralidad de ácidos nucleicos diana en que todos los ácidos nucleicos diana son diferentes. Realizaciones que utilizan una pluralidad de ácidos nucleicos diana se pueden llevar a cabo en formatos multiplex de modo que los reactivos sean suministrados simultáneamente a los ácidos nucleicos diana, por ejemplo, en una o más cámaras o en una superficie de matriz. En algunas realizaciones, la pluralidad de ácidos nucleicos diana puede incluir sustancialmente todo el genoma de un organismo particular. La pluralidad de ácidos nucleicos diana puede incluir al menos una porción del genoma de un organismo particular que incluye, por ejemplo, al menos aproximadamente 1%, 5%, 10%, 25%, 50%, 75%, 80%, 85%, 90%, 95% o 99% del genoma. En realizaciones particulares, la porción puede tener un límite superior que es a lo sumo aproximadamente 1%, 5%, 10%, 25%, 50%, 75%, 80%, 85%, 90%, 95% o 99% del genoma.

Ácidos nucleicos diana se pueden obtener de cualquier fuente. Por ejemplo, los ácidos nucleicos diana pueden prepararse a partir de moléculas de ácido nucleico obtenidas de un solo organismo o de poblaciones de moléculas de ácido nucleico obtenidas de fuentes naturales que incluyen uno o más organismos. Fuentes de moléculas de ácido nucleico incluyen, pero no se limitan a organelas, células, tejidos, órganos u organismos. Células que pueden usarse como fuentes de moléculas de ácido nucleico diana pueden ser procarióticas (células bacterianas, por ejemplo, géneros *Escherichia*, *Bacillus*, *Serratia*, *Salmonella*, *Staphylococcus*, *Streptococcus*, *Clostridium*, *Chlamydia*, *Neisseria*, *Treponema*, *Mycoplasma*, *Borrelia*, *Legionella*, *Pseudomonas*, *Mycobacterium*, *Helicobacter*, *Erwinia*, *Agrobacterium*, *Rhizobium*, y *Streptomyces*); arqueas, tales como crenarchaeota, nanoarchaeota o euryarchaeota; o eucarióticos, tales como hongos (por ejemplo, levaduras), plantas, protozoos y otros parásitos y animales (incluidos los insectos (por ejemplo, *Drosophila* spp.), nematodos (*p. ej.*, *Caenorhabditis elegans*) y mamíferos (por ejemplo, ratas, ratones, monos, primates no humanos y seres humanos).

Los ácidos nucleicos diana y los ácidos nucleicos molde pueden enriquecerse para ciertas secuencias de interés utilizando diversos métodos bien conocidos en la técnica. Se proporcionan ejemplos de métodos de este tipo en la publicación internacional N<sup>o</sup>. WO/2012/108864.

En algunas realizaciones, los ácidos nucleicos pueden enriquecerse adicionalmente durante los métodos de preparación de colecciones de moldes. Por ejemplo, los ácidos nucleicos pueden enriquecerse para ciertas secuencias, antes de la inserción de transposomas, después de la inserción de transposomas y/o después de la amplificación de ácidos nucleicos.

Además, en algunas realizaciones, los ácidos nucleicos diana y/o los ácidos nucleicos molde pueden estar altamente purificados, por ejemplo, los ácidos nucleicos pueden estar al menos aproximadamente 70%, 80%, 90%, 95%, 96%, 97%, 98%, 99% o 100% libre de contaminantes antes de su uso con los métodos proporcionados en el presente documento. En algunas realizaciones, es beneficioso usar métodos conocidos en la técnica que mantengan la calidad y el tamaño del ácido nucleico diana, por ejemplo, el aislamiento y/o la transposición directa del ADN diana pueden realizarse usando tapones de agarosa.

En algunas realizaciones, el ácido nucleico diana puede obtenerse de una muestra biológica o una muestra de paciente. La expresión "muestra biológica" o "muestra del paciente" tal como se usa en el presente documento incluye muestras tales como una o más células, tejidos o fluidos corporales. Los "fluidos corporales" pueden incluir, pero no se limitan a sangre, suero, plasma, saliva, líquido cefalorraquídeo, líquido pleural, lágrimas, líquido del conducto lactal, linfa, esputo, orina, líquido amniótico o semen. Una muestra puede incluir un fluido corporal que es "acelular". Un "fluido corporal acelular" incluye menos de aproximadamente 1% (p/p) de material celular completo. El plasma o el suero son ejemplos de fluidos corporales acelulares. Una muestra puede incluir una muestra de origen natural o sintético (es decir, una muestra celular hecha para ser acelular).

El término "plasma" tal como se usa en el presente documento se refiere al fluido acelular encontrado en la sangre. El "plasma" puede obtenerse de la sangre eliminando material celular completo de la sangre mediante métodos conocidos en la técnica (*p. ej.*, centrifugación, filtración y similares).

Ciertos métodos de preparación de ácidos nucleicos molde

Algunas realizaciones incluyen métodos de preparación de ácidos nucleicos molde. Tal como se usa en el presente documento, "ácido nucleico molde" puede referirse a un sustrato para obtener información de la secuencia. En algunas realizaciones, un ácido nucleico molde puede incluir un ácido nucleico diana, un fragmento del mismo o cualquier copia del mismo que comprenda al menos una secuencia de transposón, un fragmento del mismo o cualquier copia del mismo. En algunas realizaciones, un ácido nucleico molde puede incluir un ácido nucleico diana que comprende un adaptador de secuenciación, tal como un sitio cebador de secuenciación. En algunas realizaciones, el CE puede comprender un ácido nucleico diana.

Algunos métodos de preparación de ácidos nucleicos molde incluyen la inserción de una secuencia de transposón en un ácido nucleico diana, preparando así un ácido nucleico molde. Algunos métodos de inserción incluyen poner en contacto una secuencia de transposón proporcionada en este documento con un ácido nucleico diana en presencia de una enzima, tal como una transposasa o integrasa, bajo condiciones suficientes para la integración de la secuencia o secuencias de transposón en el ácido nucleico diana. En algunas realizaciones, un CE puede comprender dicho ácido nucleico diana.

En algunas realizaciones, la inserción de secuencias de transposón en un ácido nucleico diana puede ser no aleatoria. En algunas realizaciones, las secuencias de transposón pueden ponerse en contacto con ácidos nucleicos diana que comprenden proteínas que inhiben la integración en ciertos sitios. Por ejemplo, las secuencias de transposones pueden inhibirse de integrarse en el ADN genómico que comprende proteínas, el ADN genómico que comprende cromatina, el ADN genómico que comprende nucleosomas o el ADN genómico que comprende histonas. En algunas realizaciones, las secuencias de transposón se pueden asociar con etiquetas de afinidad para integrar la secuencia de transposón en una secuencia particular en un ácido nucleico diana. Por ejemplo, una secuencia de transposón puede estar asociada con una proteína que fija como objetivo secuencias de ácido nucleico específicas, p. ej., histonas, proteínas de unión a cromatina, factores de transcripción, factores de iniciación, etc., y anticuerpos o fragmentos de anticuerpos que se unen a secuencias de proteínas de unión a ácidos nucleicos específicos para la secuencia particular. En una realización ilustrativa, una secuencia de transposón está asociada con una etiqueta de afinidad, tal como biotina; la etiqueta de afinidad puede asociarse con una proteína de unión a ácido nucleico. En algunas realizaciones, un CE puede comprender dicho ácido nucleico diana.

Se entenderá que durante la integración de algunas secuencias de transposón en un ácido nucleico diana, varios nucleótidos consecutivos del ácido nucleico diana en el sitio de integración se duplican en el producto integrado. Por lo tanto, el producto integrado puede incluir una secuencia duplicada en cada extremo de la secuencia integrada en el ácido nucleico diana. Tal como se usa en el presente documento, la expresión "etiqueta huésped" o "etiqueta g" puede referirse a una secuencia de ácido nucleico diana que se duplica en cada uno de los extremos de una secuencia de transposón integrada. Las porciones monocatenarias de ácidos nucleicos que pueden generarse mediante la inserción de secuencias de transposones pueden repararse mediante una diversidad de métodos bien conocidos en la técnica, por ejemplo, usando ligasas, oligonucleótidos y/o polimerasas.

En algunas realizaciones, una pluralidad de las secuencias de transposón proporcionadas en el presente documento se inserta en un ácido nucleico diana. Algunas realizaciones incluyen la selección de condiciones suficientes para lograr la integración de una pluralidad de secuencias de transposones en un ácido nucleico diana de manera que la distancia media entre cada una de las secuencias de transposón integrada comprenda un cierto número de nucleótidos consecutivos en el ácido nucleico diana.

Algunas realizaciones incluyen seleccionar condiciones suficientes para lograr la inserción de una secuencia o secuencias de transposón en un ácido nucleico diana, pero no en otra secuencia o secuencias de transposón. Se puede usar una diversidad de métodos para reducir la probabilidad de que una secuencia de transposón se inserte en otra secuencia de transposón. Ejemplos de métodos de este tipo, útiles con las realizaciones proporcionadas en el presente documento, se pueden encontrar en, por ejemplo, la solicitud de patente de EE.UU. N° de publicación 2012/0208705, solicitud de patente de EE.UU. N° de publicación 2012/0208724 y la solicitud de patente internacional N° de publicación WO 2012/061832.

En algunas realizaciones, las condiciones pueden seleccionarse de modo que la distancia media en un ácido nucleico diana entre secuencias de transposón integradas sea al menos aproximadamente 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100 o más nucleótidos consecutivos. En algunas realizaciones, la distancia media en un ácido nucleico diana entre secuencias de transposón integradas es de al menos aproximadamente 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000 o más nucleótidos consecutivos. En algunas realizaciones, la distancia media en un ácido nucleico diana entre secuencias de transposón integradas es de al menos aproximadamente 1 kb, 2 kb, 3 kb, 4 kb, 5 kb, 6 kb, 7 kb, 8 kb, 90 kb, 100 kb, o más nucleótidos consecutivos. En algunas realizaciones, la distancia media en un ácido nucleico diana entre secuencias de transposón integradas es de al menos aproximadamente 100 kb, 200 kb, 300 kb, 400 kb, 500 kb, 600 kb, 700 kb, 800 kb, 900 kb, 1000 kb, o más nucleótidos consecutivos. Como se entenderá, algunas condiciones que pueden seleccionarse incluyen poner en contacto un ácido nucleico diana con un cierto número de secuencias de transposones.

Algunas realizaciones de los métodos descritos en el presente documento incluyen la selección de condiciones suficientes para lograr la integración de al menos una porción de secuencias de transposón en un ácido nucleico diana que son diferentes. En realizaciones preferidas de los métodos y composiciones descritos en este documento, cada una de las secuencias de transposón integradas en un ácido nucleico diana es diferente. Algunas condiciones que pueden seleccionarse para lograr la integración de una cierta porción de secuencias de transposón en secuencias diana que son diferentes incluyen la selección del grado de diversidad de la población de secuencias de transposón. Como se entenderá, la diversidad de secuencias de transposones surge en parte debido a la diversidad de los códigos de barras de tales secuencias de transposones. Por consiguiente, algunas realizaciones incluyen proporcionar una población de secuencias de transposones en las que al menos una porción de los códigos de barras es diferente. En algunas realizaciones, al menos aproximadamente 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 98%, 99% o 100% de códigos de barras en una población de secuencias de transposones es diferente. En algunas realizaciones, al menos una porción de las secuencias de transposón integradas en un ácido nucleico diana son las mismas.

Algunas realizaciones de la preparación de un ácido nucleico molde pueden incluir copiar las secuencias que comprenden el ácido nucleico diana. Por ejemplo, algunas realizaciones incluyen hibridar un cebador con un sitio de cebador de una secuencia de transposón integrada en el ácido nucleico diana. En algunas de estas realizaciones, el cebador puede hibridarse con el sitio del cebador y extenderse. Las secuencias copiadas pueden incluir al menos

- una secuencia de código de barras y al menos una porción del ácido nucleico diana. En algunas realizaciones, las secuencias copiadas pueden incluir una primera secuencia de código de barras, una segunda secuencia de código de barras, y al menos una porción de un ácido nucleico diana dispuesto entre ellas. En algunas realizaciones, al menos un ácido nucleico copiado puede incluir al menos una primera secuencia de código de barras de un primer ácido nucleico copiado que puede identificarse o designarse para aparearse con una segunda secuencia de código de barras de un segundo ácido nucleico copiado. En algunas realizaciones, el cebador puede incluir un cebador de secuenciación. En algunas realizaciones, los datos de secuenciación se obtienen usando el cebador de secuenciación. En más realizaciones, los adaptadores que comprenden sitios de cebador pueden ligarse a cada uno de los extremos de un ácido nucleico, y el nucleico amplificarse a partir de dichos sitios de cebador.
- Algunas realizaciones de preparar un ácido nucleico molde pueden incluir secuencias amplificadoras que comprenden al menos una porción de una o más secuencias de transposón y al menos una porción de un ácido nucleico diana. En algunas realizaciones, al menos una porción de un ácido nucleico diana puede amplificarse usando cebadores que hibridan con sitios de cebador de secuencias de transposón integradas, integradas en un ácido nucleico diana. En algunas de estas realizaciones, un ácido nucleico amplificado puede incluir una primera secuencia de código de barras y una segunda secuencia de código de barras que tiene al menos una porción del ácido nucleico diana dispuesto entre ellas. En algunas realizaciones, al menos un ácido nucleico amplificado puede incluir al menos una primera secuencia de código de barras de un primer ácido nucleico amplificado que puede identificarse para emparejarse con una segunda secuencia de código de barras de una segunda secuencia amplificada.
- Algunos métodos para preparar ácidos nucleicos molde incluyen la inserción de secuencias de transposones que comprenden enlazadores monocatenarios. En un ejemplo, las secuencias de transposón (ME-P1-enlazador-P2-ME; sitio 1 de cebador del extremo del mosaico--sitio 2 de cebador del extremo del mosaico) se insertan en un ácido nucleico diana. El ácido nucleico diana que tiene las secuencias de transposón/enlazador insertadas puede extenderse y amplificarse.
- En una realización de las composiciones y los métodos descritos en el presente documento, se usan transposomas que tienen secuencias terminales transponibles simétricas para producir un fragmento de ácido nucleico diana etiquetado en el extremo (fragmento o tagmento etiquetado). Por lo tanto, cada uno de los fragmentos tagmentados contiene extremos idénticos, que carecen de direccionalidad. Luego se puede emplear una PCR con cebador individual, usando las secuencias finales del transposón, para amplificar el número de copias del molde de  $2n$  a  $2n \cdot 2^x$ , en que  $x$  corresponde al número de ciclos de PCR. En una etapa posterior, la PCR con cebadores puede agregar secuencias adicionales, tales como secuenciar secuencias de adaptador.
- En algunas realizaciones, puede ser ventajoso que cada uno de los ácidos nucleicos moldes incorpore al menos un sitio de cebador universal. Por ejemplo, un ácido nucleico molde puede incluir primeras secuencias extremas que comprenden un primer sitio de cebador universal y segundas secuencias extremas que comprenden un segundo sitio de cebador universal. Los sitios de cebadores universales pueden tener diversas aplicaciones, tales como el uso para amplificar, secuenciar y/o identificar uno o más ácidos nucleicos molde. El primer y segundo sitios de cebador universales pueden ser iguales, sustancialmente similares, similares o diferentes. Los sitios de cebador universales pueden introducirse en ácidos nucleicos mediante diversos métodos bien conocidos en la técnica, por ejemplo, la ligadura de sitios de cebadores a ácidos nucleicos, la amplificación de ácidos nucleicos usando cebadores de cola y la inserción de una secuencia de transposón que comprende un sitio de cebador universal.

#### Transposomas

- Un "transposoma" comprende una enzima de integración tal como una integrasa o transposasa, y un ácido nucleico que comprende un sitio de reconocimiento de integración, tal como un sitio de reconocimiento de transposasa. En realizaciones proporcionadas en el presente documento, la transposasa puede formar un complejo funcional con un sitio de reconocimiento de transposasa que es capaz de catalizar una reacción de transposición. La transposasa puede unirse al sitio de reconocimiento de transposasa e insertar el sitio de reconocimiento de transposasa en un ácido nucleico diana dentro de CE en un proceso que a veces se denomina "tagmentación". En algunos de estos eventos de inserción, una cadena del sitio de reconocimiento de transposasa puede transferirse al ácido nucleico diana. En un ejemplo, un transposoma comprende una transposasa dimérica que comprende dos subunidades y dos secuencias de transposón no contiguas. En otro ejemplo, una transposasa comprende una transposasa dimérica que comprende dos subunidades y una secuencia de transposón 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 MuA transposasa y un sitio de reconocimiento de Mu transposasa que comprende secuencias extremas R1 y R2 (Mizuuchi, K., *Cell*, 35: 785, 1983; Savilahti, H, *et al.*, *EMBO J.*, 14:4893, 1995). Las secuencias de ME también se pueden usar optimizadas por un artesano experto.

Más ejemplos de sistemas de transposición que pueden usarse con ciertas realizaciones de las composiciones y los métodos proporcionados en el presente documento incluyen *Staphylococcus aureus* Tn552 (Colegio *et al.*, *J. Bacteriol.*, 183: 2384-8, 2001; Kirby C *et al.*, *Mol. Microbiol.*, 43: 173-86, 2002), Ty1 (Devine y Boeke, *Nucleic Acids*

5 *Res.*, 22: 3765-72, 1994 y publicación internacional WO 95/23875), Transposón Tn7 (Craig, N L, *Science*. 271: 1512, 1996; Craig, N L, Revisión en: *Curr Top Microbiol Immunol.*, 204:27-48, 1996), Tn/O e IS10 (Kleckner N, *et al.*, *Curr Top Microbiol Immunol.*, 204:49-82, 1996), Transposasa Mariner (Lampe D J, *et al.*, *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 bacteriana (Ohtsubo y Sekine, *Curr. Top. Microbiol Immunol.* 204: 1-26, 1996), retrovirus (Brown, *et al.*, *Proc Natl Acad Sci USA*, 86:2525-9, 1989) y retrotransposón de levadura (Boeke y Corces, *Annu Rev Microbiol.* 43:403-34, 1989) Más ejemplos incluyen IS5, Tn10, Tn903, IS911 y versiones modificadas de enzimas de la familia transposasa (Zhang *et al.*, (2009) *PLoS Genet.* 5: e1000689. Epub 16 de octubre de 2009; Wilson C. *et al.* (2007) *J. Microbiol. Methods* 71:332-5).

Más ejemplos de integrasas que pueden usarse con los métodos y las composiciones proporcionados en el presente documento incluyen integrasas retrovirales y secuencias de reconocimiento de integrasa para integrasas retrovirales de este tipo, tales como integrasas de HIV-1, HIV-2, SIV, PFV-1, RSV.

#### Secuencias de transposones

15 Algunas realizaciones de las composiciones y métodos proporcionados en el presente documento incluyen secuencias de transposones. En algunas realizaciones, una secuencia de transposón incluye al menos un sitio de reconocimiento de transposasa. En algunas realizaciones, una secuencia de transposón incluye al menos un sitio de reconocimiento de transposasa y al menos un código de barras. Secuencias de transposones útiles con los métodos y las composiciones proporcionados en el presente documento se proporcionan en la solicitud de patente de EE.UU. N° de publicación 2012/0208705, solicitud de patente de EE.UU. N° de publicación 2012/0208724 y solicitud de patente internacional N° de publicación WO 2012/061832. En algunas realizaciones, una secuencia de transposón incluye un primer sitio de reconocimiento de transposasa, un segundo sitio de reconocimiento de transposasa y un código de barras dispuesto entre ellos.

#### Transposomas con secuencias de transposones no contiguas

25 Algunos transposomas proporcionados en el presente documento incluyen una transposasa que comprende dos secuencias de transposones. En algunas de estas realizaciones, las dos secuencias de transposón no están enlazadas entre sí, en otras palabras, las secuencias de transposón no son contiguas entre sí. Ejemplos de transposomas de este tipo son conocidos en la técnica, véase, *p. ej.*, la publicación de la solicitud de Patente de EE.UU. N° 2010/0120098.

#### 30 Estructuras en bucle

En algunas realizaciones, un transposoma comprende un ácido nucleico de secuencia de transposón que se une a dos subunidades de transposasa para formar un "complejo en bucle" o un "transposoma en bucle". En un ejemplo, un transposoma comprende una transposasa dimérica y una secuencia de transposón. Los complejos en bucle pueden garantizar que los transposones se inserten en el ADN diana mientras se mantiene la información de pedido del ADN diana original y sin fragmentar el ADN diana. Como se apreciará, las estructuras en bucle pueden insertar cebadores, códigos de barras, índices y similares en un ácido nucleico diana, mientras se mantiene la conectividad física del ácido nucleico diana. En algunas realizaciones, el CE puede comprender el ácido nucleico diana. En algunas realizaciones, la secuencia de transposón de un transposoma en bucle puede incluir un sitio de fragmentación de modo que la secuencia de transposón pueda fragmentarse para crear un transposoma que comprende dos secuencias de transposón. Transposomas de este tipo son útiles para asegurar que los fragmentos de ADN diana vecinos, en los que se insertan los transposones, reciban combinaciones de códigos que pueden ensamblarse sin ambigüedades en una etapa posterior del ensayo.

#### Ciertos métodos para hacer secuencias de transposones

45 Las secuencias de transposón proporcionadas en el presente documento pueden prepararse mediante una diversidad de métodos. Métodos ilustrativos incluyen síntesis directa y métodos de extensión de horquilla. En algunas realizaciones, las secuencias de transposón pueden prepararse por síntesis directa. Por ejemplo, una secuencia de transposón que comprende un ácido nucleico puede prepararse por métodos que comprenden síntesis química. Métodos de este tipo son bien conocidos en la técnica, *p.ej.*, síntesis en fase sólida usando precursores de fosforamidita tales como los derivados de 2'-desoxinucleósidos, ribonucleósidos o análogos de nucleósidos protegidos. Se pueden encontrar métodos ilustrativos para preparar la secuencia del transposón en, por ejemplo, la solicitud de patente de EE.UU. N° de publicación 2012/0208705, solicitud de patente de EE.UU. N° de publicación 2012/0208724 y solicitud de patente internacional N° de publicación WO 2012/061832

55 En algunas realizaciones que comprenden transposomas en bucle se pueden preparar secuencias de transposones que comprenden un enlazador monocatenario. En algunas realizaciones, el enlazador acopla las secuencias de transposón de un transposoma de tal manera que una secuencia de transposón que comprende una primera secuencia de reconocimiento de transposasa está acoplada a una segunda secuencia de transposón que comprende una segunda secuencia de reconocimiento de transposasa en una orientación de 5' a 3'. En algunas realizaciones, el enlazador acopla una secuencia de transposón que comprende una primera secuencia de

reconocimiento de transposasa a una segunda secuencia de transposón que comprende una segunda secuencia de reconocimiento de transposasa en una orientación de 5' a 5' o en una orientación de 3' a 3'. El acoplamiento de secuencias de transposones de un transposoma en una orientación de 5' a 5' o en una orientación de 3' a 3' puede ser ventajoso para evitar que los elementos de reconocimiento de transposasa, en particular los elementos de mosaico (ME o M), interactúen entre sí. Las secuencias de transposón acopladas pueden prepararse preparando secuencias de transposón que comprenden un grupo aldehído o un grupo oxiamina. Los grupos aldehído y oxiamina pueden interactuar para formar un enlace covalente, acoplando así las secuencias del transposón.

En algunas realizaciones, se pueden preparar transposomas que comprenden secuencias complementarias. En una realización, una transposasa se carga con secuencias de transposón que comprenden colas complementarias. Las colas se hibridan para formar una secuencia de transposón enlazada. La hibridación puede producirse en condiciones diluidas para disminuir la probabilidad de hibridación entre transposomas.

Inserción fijada como objetivo

En algunas realizaciones de los métodos y las composiciones proporcionados en el presente documento, las secuencias de transposón se pueden insertar en secuencias fijadas como objetivo particulares de un ácido nucleico diana. La transposición en ADNdc puede ser más eficiente que en dianas de ADNcs. En algunas realizaciones, el ADNdc se desnaturaliza en ADNcs y se reasocia con sondas de oligonucleótidos (20-200 bases). Estas sondas crean sitios de ADNdc que pueden usarse eficientemente como sitios de integración con transposomas proporcionados en el presente documento. En algunas realizaciones, el ADNdc puede ser fijado como objetivo usando la formación del bucle D con sondas de oligo recubiertas con recA, y la posterior formación de triplex. En algunas de realizaciones de este tipo, el bucle D es un sustrato preferido para los transposomas que comprenden la transposasa Tn4430. En más realizaciones, las regiones de interés en ADNdc pueden ser fijadas como objetivo usando proteínas de unión a ADN específicas de secuencia tales como complejos de dedos de zinc y otros ligandos de afinidad a regiones de ADN específicas.

En algunas realizaciones, los transposomas que comprenden una transposasa que tiene un sustrato preferido de posiciones emparejadas erróneamente en un ácido nucleico diana pueden usarse para fijar como objetivo la inserción en el ácido nucleico diana. Por ejemplo, algunas transposasas de MuA, tales como HYPERMU (Epicenter), prefieren dianas emparejadas erróneamente. En algunas de tales realizaciones, las sondas de oligonucleótidos que comprenden un emparejamiento erróneo se reasocian en un ácido nucleico diana monocatenario. Los transposomas que comprenden transposasas de MuA, tales como HYPERMU, se pueden usar para fijar como objetivo las secuencias emparejadas erróneamente del ácido nucleico diana.

Elemento de preservación de contigüidad (CE)

Un elemento de preservación de contigüidad (CE) es una entidad física que preserva al menos dos, o más, o todos los analitos en proximidad (o contigüidad) a través de una o más etapas de ensayo y proporciona acceso a reactivos de ensayo y puede agruparse y dividirse múltiples veces. sin perder la proximidad de los analitos.

En algunas realizaciones, el CE puede ser un soporte sólido. En una realización, el CE puede ser una emulsión o gotita. En algunas realizaciones, el CE es gel, hidrogel o perla de gel. En algunas realizaciones, el CE puede comprender un soporte sólido tal como perlas. En algunas realizaciones, las perlas pueden comprender, además, anticuerpos, oligonucleótidos y/o códigos de barras. En otra realización, el CE puede constituir una nanobola de ADN creada por WGA, RCA o condensación de cualquier reactivo de ácido nucleico.

En algunas realizaciones, se puede hacer un CE incrustando el ácido nucleico de las células o de una célula individual, o el producto de amplificación del mismo (de WGA, etc.) en una matriz polimérica tal como agarosa, poliacrilamida, alginato, etc. En algunas realizaciones, la contigüidad del contenido de las células o de una célula individual dentro de un CE se mantiene preservando la proximidad física de los componentes entre sí mediante encapsulación (tal como en una matriz polimérica), inmovilización en una perla o atrapamiento, manteniendo efectivamente información de contigüidad dentro del CE a través de rondas repetidas de agrupación y redistribución. La característica de que una colección de CE puede agruparse y dividirse independientemente, reaccionar con reactivos de ensayo, agruparse y dividirse nuevamente, etc., manteniendo sin embargo la contigüidad de los analitos que constituyen un CE individual permite la indexación combinatoria a través de diferentes etapas de división y agrupación.

En algunas realizaciones, los analitos en el elemento que preserva la contigüidad son accesibles para reactivos de ensayo que incluyen soluciones acuosas, enzimas (p. ej., fragmentasas, polimerasas, ligasas, transposasas, quinasas, endonucleasas de restricción, proteasas, fosfatasas, lipasas), adaptadores de ácido nucleico, códigos de barras de ácido nucleico, marcadores.

En algunas realizaciones, el CE comprende células o una célula individual. En algunas realizaciones, el CE comprende ácido nucleico de células o de una célula individual, tal como ADN, ARNm o ADNc; macromoléculas de células o de una célula individual, incluidas proteínas, polisacáridos, lípidos y ácidos nucleicos, así como moléculas pequeñas, tales como metabolitos primarios, metabolitos secundarios y productos naturales de células o de una célula individual. En algunas realizaciones, el ácido nucleico sufre amplificación tal como amplificación por PCR o del

genoma completo antes de formar el CE que comprende el ácido nucleico. En algunas realizaciones, el análisis del ADN y el ARNm se puede realizar en paralelo.

5 En algunas realizaciones, uno o más analitos de un CE están marcados con uno o más marcadores. Marcadores ilustrativos incluyen, pero no se limitan a códigos de barras o índices de ADN, marcadores fluorescentes, marcadores quimioluminiscentes, códigos de barras o índices de ARN, marcadores radioactivos, anticuerpos que comprenden un marcador, perlas que comprenden un marcador.

10 En algunas realizaciones, un método puede incluir las etapas de (a) compartimentar el CE que comprende ácido nucleico diana en una pluralidad de primeros recipientes; (b) proporcionar un primer índice al ácido nucleico diana de cada primer recipiente, obteniendo así un primer ácido nucleico indexado; (c) combinar los primeros ácidos nucleicos indexados; (d) compartimentar los primeros ácidos nucleicos molde indexados en una pluralidad de segundos recipientes; (e) proporcionar un segundo índice al primer ácido nucleico molde indexado de cada segundo recipiente, obteniendo así un segundo ácido nucleico indexado. Las etapas a-e se pueden continuar con ciclos adicionales de una o más etapas de la serie a-e para derivar compartimentos virtuales adicionales. Este método de indexación combinatoria se puede utilizar para crear efectivamente un gran número de compartimentos virtuales a partir de un número limitado de compartimientos físicos.

20 En algunas realizaciones, un método puede incluir las etapas de (a) proporcionar un CE que comprende analitos de ácido no nucleico (p, ej., proteínas) con informadores de ácido nucleico unidos; (b) compartimentar el CE en una pluralidad de primeros recipientes; (c) proporcionar un primer índice a los informadores de ácido nucleico diana de cada primer recipiente, obteniendo así un primer informador de ácido nucleico diana indexado; (c) combinar los primeros informadores de ácido nucleico indexado; (d) compartimentar los primeros CE indexados en una pluralidad de segundos recipientes ; (e) proporcionar un segundo índice a los primeros informadores de ácido nucleico indexados de cada segundo recipiente, obteniendo así un segundo informador de ácido nucleico indexado. Las etapas a-e se pueden continuar con ciclos adicionales de una o más etapas de la serie a-e para derivar compartimentos virtuales adicionales. La etapa de compartimentación puede incluir además la amplificación de ácido nucleico o la etapa de captura, tal como PLA, PEA u otra técnica que captura o amplifica ácidos nucleicos.

25 En algunas realizaciones, un tejido incrustado en parafina fijado con formalina puede dividirse en secciones, con cada sección añadida a un CE. Cada uno de los CE puede ser analizado posteriormente por contenido o secuencia y en una etapa posterior se puede obtener un mapa en 2D o 3D del contenido de cada frotis.

30 En algunas realizaciones, un ácido nucleico o ácidos nucleicos pueden incrustarse en una matriz que limita los ácidos nucleicos a un espacio definido, pero permite el acceso a reactivos para realizar etapas que incluyen, pero no se limitan a, amplificación (PCR, amplificación de genoma completo, extensión del cebador aleatorio, etc.), ligadura, transposición, hibridación, digestión de restricción y mutagénesis de ADN. Ejemplos de mutagénesis incluyen, pero no se limitan a, extensión propensa a errores, alquilación, conversión de bisulfito y desaminasas inducidas por activación (citidina), etc.

35 En algunas realizaciones, los métodos y las composiciones que usan CE pueden combinarse con enfoques de ensamblaje mutagénico para mejorar en gran medida el ensamblaje de la información de la secuencia de ADN. El ADN genómico puede fragmentarse y dividirse en una pluralidad de CE, comprendiendo cada CE una fracción del genoma. Las diferentes fracciones del genoma reciben diferentes códigos de barras, lo que permite que las fracciones del genoma se ensamblen de forma independiente. Uno de los desafíos más grandes es el ensamblaje de repeticiones. Un método para ensamblar repeticiones se describe en Levy, D. y Wigler, M. (2014) Facilitated sequence counting and assembly by template mutagenesis. Proc. of the Natl. Acad. Sci. 111 (43). E4632-E4637. ISSN 0027-8424. Los enfoques de ensamblaje también se discuten en el documento US20140024537, titulado: Methods And Systems for Determining Haplotypes And Phasing of Haplotypes.

45 Para métodos que combinan la partición de fragmentos de ADN con una mutagénesis o enfoque relacionado, la partición se puede realizar con CE, pocillos, índices, índices virtuales, compartimentos físicos, gotitas, etc. La mutagénesis se puede realizar mediante varios métodos, incluidos, pero no limitados a la extensión propensa a errores, alquilación, conversión de bisulfito y desaminasas inducidas por activación (citidina), etc. El método de división del ácido nucleico en CE y el enfoque de mutagénesis pueden ser útiles en los caos en los que los métodos convencionales dificultan el ensamblaje de repeticiones o regiones difíciles.

50 En algunas realizaciones, los métodos recogidos en el presente documento pueden usarse para el ajuste de fases variante, ensamblaje del genoma (*de novo*), rastreo de poblaciones de células para determinar la heterogeneidad en la población y determinar las diferencias de célula a célula.

55 En algunas realizaciones, el ADNc de las células o de una célula individual se aísla en recipientes y se convierte en un CE que se indexa a través del enfoque de compartimentación virtual tal como se describió anteriormente. Esto permite la expresión génica y el perfilado del transcrito de un número de 1000, 10.000, 100.000 e incluso un mayor número de diferentes colecciones de células individuales indexadas.

En algunas realizaciones, el número de células individuales que se pueden analizar es aproximadamente el 10% del número total de compartimentos virtuales debido al muestreo de Poisson. Para un esquema de indexación de cuatro

niveles con compartimentos de 96 pocillos en cada etapa, se puede analizar un total de  $10\% \times 96 \times 96 \times 96 =$  más de 8 millones de células individuales en un experimento usando un total de  $4 \times 96 = 384$  compartimentos físicos. En el ejemplo de la Fig. 3, se utilizan cuatro etapas combinatorias de dilución y agrupación para crear un gran número de compartimentos virtuales (un conjunto de moléculas o elementos de colección de ADN que contienen una combinación de índice única). En este ejemplo, el recipiente de ADN contiguo se crea mediante la encapsulación de los contenidos de una célula individual en una matriz polimérica (p. ej., PAM = poliacrilamida). En una realización particular preferida para el análisis genómico, los contenidos de ADN genómico de la célula individual se amplifican mediante MDA (una reacción de amplificación de desplazamiento múltiple WGA). Este producto MDA de una célula individual constituye el recipiente de ADN que procede a través del esquema de indexación combinatoria. Para la expresión génica se puede hacer una preparación de ADNc de una célula individual a partir del recipiente de una célula individual tal como se describe por Picelli (Picelli, 2014). En la realización preferida, los índices iniciales se unen al ADN genómico o ADNc a través de técnicas de preparación de colecciones estándar usando fragmentación (enzimática) y ligadura de adaptadores, o mediante tagmentación usando complejos de transposasa. En la realización preferida, los índices posteriores se unen a la colección mediante ligadura o PCR. Se prefiere la ligadura, ya que es fácil agregar adaptadores indexados de una manera secuencial. La etapa final puede involucrar solo PCR indexada o ligadura y PCR.

En algunas realizaciones, el ácido nucleico diana está protegido contra histonas/proteínas (véase Buenrostro et al. Nature Methods 10, 1213-1218 (2013) doi:10.1038/nmeth.2688.). Las aplicaciones incluyen perfiles epigenómicos y el análisis de cromatina abierta y proteínas de unión al ADN y posición de nucleosomas.

En algunas realizaciones, los elementos que preservan la contigüidad pueden comprender una célula individual y el ácido nucleico de la célula puede amplificarse. Posteriormente, cada uno de los elementos de preservación de contigüidad puede indexarse de manera única a través del esquema de indexación combinatoria. Las lecturas de secuenciación corta se pueden agrupar en función de un índice único. Las lecturas sintéticas largas pueden ser secuencialmente ensambladas *de novo* en base a un índice único (McCoy et al. Plosone 2014 (DOI: 10.1371/journal.pone.0106689) Illumina TruSeq Synthetic Long-Reads Empower De Novo Assembly and Resolve Complex, High-Repetitive Transposable Elements).

En algunas realizaciones, CE puede comprender el contenido de una célula, por ejemplo, proteínas, organelas, ARN, ADN, ribosomas, anticuerpos, esteroides, estructuras especializadas, glicanos, lípidos, moléculas pequeñas, moléculas que pueden afectar a una vía biológica, monosacáridos y polisacáridos, alcaloides, metabolitos primarios y secundarios.

En algunas realizaciones, las organelas dentro del CE pueden estar teñidos diferencialmente. Ejemplos de reactivos de tinción de organelas son las proteínas fluorescentes que fijan como objetivo organelas (Cellular Lights™), las manchas clásicas de organelas o conjugados de colorantes que selectivamente o no selectivamente pueden marcar organelas o estructuras celulares.

En algunas realizaciones, un analito de interés en un CE es una proteína. Las proteínas se pueden marcar con códigos de barras o marcadores alternativos. El código de barras o los marcadores se pueden leer utilizando matrices tradicionales o métodos basados en secuencias. Los enfoques de ligadura de proximidad y las secuencias de índice de anticuerpos se pueden usar para detectar proteínas (Fredriksson et al. Nature Biotechnology 20, 473-477 (2002)) junto con la detección de las secuencias de códigos de barras para establecer la identidad y abundancia de las proteínas en cada una de las células individuales. Las proteínas se pueden marcar por diversos métodos ([www.piercenet.com/cat/protein-antibody-labeling](http://www.piercenet.com/cat/protein-antibody-labeling)) conocidos por un trabajador calificado que incluye estrategias de marcaje químico específicas del sitio *in vivo* e *in vitro*.

Ligadura de proximidad (Duo-link PLA, [www.sigmaaldrich.com/life-science/molecular-biology/molecular-biology-products.html?TablePage=112232138](http://www.sigmaaldrich.com/life-science/molecular-biology/molecular-biology-products.html?TablePage=112232138), Ensayo de ligadura de proximidad multiplexada documento EP 2714925 A1) es un ejemplo para la detección de proteínas, interacciones proteína-proteína y modificaciones post-traduccionales que se pueden adaptar para su uso en un elemento de preservación de contigüidad. Este método puede usarse para detectar y cuantificar una proteína específica o un complejo de proteínas en un elemento de preservación de contigüidad. Un ejemplo de un flujo de trabajo es el siguiente: (1) hacer un elemento o elementos que preservan la contigüidad, (2) lavar y agregar un par o pares de anticuerpos primarios específicos para la proteína de interés, (3) lavar y teñir con anticuerpos marcados con código de barras. Cada una de las poblaciones de elementos que preservan la contigüidad en un recipiente recibe un anticuerpo marcado con un código de barras diferente. Mediante la ligadura de proximidad del par o los pares de anticuerpos primarios, se pueden formar productos amplificables que contienen un código de barras único para una proteína específica. Un código de barras puede ser específico para la proteína de interés, mientras que otros códigos de barras se usan para asignar la proteína a un elemento o célula específico que preserva la contigüidad. A través de una o más etapas de división y agrupación, las fracciones se pueden marcar de manera diferencial. Como tal, el contenido de elementos de preservación de contigüidad individuales puede analizarse sin la necesidad de procesar cada uno de los elementos de preservación de contigüidad individualmente en muchas etapas paralelas. Es particularmente una gran ventaja procesar 10, 100, 1000, 10.000, 100.000, 1.000.000, 10.000.000, 100.000.000 y más elementos de preservación de contigüidad de tal manera. Esteroides y moléculas pequeñas se pueden detectar de manera similar a la descrita para las proteínas. Los anticuerpos marcados con códigos de barras se pueden desarrollar para los esteroides (Hum Reprod. enero de

1988; 3(1):63-8. Anticuerpos contra esteroides. Bösze P et al.. Alternativamente, se han descrito colorantes fluorescentes y conjugados radiactivos ([www.jenabioscience.com/cms/en/1/catalog/2305\\_fluorescent\\_hormones.html](http://www.jenabioscience.com/cms/en/1/catalog/2305_fluorescent_hormones.html)). Estos conjugados de anticuerpos para esteroides pueden procesarse tal como se describe anteriormente. Se pueden usar diversos métodos para detectar uno o más componentes del elemento de preservación de contigüidad. Uno o más componentes del elemento de preservación de contigüidad pueden marcarse con sondas quimioluminiscentes, fluorescentes, radiactivas, etiquetas de ADN, códigos de barras e índices. Se pueden utilizar estrategias de amplificación para potenciar la señal. Por ejemplo, la amplificación de círculo rodante (RCA) se puede utilizar para detectar analitos. Los productos de RCA pueden detectarse posteriormente mediante decodificadores de secuenciación fluorescentes (sondas). Adicionalmente, se pueden utilizar micromatrices, matrices de proteínas, secuenciación, secuenciación de nano-poros, secuenciación de próxima generación, electroforesis capilar, matrices de perlas para la lectura.

Establecer la contigüidad del contenido de una célula

En algunas realizaciones, la contigüidad del contenido de células o de una célula individual, por ejemplo, pero sin limitarse a ADN, ARN, proteínas, organelas, metabolitos, moléculas pequeñas, se puede preservar en un elemento de preservación de contigüidad (CE). Se puede crear un CE mediante varios métodos que incluyen, pero no se limitan a encapsular el contenido dentro de una gotita, incrustar el contenido en una matriz polimérica (después de la encapsulación) y unir el contenido a una perla. En la realización preferida, el CE es permeable a reactivos de ensayo tales como tampones acuosos, enzimas (polimerasas, ligasas, transposasas, etc.), nucleótidos, adaptadores de oligonucleótidos, transposones y cebadores, etc. Las colecciones indexadas se crean a partir de este CE tal como se describe arriba. Las rondas repetidas de dilución en compartimentos físicos, la unión de índices específicos para el compartimento, la agrupación y la redilución en compartimentos adicionales conducen a una creación exponencial de muchos compartimentos virtuales. Si se diseña adecuadamente, el contenido de cada uno de los CE, al final, se indexará virtualmente con un código de barras único. Como un ejemplo en la Fig. 1, un esquema de indexación de cuatro niveles conduce a un gran número de compartimentos e índices virtuales (> 84 millones) con solo  $4 \times 96 = 384$  compartimentos físicos totales. En la realización preferida, se añaden índices específicos para el compartimento en cada nivel de compartimentación mediante tagmentación, ligadura o PCR. En la realización preferida, cada uno de los compartimentos físicos en cada etapa tiene un índice único. La compartimentación posterior puede usar los mismos o diferentes índices. Si se utilizan los mismos índices de un nivel de compartimentación al siguiente, la posición del índice dentro de la serie de secuencias final identificará el compartimento y el nivel de compartimentación.

Análisis de componentes celulares utilizando gotitas

En una realización, el CE puede ser una emulsión o gotita. En una realización, el CE es una gotita en contacto con aceite. En un ejemplo, el CE que comprende ácido nucleico incluye la dilución y el reparto de una muestra de ácido nucleico en gotitas, compartimentos o perlas. En una realización, la gotita comprende células o una célula individual. En una realización, el CE que comprende células individuales incluye la dilución y el reparto de una célula individual en gotitas, compartimentos o perlas.

En algunas realizaciones, una "Gotita" puede ser un volumen de líquido en un actuador de gotitas que está al menos parcialmente limitado por el fluido de relleno. Por ejemplo, una gotita puede estar completamente rodeada por fluido de relleno o puede estar limitada por fluido de relleno y una o más superficies de un actuador de gotas. Las gotitas pueden adoptar una gran diversidad de formas; ejemplos no limitativos incluyen generalmente forma de disco, en forma de babosa, esfera truncada, elipsoide, esférica, esfera parcialmente comprimida, hemisférica, ovoide, cilíndrica y diversas formas configuradas durante las operaciones de las gotitas, tales como la fusión o división o formadas como resultado del contacto de formas de este tipo con una o más superficies de un actuador de gotitas. Los actuadores de gotitas se utilizan para realizar una amplia diversidad de operaciones con las gotitas. Un actuador de gotitas incluye típicamente dos sustratos separados por un espacio. Los sustratos incluyen electrodos para realizar operaciones con las gotitas. El espacio se llena típicamente con un fluido de relleno que es inmisible con el fluido que se va a manipular en el actuador de gotitas. Las superficies expuestas al espacio son típicamente hidrófobas. El análisis del material genético (genómica) y su expresión (genómica funcional), proteómica, análisis de colección combinatoria y otras aplicaciones bioanalíticas multiplexadas se pueden realizar en gotitas y las siguientes operaciones se pueden llevar a cabo en el actuador de gotitas de análisis. Métodos para manipular las gotitas usando el actuador de gotitas se describen en las publicaciones de solicitud de EE.UU. 20100130369 y 20130203606, respectivamente.

"Actuador de gotitas" significa un dispositivo para manipular gotitas. Para ejemplos de actuadores de gotitas, véase Pamula et al., Patente de EE.UU. N° 6.911.132, titulado "Apparatus for Manipulating Droplets by Electrowetting-Based Techniques", expedida el 28 de junio de 2005; Pamula et al., publicación de patente de EE.UU. N° 20060194331, titulado "Apparatuses and Methods for Manipulating Droplets on a Printed Circuit Board", expedida el 31 de agosto de 2006; Pollack et al., Publicación de patente internacional N° WO/2007/120241, titulada "Droplet-Based Biochemistry", expedida el 25 de octubre de 2007; Shenderov Patente de EE.UU. N° 6.773.566, titulada "Electrostatic Actuators for Microfluidics and Methods for Using Same", expedida el 10 de agosto de 2004; Shenderov, Patente de EE.UU. N° 6.565.727, titulada "Actuators for Microfluidics Without Moving Parts", expedida el 20 de mayo de 2003; Kim et al., Publicación de patente de EE.UU. N° 20030205632, titulada "Electrowetting-driven

Micropumping", expedida el 6 de noviembre de 2003; Kim et al., Publicación de patente de EE.UU. N° 20060164490, titulada "Method and Apparatus for Promoting the Complete Transfer of Liquid Drops from a Nozzle", expedida el 27 de julio de 2006; Kim et al., Publicación de patente de EE.UU. N° 20070023292, titulada "Small Object Moving on Printed Circuit Board", expedida el 1 de febrero de 2007; Shah et al., Publicación de patente de EE.UU. N° 20090283407, titulada "Method for Using Magnetic Particles in Droplet Microfluidics", publicada el 19 de noviembre de 2009; Kim et al., Publicación de patente de EE.UU. N° 20100096266, titulada "Method and Apparatus for Real-time Feedback Control of Electrical Manipulation of Droplets on Chip", publicada el 22 de abril de 2010; Velev, Patente de EE.UU. N° 7.547.380, titulada "Droplet Transportation Devices and Methods Having a Fluid Surface", expedida el 16 de junio de 2009; Sterling et al., Patente de EE.UU. N° 7.163.612, titulada "Method, Apparatus and Article for Microfluidic Control via Electrowetting, for Chemical and Biological Assay and the Like", expedida el 16 de enero de 2007; Becker et al., Patente de EE.UU. N° 7.641.779, titulada "Method and Apparatus for Programmable Fluidic Processing", expedida el 5 de enero de 2010; Becker et al., Patente de EE.UU. N° 6.977.033, titulada "Method and Apparatus for Programmable Fluidic Processing", expedida el 20 de diciembre de 2005; Decre et al., Patente de EE.UU. N° 7.328.979, titulada "System for Manipulation of a Body of Fluid", expedida el 12 de febrero de 2008; Yamakawa et al., Publicación de patente de EE.UU. N° 20060039823, titulada "Chemical Analysis Apparatus", publicada el 23 de febrero de 2006; Wu, Publicación de patente de EE.UU. N° 20110048951, titulada "Digital Microfluidics Based Apparatus for Heta-exchanging Chemical Processes", publicada el 3 de marzo de 2011; Fouillet et al., Publicación de patente de EE.UU. N° 20090192044, titulada "Electrode Addressing Method", publicada el 30 de julio de 2009; Fouillet et al., Patente de EE.UU. N° 7.052.244, titulada "Device for Displacement of Small Liquid Volumes Along a Micro-catenary Line by Electrostatic Forces", expedida el 30 de mayo de 2006; Marchand et al., Publicación de patente de EE.UU. N° 20080124252, titulada "Droplet Microreactor", publicada el 29 de mayo de 2008; Adachi et al., Publicación de patente de EE.UU. N° 20090321262, titulada "Liquid Transfer Device", publicada el 31 de diciembre de 2009; Roux et al., Publicación de patente de EE.UU. N° 20050179746, titulada "Device for Controlling the Displacement of a Drop Between Two or Several Solid Substrates", publicada el 18 de agosto de 2005; y Dhindsa et al., "Virtual Electrowetting Channels: Electronic Liquid Transport with Continuous Channel Functionality", Lab Chip, 10: 832-836 (2010).

Ciertos actuadores de gotitas incluirán uno o más sustratos dispuestos con un espacio de operaciones de gotitas entre ellos y electrodos asociados (p. ej., en capas, unidos y/o incrustados) en uno o más sustratos y dispuestos para realizar una o más operaciones de gotas. Por ejemplo, ciertos actuadores de gotitas incluirán un sustrato base (o inferior), electrodos de operación de gotitas asociados con el sustrato, una o más capas dieléctricas encima del sustrato y/o los electrodos y, opcionalmente, una o más capas hidrófobas encima del sustrato, capas dieléctricas y/o los electrodos que forman una superficie de operaciones de gotitas. También se puede proporcionar un sustrato superior, que está separado de la superficie de operaciones de gotitas por un espacio, comúnmente denominado espacio de operaciones de gotitas. Se discuten diversas disposiciones de electrodos en los sustratos superior y/o inferior en las patentes y solicitudes mencionadas anteriormente y se discuten ciertas disposiciones de electrodos novedosas en la descripción de la presente divulgación. Durante las operaciones de gotitas, se prefiere que las gotitas permanezcan en contacto continuo o contacto frecuente con un electrodo de tierra o de referencia. Se puede asociar un electrodo de tierra o de referencia con el sustrato superior orientado hacia el espacio, el sustrato inferior orientado hacia el espacio, en el espacio. Cuando se proporcionan electrodos en ambos sustratos, los contactos eléctricos para acoplar los electrodos a un instrumento actuador de gotitas para controlar o monitorear los electrodos pueden estar asociados con una o ambas placas. En algunos casos, los electrodos en un sustrato están acoplados eléctricamente al otro sustrato, de modo que solo un sustrato está en contacto con el actuador de gotitas. En una realización, un material conductor (p. ej., un epoxi, tal como el Sistema de polímero EP79 MASTER BOND™, disponible de Master Bond, Inc., Hackensack, NJ) proporciona la conexión eléctrica entre electrodos en un sustrato y rutas eléctricas en los otros sustratos, por ejemplo, un electrodo de tierra en un sustrato superior puede estar acoplado a una ruta eléctrica en un sustrato inferior por un material conductor de este tipo. Cuando se usan múltiples sustratos, se puede proporcionar un espaciador entre los sustratos para determinar la altura del espacio entre ellos y definir depósitos dispensadores en el actuador. La altura del separador puede ser, por ejemplo, de al menos aproximadamente 5 µm, 100 µm, 200 µm, 250 µm, 275 µm o más. Alternativa o adicionalmente, la altura del separador puede ser como máximo de aproximadamente 600 µm, 400 µm, 350 µm, 300 µm o menos. El separador puede, por ejemplo, estar formado por una capa de proyecciones que forman los sustratos superior o inferior, y/o un material insertado entre los sustratos superior e inferior. Se pueden proporcionar una o más aberturas en el uno o más sustratos para formar una ruta de fluido a través de la cual se puede suministrar líquido en el espacio de operaciones de gotitas. En algunos casos, la una o más aberturas pueden estar alineadas para la interacción con uno o más electrodos, p. ej., alineadas de manera que el líquido que fluye a través de la abertura se acerque lo suficiente a uno o más electrodos de operaciones de gotitas para permitir que una operación de gotitas sea efectuada por los electrodos de operaciones de gotitas utilizando el líquido. Los sustratos base (o de fondo) y superiores se pueden formar en algunos casos como un componente integral. Se pueden proporcionar uno o más electrodos de referencia en los sustratos base (o fondo) y/o superiores y/o en el espacio. Se proporcionan ejemplos de disposiciones de electrodos de referencia en las patentes y solicitudes de patente mencionadas anteriormente. En diversas realizaciones, la manipulación de gotitas por parte de un accionador de gotitas puede estar mediada por electrodo, p. ej., mediada por electrohumectación o mediada por dielectroforesis o mediada por la fuerza Coulombiana. Ejemplos de otras técnicas para controlar las operaciones de gotitas que se pueden usar en los actuadores de gotitas de la presente divulgación incluyen el uso de dispositivos que inducen una presión fluidica hidrodinámica, tales como los que operan sobre la base de principios mecánicos (p. ej., bombas de jeringas

externas, bombas de membrana neumática, bombas de membrana vibratoria, dispositivos de vacío, fuerzas centrífugas, bombas piezoeléctricas/ultrasónicas y fuerzas acústicas); principios eléctricos o magnéticos (p. ej., flujo electro-osmótico, bombas electrocinéticas, tapones ferrofluídicos, bombas electrohidrodinámicas, atracción o repulsión utilizando fuerzas magnéticas y bombas magnetohidrodinámicas); principios termodinámicos (p. ej., generación de burbujas de gas/expansión de volumen inducida por cambio de fase); otros tipos de principios de humectación de la superficie (p. ej., electrohumectación y optoelectrohumectación, así como gradientes de tensión superficial inducidos química, térmica, estructural y radiactivamente); gravedad; tensión superficial (p. ej., acción capilar); fuerzas electrostáticas (p. ej., flujo electro-osmótico); flujo centrífugo (sustrato dispuesto en un disco compacto y rotado); fuerzas magnéticas (p. ej., los iones oscilantes causan flujo); fuerzas magnetohidrodinámicas; y vacío o presión diferencial. En ciertas realizaciones, se pueden emplear combinaciones de dos o más de las técnicas anteriores para conducir una operación de gotitas en un accionador de gotitas de la presente divulgación. De manera similar, uno o más de los anteriores pueden usarse para suministrar líquido a un espacio de operaciones de gotitas, p. ej., desde un depósito en otro dispositivo o desde un depósito externo del actuador de gotitas (p. ej., un depósito asociado con un sustrato del actuador de gotitas y un ruta de flujo desde el depósito hacia la espacio de operaciones de gotitas). Las superficies de operaciones de gotitas de ciertos actuadores de gotitas de la presente divulgación pueden estar hechas de materiales hidrófobos o pueden estar recubiertas o tratadas para hacerlas hidrófobas. Por ejemplo, en algunos casos, una parte o la totalidad de las superficies de operaciones de gotitas pueden derivatizarse con materiales o productos químicos de baja energía superficial, p. ej., por deposición o usando síntesis in situ usando compuestos tales como compuestos poli- o per-fluorados en solución o monómeros polimerizables. Ejemplos incluyen TEFLON® AF (disponible de DuPont, Wilmington, DE), miembros de la familia de materiales cytop, recubrimientos en la familia FLUOROPEL® de recubrimientos hidrófobo y superhidrófobos (disponibles de Cytonix Corporation, Beltsville, MD), recubrimientos de silano, recubrimientos de fluorosilano, derivados de fosfonato hidrófobos (p. ej., los vendidos por Aculon, Inc) y recubrimientos electrónicos NOVEC™ (disponibles de 3M Company, St. Paul, MN), otros monómeros fluorados para la deposición química de vapor mejorada con plasma (PECVD) y organosiloxano (p. ej., SiOC) para PECVD. En algunos casos, la superficie de operaciones de gotitas puede incluir un recubrimiento hidrófobo que tiene un espesor que varía de aproximadamente 10 nm a aproximadamente 1.000 nm. Además, en algunas realizaciones, el sustrato superior del actuador de gotitas incluye un polímero orgánico conductor de la electricidad, que luego se reviste con un recubrimiento hidrófobo o se trata de otro modo para hacer que las operaciones de gotitas sean hidrófobas en la superficie. Por ejemplo, el polímero orgánico conductor de la electricidad que se deposita sobre un sustrato plástico puede ser poli(3,4-etilendioxitiofeno) poli(estirenosulfonato) (PEDOT: PSS). Otros ejemplos de polímeros orgánicos conductores de la electricidad y capas conductoras alternativas se describen en Pollack et al., Publicación de patente internacional N° WO/2011/002957, titulada "Droplet Actuator Devices and Methods", publicada el 6 de enero de 2011. Uno o ambos sustratos pueden fabricarse utilizando una placa de circuito impreso (PCB), vidrio, vidrio recubierto con óxido de indio y estaño (ITO) y/o materiales semiconductores como el sustrato. Cuando el sustrato es vidrio recubierto con ITO, el recubrimiento de ITO tiene preferiblemente un espesor de al menos aproximadamente 20 nm, 50 nm, 75 nm, 100 nm o más. Alternativa o adicionalmente, el grosor puede ser como máximo de aproximadamente 200 nm, 150 nm, 125 nm o menos. En algunos casos, el sustrato superior y/o inferior incluye un sustrato de PCB que está recubierto con un dieléctrico, tal como un dieléctrico de poliimida, que en algunos casos también puede estar recubierto o tratado de otro modo para hacer que las operaciones de gotitas sean hidrofóbicas en la superficie. Cuando el sustrato incluye una PCB, los siguientes materiales son ejemplos de materiales adecuados: MITSUI™ BN-300 (disponible de MITSUI Chemicals America, Inc., San Jose, CA); ARLON™ 11N (disponible de Arlon, Inc, Santa Ana, CA); NELCO® N4000-6 y N5000-30/32 (disponible de Park Electrochemical Corp., Melville, NY); ISOLA™ FR406 (disponible de Isola Group, Chandler, AZ), especialmente IS620; familia de fluoropolímeros (adecuada para la detección de fluorescencia, ya que tiene baja fluorescencia de fondo); familia de poliimida; poliéster; poli(naftalato de etileno); policarbonato; polieteretercetona; polímero de cristal líquido; copolímero de cicloolefina (COC); polímero de cicloolefina (COP); aramida; refuerzo de aramida no tejido THERMOUNT® (disponible de DuPont, Wilmington, DE); fibra de marca NOMEX® (disponible de DuPont, Wilmington, DE); y papel. Diversos materiales también son adecuados para su uso como el componente dieléctrico del sustrato. Ejemplos incluyen: dieléctrico depositado por vapor, tal como PARYLENE™ C (especialmente en vidrio), PARYLENE™ N y PARYLENE™ HT (para alta temperatura, ~300°C) (disponible de Parylene Coating Services, Inc., Katy, TX); recubrimientos de TEFLON® AF; citopático; máscaras de soldadura, tales como máscaras de soldadura líquidas fotoimpresables (p. ej., en PCB), tales como las series TAIYO™ PSR4000, TAIYO™ PSR y AUS (disponibles de Taiyo America, Inc. Carson City, NV) (buenas características térmicas para aplicaciones que involucran un control térmico), y PROBIMER™ 8165 (buenas características térmicas para aplicaciones que involucran un control térmico (disponible de Huntsman Advanced Materials Americas Inc., Los Angeles, CA); máscara de soldadura de película seca, tal como las de la línea de máscara de soldadura de película seca VACREL® (disponible de DuPont, Wilmington, DE); dieléctricos de película, tales como película de poliimida (p. ej., película de poliimida KAPTON®, disponible de DuPont, Wilmington, DE), polietileno y fluoropolímeros (p. ej., FEP), politetrafluoroetileno; poliéster; poli(naftalato de etileno); copolímero de cicloolefina (COC); polímero de cicloolefina (COP); cualquier otro material de sustrato de PCB enumerado anteriormente; resina de matriz negra; polipropileno; y materiales de circuito flexible negro, tales como DuPont™ Pyralux® HXC y DuPont™ Kapton® MBC (disponible de DuPont, Wilmington, DE). La tensión y la frecuencia de transporte de gotitas se pueden seleccionar para el rendimiento con reactivos utilizados en protocolos de ensayo específicos. Los parámetros de diseño pueden variar, p. ej., el número y la ubicación de los depósitos en el actuador, el número de conexiones de electrodos independientes, el tamaño (volumen) de los diferentes depósitos, la ubicación de imanes/zonas de lavado de perlas, el tamaño del electrodo, el paso entre

electrodos y la altura del espacio (entre sustratos superior e inferior) se puede variar para usar con reactivos específicos, protocolos, volúmenes de gotitas, etc. En algunos casos, un sustrato de la presente divulgación puede derivatizarse con materiales o productos químicos de baja energía superficial, p. ej., usando deposición o síntesis in situ utilizando compuestos poli- o per-fluorados en solución o monómeros polimerizables. Ejemplos incluyen revestimientos de TEFLON®AF y revestimientos de FLUOROPEL® para revestimiento por inmersión o pulverización, otros monómeros fluorados para la deposición de vapor químico potenciado por plasma (PECVD) y organosiloxano (p. ej., SiOC) para PECVD. Adicionalmente, en algunos casos, una parte o la totalidad de la superficie de operaciones de gotitas puede estar recubierta con una sustancia para reducir el ruido de fondo, tal como la fluorescencia de fondo de un sustrato de PCB. Por ejemplo, el recubrimiento reductor de ruido puede incluir una resina de matriz negra, tal como las resinas de matriz negra disponibles de Toray industries, Inc., Japón. Electrodo de un actuador de gotitas son controlados típicamente por un controlador o un procesador, que se proporciona como parte de un sistema, que puede incluir funciones de procesamiento, así como almacenamiento de datos y software y capacidades de entrada y salida. Se pueden proporcionar reactivos en el actuador de gotitas en el espacio de operaciones de gotitas o en un depósito acoplado de forma fluida al espacio de operaciones de gotitas. Los reactivos pueden estar en forma líquida, p. ej., gotitas, o pueden proporcionarse en una forma reconstituible en el espacio de operaciones de gotitas o en un depósito acoplado de forma fluida al espacio de operaciones de gotitas. Reactivos reconstituibles pueden combinarse típicamente con líquidos para la reconstitución. Un ejemplo de reactivos reconstituibles adecuados para usar con los métodos y el aparato expuestos en el presente documento incluye los descritos en Meathrel et al., patente de EE.UU. N° 7.727.466, titulada "Disintegrable Films for Diagnostic Devices", expedida el 1 de junio de 2010.

"Operación de gotitas" significa cualquier manipulación de una gotita en un actuador de gotitas. Una operación de gotitas puede, por ejemplo, incluir: cargar una gotita en el actuador de gotitas; dispensar una o más gotitas de una gotita fuente; disociar, separar o dividir una gotita en dos o más gotitas; transportar una gotita de un lugar a otro en cualquier dirección; fusionar o combinar dos o más gotitas en una gotita individual; diluir una gotita; mezclar una gotita; agitar una gotita; deformar una gotita; retener una gotita en posición; incubar una gotita; calentar una gotita; vaporizar una gotita; enfriar una gotita; desechar una gotita; transportar una gotita fuera de un actuador de gotitas; otras operaciones de gotitas descritas en el presente documento; y/o cualquier combinación de lo anterior. Los términos "fusionar", "fusionando", "combinar", "combinando" y similares se usan para describir la creación de una gotita a partir de dos o más gotitas. Debe entenderse que cuando dicho término se usa en referencia a dos o más gotitas, puede usarse cualquier combinación de operaciones de gotitas que sea suficiente para dar como resultado la combinación de las dos o más gotitas en una gotita. Por ejemplo, "fusionar la gotita A con la gotita B" se puede lograr transportando la gotita A en contacto con una gotita estacionaria B, transportando la gotita B en contacto con una gotita estacionaria A o transportando las gotitas A y B en contacto entre sí. Los términos "dividir", "separar" y "dividir" no pretenden implicar ningún resultado particular con respecto al volumen de las gotitas resultantes (es decir, el volumen de las gotitas resultantes puede ser el mismo o diferente) o el número de gotitas resultantes (el número de gotitas resultantes puede ser 2, 3, 4, 5 o más). El término "mezclar" se refiere a operaciones de gotitas que dan como resultado una distribución más homogénea de uno o más componentes dentro de una gotita. Ejemplos de operaciones de "carga" de gotitas incluyen carga de microdiálisis, carga asistida por presión, carga robótica, carga pasiva y carga de pipeta. Las operaciones de gotitas pueden estar mediadas por electrodos. En algunos casos, las operaciones de gotitas se ven facilitadas aún más por el uso de regiones hidrófilas y/o hidrófobas en las superficies y/o por obstáculos físicos. Para ejemplos de operaciones de gotitas, véanse las patentes y solicitudes de patentes citadas anteriormente bajo la definición de "actuador de gotitas". Las técnicas de detección de imágenes o impedancia o capacitancia a veces se pueden usar para determinar o confirmar el resultado de una operación de gotita. Ejemplos de tales técnicas se describen en Sturmer et al., Publicación de patente de EE.UU. N° 20100194408, titulada "Capacitance Detection in a Droplet Actuator", publicada el 5 de agosto de 2010.

En términos generales, las técnicas de detección o de formación de imágenes pueden usarse para confirmar la presencia o ausencia de una gotita en un electrodo específico. Por ejemplo, la presencia de una gotita distribuida en el electrodo de destino después de una operación de distribución de gotitas confirma que la operación de distribución de gotitas fue efectiva. De manera similar, la presencia de una gotita en un punto de detección en una etapa apropiada en un protocolo de ensayo puede confirmar que un conjunto previo de operaciones de gotitas ha producido con éxito una gotita para la detección. El tiempo de transporte de gotitas puede ser bastante rápido. Por ejemplo, en diversas realizaciones, el transporte de una gotita de un electrodo al siguiente puede exceder aproximadamente 1 s, o aproximadamente 0,1 s, o aproximadamente 0,01 s, o aproximadamente 0,001 s. En una realización, el electrodo se hace funcionar en modo de CA, pero se cambia al modo CC para la formación de imágenes. Es útil para llevar a cabo operaciones de gotitas para que el área de huella de la gotita sea similar al área de humectación eléctrica; en otras palabras, las gotitas 1x, 2x 3x se hacen funcionar controladas de manera útil utilizando 1, 2 y 3 electrodos, respectivamente. Si la huella de la gotita es mayor que el número de electrodos disponibles para realizar una operación de gotitas en un momento dado, la diferencia entre el tamaño de la gotita y el número de electrodos no debería ser mayor que 1; en otras palabras, una gotita 2x se controla útilmente usando 1 electrodo y una gotita 3x se controla útilmente usando 2 electrodos. Cuando las gotitas incluyen perlas, es útil que el tamaño de la gotita sea igual al número de electrodos que controlan la gotita, p. ej., transportando la gotita.

En algunos aspectos, se puede preparar una colección de ácidos nucleicos a partir de células o de una célula individual usando CEs tales como gotitas. En algunas realizaciones, las células pueden suspenderse en un tampón.

En algunas realizaciones, la suspensión de células puede introducirse en un actuador de gotitas. Usando operaciones de gotitas mediadas por electrodos, se puede dispensar una serie de gotitas que comprenden suspensión de células de tal manera que cada una de las gotitas comprenda una célula individual. Usando operaciones de gotitas mediadas por electrodos, puede dispensarse una serie de gotitas de reactivos que comprenden tampón de lisis celular (gotitas de tampón de lisis). Las gotitas de tampón de lisis y la serie de gotitas de suspensión de células que comprenden células individuales se pueden combinar usando operaciones mediadas por electrodos para formar una gotita de lisado celular de modo que la gotita de lisado celular comprenda componentes de las células individuales. Los reactivos de reacción que comprenden códigos de barras de ácido nucleico únicos, transposones y enzimas adecuadas (p. ej., fragmentasas, polimerasas, ligasas, transposasas, transcriptasas inversas, etc.) pueden introducirse en un actuador de gotitas. En algunas realizaciones, los transposones y/o los códigos de barras pueden comprender sitios de unión de cebadores. Usando operaciones de gotitas mediadas por electrodos, se puede dispensar una serie de gotitas de reactivos que comprenden reactivos de reacción de modo que cada una de las gotitas de reactivos comprenda códigos de barras de ácido nucleico únicos y enzimas adecuadas. Las gotitas de lisado celular y las gotitas de reactivo se pueden combinar usando operaciones mediadas por electrodos para formar una matriz de la primera gotita con código de barras en la que las enzimas de las gotitas de reactivo actúan sobre el ácido nucleico de una célula individual de modo que los ácidos nucleicos comprendan un código de barras. En algunas realizaciones, el ARNm dentro de las gotitas de lisado celular se puede transcribir inversamente cuando las gotitas de lisado celular y las gotitas de reactivo se combinan y el ADNc puede comprender códigos de barras. En algunas realizaciones, los códigos de barras pueden comprender sitios de unión de cebadores e índices moleculares únicos. Usando operaciones de gotitas mediadas por electrodos, la primera gotita con código de barras puede combinarse adicionalmente con gotitas de reactivos para generar series de segundas gotitas de códigos de barras, terceras gotitas de códigos de barras, etc. En algunas realizaciones, para cada una de las rondas de combinación, los códigos de barras son diferentes. Por lo tanto, múltiples rondas de combinación de las gotitas de código de barras con gotitas de reactivo generarán códigos de barras combinatorios. Al final, el ácido nucleico de las diferentes gotitas se puede agrupar y secuenciar. La información de secuenciación puede revelar información de secuenciación del ácido nucleico de la célula y, opcionalmente, también identificar la fuente de los ácidos nucleicos (p. ej., células o una célula individual). Dicha información es valiosa si el ácido nucleico comprende una mutación asociada con una enfermedad tal como la enfermedad genética heredada o el cáncer.

En algunos aspectos, los métodos de la presente solicitud pueden emplearse para la proteómica. Se puede hacer una serie de gotitas que contienen perlas introduciendo la suspensión de perlas en un actuador de gotitas para dispensar un conjunto de gotitas de la suspensión de perlas de modo que cada una de las gotitas en la serie de gotitas comprenda una perla individual (véase la Publicación de patente de EE.UU. 20100130369.). Las perlas pueden comprender anticuerpos u otras sondas de afinidad (véase Immobilized Biomolecules in Analysis. A Practical Approach. Cass T, Ligler F S, eds. Oxford University Press, Nueva York, 1998, págs. 1-14.). En algunas realizaciones, los anticuerpos pueden ser específicos para epítopos de la superficie celular. En algunas realizaciones, los anticuerpos pueden ser monoclonales y en otras realizaciones, los anticuerpos pueden ser policlonales. Usando operaciones de gotitas mediadas por electrodos, una serie de gotitas de suspensión de perlas se puede combinar con una serie de gotitas que comprende células individuales para producir una serie de gotitas de perlas de tal manera que los anticuerpos en las perlas se unen a las proteínas de la superficie celular. En algunas realizaciones, los anticuerpos pueden ser específicos para proteínas dentro de una célula. Usando operaciones de gotitas mediadas por electrodos, se puede combinar una serie de gotitas de suspensión de perlas con una serie de gotitas que comprenden lisados de células individuales de modo que los anticuerpos en las perlas se unan a las proteínas dentro de una célula para producir una serie de proteínas en las gotitas de perlas. Opcionalmente, usando operaciones de gotitas mediadas por electrodos, la serie de proteínas en las gotitas de perlas se puede combinar con una serie de gotitas de reactivos que comprenden reactivos de marcaje de proteínas de modo que las proteínas se puedan marcar de manera única. Las proteínas unidas pueden detectarse a partir de los marcadores asociados o por otros medios (electroforesis en gel de SDS-poliacrilamida, ELISA, etc.). Se puede determinar la identidad de la proteína y la fuente de la proteína. En algunas realizaciones, los datos proteómicos pueden correlacionarse con datos de secuenciación.

En algunas realizaciones, los anticuerpos pueden ser específicos para otras biomoléculas y no limitarse a una proteína. Biomoléculas de este tipo pueden incluir, pero no se limitan a polisacáridos o lípidos. En algunas realizaciones, la identidad y la fuente de biomoléculas de este tipo pueden correlacionarse con los datos de secuencia generados anteriormente.

#### 55 Análisis celular *in situ*

En algunas realizaciones, las células y sus componentes pueden analizarse *in situ*. En algunas realizaciones, se puede permitir que las células pasen a través de una celda de flujo.

Como se usa en este documento, la expresión "celda de flujo" pretende significar una cámara que tiene una superficie a través de la cual pueden fluir uno o más reactivos fluidos. Generalmente, una celda de flujo tendrá una abertura de entrada y una abertura de salida para facilitar el flujo de fluido. Se describen ejemplos de células de flujo y sistemas fluidicos relacionados y plataformas de detección que pueden usarse fácilmente en los métodos de la presente descripción, por ejemplo, en Bentley et al., *Nature* 456:53-59 (2008), documentos 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.

En algunas realizaciones, las celdas de flujo pueden alojar matrices. Las matrices utilizadas para la secuenciación de ácidos nucleicos tienen a menudo patrones espaciales aleatorios de características de ácido nucleico. Por ejemplo, las plataformas de secuenciación HiSeq™ o MiSeq™ disponibles de Illumina Inc. (San Diego, CA) utilizan celda de flujo sobre las cuales se forman matrices de ácidos nucleicos mediante siembra aleatoria seguida de amplificación en puente. Sin embargo, las matrices estampadas también se pueden usar para la secuenciación de ácidos nucleicos u otras aplicaciones analíticas. Ejemplos de matrices estampadas, métodos para su fabricación y métodos para su uso se exponen en los documentos de EE.UU. N.º. de serie 13/787.396; EE.UU. N.º. de serie 13/783.043; de EE.UU. N.º. de serie 13/784.368; Publicación de solicitud de patente de EE.UU. 2013/0116153 A1; y Publicación de solicitud de patente N.º. 2012/0316086 A1. Las características de tales matrices estampadas se pueden usar para capturar una molécula molde de ácido nucleico individual para sembrar la formación posterior de una colonia homogénea, por ejemplo, mediante amplificación de puente. Matrices estampadas de este tipo son particularmente útiles para aplicaciones de secuenciación de ácido nucleico.

En algunas realizaciones, la superficie de la celda de flujo puede comprender restos de captura tales como anticuerpos para inmovilizar las células que la atraviesan en la superficie de la celda de flujo. En algunas realizaciones, los anticuerpos en la superficie de la celda de flujo pueden unirse específicamente a las proteínas de la superficie celular. En algunas realizaciones, los anticuerpos pueden unirse específicamente a las proteínas de la superficie celular de las células cancerosas, enriqueciendo así las células cancerosas en la superficie de la celda de flujo.

En algunas realizaciones, las células pueden clasificarse en diversos tipos mediante tecnología de clasificación celular conocida en la técnica antes de pasar las células a la celda de flujo. Tecnología ilustrativa de clasificación celular incluye, pero no se limita a clasificación celular activada por fluorescencia, o FACS que utiliza citometría de flujo, clasificación celular activada magnéticamente (MACS) (Miltenyi Biotec Inc., San Diego, CA), o mediante la técnica de separación celular libre de columnas en la que se coloca un tubo de células marcadas dentro de un campo magnético. Las células seleccionadas positivamente quedan retenidas en el tubo, mientras que las células seleccionadas negativamente están en la suspensión líquida (STEMCELL Technologies Inc., Vancouver, BC, Canadá).

En algunas realizaciones, las células que pasan a través de la celda de flujo pueden lisarse dentro de la celda de flujo y liberar así el ácido nucleico de las células (ADN y ARN) en la celda de flujo. En algunas realizaciones, las células se inmovilizan en la celda de flujo antes de la lisis. Métodos de lisis celular son conocidos en la técnica e incluyen, pero no se limitan a tratamiento con ultrasonidos, tratamiento con proteasa, por choque osmótico, tratamiento con alto contenido de sal. En algunas realizaciones, todo el ARN se puede transcribir inversamente. En algunas realizaciones, se pueden introducir códigos de barras únicos en el ácido nucleico de las células, por ejemplo, el ADN, ARN o ADNc. Métodos para introducir códigos de barras en un ácido nucleico se analizan anteriormente e incluyen, pero no se limitan a la tagmentación usando tecnología Nextera™, ligasas, polimerasas. En algunas realizaciones, los códigos de barras pueden ser útiles para la identificación de la fuente celular. En algunas realizaciones, los códigos de barras pueden tener sitios de unión de cebadores. En algunas realizaciones, se pueden introducir múltiples códigos de barras en el ácido nucleico. En algunas realizaciones, los múltiples códigos de barras son diferentes entre sí. En algunas realizaciones, el ácido nucleico con códigos de barras puede difundirse; agrupados nuevamente y se pueden introducir códigos de barras adicionales. En algunas realizaciones, después o durante la introducción de los códigos de barras, el ácido nucleico puede fragmentarse. En algunas realizaciones, el ácido nucleico fragmentado puede amplificarse antes de difundirse en la celda de flujo. En algunas realizaciones, el ácido nucleico fragmentado que comprende los códigos de barras puede difundirse a una parte diferente de la celda de flujo que comprende sondas de captura e inmovilizarse en la celda de flujo. En algunas realizaciones, el ácido nucleico fragmentado inmovilizado puede someterse a amplificación en puente.

En algunas realizaciones de los aspectos anteriores, la célula que pasa a través de la celda de flujo es una célula individual. En algunas realizaciones, se puede evaluar todo el transcriptoma. En algunas realizaciones, el ADN y el ARN de las células o de una célula individual pueden evaluarse simultáneamente en cuanto a la información de la secuencia. En algunas realizaciones, las proteínas de las células o de una célula individual pueden evaluarse en cuanto a la identidad y en cuanto a la información de la secuencia. En algunas realizaciones, se pueden evaluar otros analitos de células o de una célula individual tales como lípidos, carbohidratos, organelas celulares.

#### Fragmentación de ácidos nucleicos molde

Algunas realizaciones de la preparación de un ácido nucleico molde pueden incluir la fragmentación de un ácido nucleico diana. En algunas realizaciones, adaptadores indexados o con código de barras están unidos al ácido nucleico diana fragmentado. Los adaptadores se pueden unir usando cualquier número de métodos bien conocidos en la técnica, tales como ligadura (enzimática o química), tagmentación, extensión de polimerasa, etcétera. En algunas realizaciones, la inserción de transposomas que comprenden secuencias de transposón no contiguas puede dar como resultado la fragmentación de un ácido nucleico diana. En algunas realizaciones que comprenden transposomas en bucle, un ácido nucleico diana que comprende secuencias de transposón puede fragmentarse en

los sitios de fragmentación de las secuencias de transposón. Otros ejemplos de métodos útiles para fragmentar ácidos nucleicos diana útiles con las realizaciones proporcionadas en el presente documento se pueden encontrar, por ejemplo, en la solicitud de patente de EE.UU. N° de publicación 2012/0208705, solicitud de patente de EE.UU. N° de publicación 2012/0208724 y solicitud de patente internacional N° de publicación WO 2012/061832.

## 5 Etiquetado de moléculas individuales

También se describen métodos para etiquetar moléculas de modo que las moléculas individuales puedan ser rastreadas e identificadas. Los datos a granel se pueden luego desconvolucionar y convertir de nuevo en la molécula individual. La capacidad de distinguir moléculas individuales y relacionar la información con la molécula de origen es especialmente importante cuando los procesos de la molécula original para dar el producto final cambian la representación (estequiométrica) de la población original. Por ejemplo, la amplificación conduce a la duplicación (*p. ej.*, duplicados de PCR o amplificación sesgada) que puede sesgar la representación original. Esto puede alterar la llamada del estado de metilación, el número de copias, la relación alélica debido a la amplificación no uniforme y/o al sesgo de amplificación. Al identificar moléculas individuales, el etiquetado del código distingue entre moléculas idénticas después del procesamiento. Como tal, las duplicaciones y el sesgo de amplificación se pueden filtrar, lo que permite la determinación precisa de la representación original de una molécula o población de moléculas.

Una ventaja de etiquetar de manera única moléculas individuales es que las moléculas idénticas en el conjunto original se identifican de manera única en virtud de su etiquetado. En otros análisis posteriores, ahora se pueden distinguir estas moléculas marcadas de manera única. Esta técnica puede explotarse en esquemas de ensayo en los que se emplea la amplificación. Por ejemplo, se sabe que la amplificación distorsiona la representación original de una población mixta de moléculas. Si no se utilizara un etiquetado único, la representación original (como el número de copias o la relación alélica) debería tener en cuenta los sesgos (conocidos o desconocidos) para cada una de las moléculas en la representación. Con un etiquetado único, la representación se puede determinar con precisión eliminando duplicados y contando la representación original de las moléculas, cada una con una etiqueta única. Por lo tanto, los ADNc se pueden amplificar y secuenciar, sin temor a sesgos, ya que los datos se pueden filtrar de modo que solo se seleccionen secuencias auténticas o secuencias de interés para su posterior análisis. Se pueden construir lecturas precisas tomando el consenso en muchas lecturas con el mismo código de barras.

En algunas realizaciones de las composiciones y los métodos descritos en el presente documento, se prefiere etiquetar la población original en las primeras fases del ensayo, aunque el etiquetado puede ocurrir en fases posteriores si las etapas anteriores no introducen sesgo o no son importantes. En cualquiera de estas aplicaciones, la complejidad de las secuencias de códigos de barras debe ser mayor que la cantidad de moléculas individuales a etiquetar. Esto asegura que diferentes moléculas diana reciban etiquetas diferentes y únicas. Como tal, es deseable un conjunto de oligonucleótidos aleatorios de cierta longitud (*p. ej.*, 5, 10, 20, 30, 40, 50, 100 o 200 nucleótidos de longitud). Una agrupación aleatoria de etiquetas representa una gran complejidad de etiquetas con espacio de código de 4<sup>n</sup>, en donde *n* es el número de nucleótidos. Códigos adicionales (ya sean diseñados o aleatorios) se pueden incorporar en diferentes fases para servir como una verificación adicional, como una verificación de paridad para corrección de errores.

En una realización de las composiciones y los métodos descritos en el presente documento, moléculas individuales (tal como ADN diana) se unen a marcadores únicos, tales como secuencias de oligo únicas y/o códigos de barras. La fijación de los marcadores puede producirse mediante ligadura, química de acoplamiento, adsorción, inserción de secuencias de transposones, etc. Otros medios incluyen amplificación (tal como por PCR, RCA o LCR), copiado (tal como la adición por una polimerasa) e interacciones no covalentes.

Métodos específicos comprenden la inclusión de códigos de barras (*p. ej.*, secuencias diseñadas o aleatorias) en cebadores de PCR de modo que cada uno de los moldes recibirá un código individual dentro del espacio de código, proporcionando con ello amplicones únicos que pueden ser discriminados de otros amplicones. Este concepto se puede aplicar a cualquier método que use amplificación de polimerasa, tales como los ensayos GoldenGate™ y los ensayos descritos en las Patentes de EE.UU. N°. 7.582.420, N°. 7.955.794 y N°. 8.003.354. Secuencias diana etiquetadas con código pueden circularizarse y amplificarse mediante métodos tales como la amplificación de círculo rodante para proporcionar amplicones etiquetados con código. De manera similar, el código también se puede agregar a ARN

## 50 Métodos de análisis de ácidos nucleicos molde

Algunas realizaciones de la tecnología descrita en el presente documento incluyen métodos para analizar ácidos nucleicos molde. En tales realizaciones, la información de secuenciación puede obtenerse a partir de ácidos nucleicos molde y esta información puede usarse para generar una representación de secuencia de uno o más ácidos nucleicos diana.

En algunas realizaciones de los métodos de secuenciación descritos en el presente documento, se puede usar una estrategia de lectura enlazada. Una estrategia de lectura enlazada puede incluir la identificación de datos de secuencia que enlaza al menos dos lecturas de secuenciación. Por ejemplo, una primera lectura de secuenciación puede contener un primer marcador, y una segunda lectura de secuenciación puede contener un segundo marcador.

Los marcadores primero y segundo pueden identificar los datos de secuenciación de cada una de las lecturas de secuenciación para que sean adyacentes en una representación de secuencia del ácido nucleico diana. En algunas realizaciones de las composiciones y los métodos descritos en el presente documento, los marcadores pueden comprender una primera secuencia de código de barras y una segunda secuencia de código de barras, en que la primera secuencia de código de barras puede emparejarse con la segunda secuencia de código de barras. En otras realizaciones, los marcadores pueden comprender una primera etiqueta de huésped y una segunda etiqueta de huésped. En más realizaciones, los marcadores pueden comprender una primera secuencia de código de barras con una primera etiqueta de huésped, y una segunda secuencia de código de barras con una segunda etiqueta de huésped.

Una realización ilustrativa de un método para secuenciar un ácido nucleico molde puede comprender las siguientes etapas: (a) secuenciar la primera secuencia de código de barras usando un cebador de secuenciación que se hibrida con el primer sitio de cebador; y (b) secuenciar la segunda secuencia de código de barras usando un cebador de secuenciación que se hibrida con el segundo cebador. El resultado son dos lecturas de secuencia que ayudan a enlazar el ácido nucleico molde con sus vecinos genómicos. Dadas las lecturas lo suficientemente largas y los fragmentos de la colección lo suficientemente cortos, estas dos lecturas se pueden fusionar de manera informal para hacer una lectura larga que cubra todo el fragmento. Usando las lecturas de secuencia de código de barras y la secuencia duplicada de 9 nucleótidos presente desde la inserción, las lecturas ahora se pueden enlazar a sus vecinos genómicos para formar "lecturas enlazadas" mucho más largas *in silico*.

Como se entenderá, una colección que comprende ácidos nucleicos molde puede incluir fragmentos de ácido nucleico duplicados. La secuenciación de fragmentos duplicados de ácido nucleico es ventajosa en métodos que incluyen la creación de una secuencia consenso para fragmentos duplicados. Métodos de este tipo pueden aumentar la precisión para proporcionar una secuencia consenso para un ácido nucleico molde y/o una colección de ácidos nucleicos molde.

En algunas realizaciones de la tecnología de secuenciación descrita en el presente documento, el análisis de la secuencia se realiza en tiempo real. Por ejemplo, la secuenciación en tiempo real se puede realizar adquiriendo y analizando simultáneamente datos de la secuenciación. En algunas realizaciones, un proceso de secuenciación para obtener datos de la secuenciación puede terminarse en diversos puntos, incluyendo después de que se obtenga al menos una parte de los datos de secuencia de ácido nucleico diana o antes de que se secuencie la lectura completa de ácido nucleico. Métodos, sistemas y realizaciones adicionales ilustrativos se proporcionan en la Publicación de Patente Internacional N° WO 2010/062913.

En una realización ilustrativa de un método para ensamblar lecturas de secuenciación cortas usando una estrategia de lectura enlazada, las secuencias de transposón que comprenden códigos de barras se insertan en el ADN genómico, se prepara una colección y se obtienen datos de secuenciación para la colección de ácidos nucleicos molde. Bloques de molde se pueden ensamblar identificando códigos de barras emparejados y luego se ensamblan contigios más grandes. En una realización, las lecturas ensambladas pueden ensamblarse adicionalmente en contigios más grandes a través del emparejamiento de código usando lecturas superpuestas.

Algunas realizaciones de la tecnología de secuenciación descrita en el presente documento incluyen características de detección y corrección de errores. Ejemplos de errores pueden incluir errores en llamadas base durante un proceso de secuenciación y errores en el ensamblaje de fragmentos en contigios más grandes. Como se entendería, la detección de errores puede incluir detectar la presencia o probabilidad de errores en un conjunto de datos y, como tal, detectar la ubicación de un error o el número de errores puede no ser necesario. Para la corrección de errores, es útil la información sobre la ubicación de un error y/o el número de errores en un conjunto de datos. Métodos para la corrección de errores son bien conocidos en la técnica. Ejemplos incluyen el uso de distancias de Hamming y el uso de un algoritmo de suma de verificación (Véase, p. ej., la Publicación de la solicitud de patente de EE.UU. N° 2010/0323348; la Patente de EE.UU. N° 7.574.305; y la patente de EE.UU. N° 6.654.696).

#### Colecciones anidadas

Un método alternativo implica los métodos de etiquetado de unión anteriores y la preparación de colecciones de secuenciación anidadas. Las sub-colecciones anidadas se crean a partir de fragmentos de ADN etiquetados con código. Esto puede permitir el marcaje de transposones menos frecuentes en todo el genoma. También puede crear una mayor diversidad de lecturas de secuenciación (anidadas). Estos factores pueden conducir a una mejor cobertura y precisión.

El sub-muestreo y la amplificación del genoma completo pueden crear muchas copias de una determinada población de moléculas iniciales. Luego, los fragmentos de ADN se generan por fragmentación específica para el transposón, en que cada uno de los fragmentos recibe un código que le permite enlazar el fragmento de nuevo al vecino original que tiene un código coincidente (ya sea idéntico, complementario o enlazado de otro modo de manera informática). Los fragmentos etiquetados se fragmentan al menos una segunda vez mediante métodos aleatorios o métodos específicos para la secuencia, tales como la digestión enzimática, el cizallamiento aleatorio, el cizallamiento basado en transposones u otros métodos, creando así sub-colecciones de los fragmentos de ADN etiquetados con código. En una variación útil del método descrito previamente, los fragmentos etiquetados con código pueden aislarse

preferentemente usando transposones que contienen una biotina u otra funcionalidad de afinidad para fines de enriquecimiento aguas abajo. La posterior preparación de la colección convierte los fragmentos de ADN anidados en moldes de secuenciación. La secuenciación de extremos emparejados da como resultado la determinación de la secuencia de la etiqueta con código de los fragmentos de ADN y del ADN diana. Dado que se crean colecciones anidadas para la misma etiqueta con código, se pueden secuenciar fragmentos de ADN largos con lecturas cortas.

#### Métodos de secuenciación

Los métodos y la composición descritos en el presente documento pueden usarse junto con una diversidad de técnicas de secuenciación. En algunas realizaciones, el proceso para determinar la secuencia de nucleótidos de un ácido nucleico diana puede ser un proceso automatizado.

Algunas realizaciones de los métodos de secuenciación descritos en el presente documento incluyen secuenciación mediante tecnologías de síntesis (SBS), por ejemplo, técnicas de pirosecuenciación. La pirosecuenciación detecta la liberación de pirofosfato inorgánico (PPi) a medida que se incorporan nucleótidos particulares en la cadena naciente (Ronaghi *et al.*, *Analytical Biochemistry* 242(1):84-9 (1996); Ronaghi, M. *Genome Res.* 11(1):3-11 (2001); Ronaghi *et al.*, *Science* 281(5375):363 (1998); patente de EE.UU. N° 6.210.891; patente de EE.UU. N° 6.258.568 y patente de EE.UU. N° 6.274.320).

En otro tipo de ejemplo 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 tal como se describe, por ejemplo, en la patente de EE.UU. N° 7.427.67, la patente de EE.UU. N° 7.414.1163 y la patente de EE.UU. N° 7.057.026. Este enfoque, que está siendo comercializado por Illumina Inc., también se describe en las Publicaciones de solicitud de patente internacional N° WO 91/06678 y WO 07/123744. La disponibilidad de terminadores marcados con fluorescencia, en los que tanto la terminación se puede revertir como el marcador fluorescente se puede escindir, facilita la secuenciación eficiente de terminación reversible cíclica (CRT). Las polimerasas también pueden ser co-modificadas para incorporar y extenderse eficientemente a partir de estos nucleótidos modificados.

Sistemas y métodos de SBS ilustrativos adicionales que se pueden utilizar con los métodos y las composiciones descritos en el presente documento se describen en la Publicación de solicitud de patente de EE.UU. N° 2007/0166705, Publicación de la solicitud de patente de EE.UU. N° 2006/0188901, Patente de EE.UU. N° 7057026, Publicación de solicitud de patente de EE.UU. N° 2006/0240439, Publicación de solicitud de patente de EE.UU. N° 2006/0281109, publicación PCT N° WO 05/065814, Publicación de solicitud de patente de EE.UU. N° 2005/0100900, publicación PCT N° WO 06/064199 y Publicación PCT N° WO 07/010251.

Algunas realizaciones de la tecnología de secuenciación descrita en el presente documento pueden utilizar la secuenciación mediante técnicas de ligadura. Dichas técnicas utilizan ADN ligasa para incorporar nucleótidos e identificar la incorporación de dichos nucleótidos. Los sistemas y métodos de SBS ilustrativos que se pueden utilizar con las composiciones y los métodos descritos en el presente documento se describen en la patente de EE.UU. N° 6.969.488, la patente de EE.UU. N° 6.172.218 y la patente de EE.UU. N° 6.306.597.

Los métodos de secuenciación descritos en el presente documento pueden llevarse a cabo ventajosamente en formatos multiplex de manera que se manipulen simultáneamente múltiples ácidos nucleicos diana diferentes. En realizaciones particulares, se pueden tratar diferentes ácidos nucleicos diana en un recipiente de reacción común o en una superficie de un sustrato particular. Esto permite el suministro conveniente de reactivos de secuenciación, la separación de reactivos que no han reaccionado y la detección de eventos de incorporación de manera múltiple. En realizaciones que usan ácidos nucleicos diana unidos a la superficie, los ácidos nucleicos diana pueden estar en un formato de matriz. En un formato de matriz, los ácidos nucleicos diana se pueden acoplar típicamente a una superficie de una manera espacialmente distinguible. Por ejemplo, los ácidos nucleicos diana pueden unirse mediante fijación covalente directa, fijación a una perla u otra partícula o asociarse con una polimerasa u otra molécula que está fijada a la superficie. La matriz puede incluir una copia única de un ácido nucleico diana en cada sitio (a la que también se alude como característica) o múltiples copias que tienen la misma secuencia pueden estar presentes en cada uno de los sitios o características. Se pueden producir múltiples copias mediante métodos de amplificación tales como amplificación en puente o PCR en emulsión tal como se describe con más detalle en el presente documento.

Los métodos establecidos en el presente documento pueden usar matrices que tienen características en cualquiera de una diversidad de densidades que incluyen, por ejemplo, al menos aproximadamente 10 características/cm<sup>2</sup>, 100 características/cm<sup>2</sup>, 500 características/cm<sup>2</sup>, 1.000 características/cm<sup>2</sup>, 5.000 características/cm<sup>2</sup>, 10.000 características/cm<sup>2</sup>, 50.000 características/cm<sup>2</sup>, 100.000 características/cm<sup>2</sup>, 1.000.000 características/cm<sup>2</sup>, 5.000.000 características/cm<sup>2</sup>, 10<sup>7</sup> características/cm<sup>2</sup>, 5 x 10<sup>7</sup> características/cm<sup>2</sup>, 10<sup>8</sup> características/cm<sup>2</sup>, 5 x 10<sup>8</sup> características/cm<sup>2</sup>, 10<sup>9</sup> características/cm<sup>2</sup>, 5x 10<sup>9</sup> características/cm<sup>2</sup>, o mayores.

#### Métodos para reducir las tasas de error en la secuenciación de datos

Algunas realizaciones de los métodos y las composiciones proporcionadas en el presente documento incluyen la reducción de las tasas de error en la secuenciación de datos. En algunas de tales realizaciones, las cadenas sentido

y antisentido de un ácido nucleico bicatenario diana están asociadas cada una con un código de barras diferente. Cada una de las cadenas se amplifica, la información de la secuencia se obtiene de múltiples copias de las cadenas amplificadas y se genera una representación de secuencia consenso del ácido nucleico diana a partir de la información de la secuencia redundante. Por lo tanto, la información de la secuencia puede originarse e identificarse a partir de cada una de las cadenas. En consecuencia, los errores de secuencia pueden identificarse y reducirse cuando la información de la secuencia que se origina en una cadena es inconsistente con la información de la secuencia de la otra cadena.

En algunas realizaciones, las cadenas sentido y antisentido de un ácido nucleico diana están asociadas con un código de barras diferente. Los códigos de barras pueden estar asociados con el ácido nucleico diana mediante una diversidad de métodos que incluyen la ligadura de adaptadores y la inserción de secuencias de transposones. En algunas de tales realizaciones, un adaptador Y se puede ligar a al menos un extremo de un ácido nucleico diana. El adaptador en Y puede incluir una secuencia bicatenaria y cadenas no complementarias, comprendiendo cada una de las cadenas un código de barras diferente. El ácido nucleico diana con el adaptador Y ligado puede amplificarse y secuenciarse de modo que cada uno de los códigos de barras pueda usarse para identificar las cadenas sentido o antisentido originales. Un método similar se describe en Kinde I. et al., (2011) PNAS 108: 9530-9535. En algunas realizaciones, las cadenas sentido y antisentido de un ácido nucleico diana se asocian con un código de barras diferente insertando secuencias de transposón proporcionadas en el presente documento. En algunas de estas realizaciones, las secuencias de transposón pueden comprender códigos de barras no complementarios.

Algunas realizaciones de métodos de este tipo incluyen obtener información de la secuencia de una cadena de un ácido nucleico bicatenario diana, que comprende (a) obtener datos de la secuencia de un ácido nucleico molde que comprende un primer adaptador de secuenciación y un segundo adaptador de secuenciación que tiene al menos una porción del ácido nucleico diana bicatenario dispuesto entre ellos, en donde: (i) el primer adaptador de secuenciación comprende un primer código de barras bicatenario, un primer sitio de cebador monocatenario y un segundo sitio de cebador monocatenario, en donde los sitios de cebador primero y segundo no son complementarios, y (ii) el segundo adaptador de la secuenciación que comprende un segundo código de barras bicatenario, un tercer sitio de cebador monocatenario y un cuarto sitio de cebador monocatenario, en el que los sitios de cebador tercero y cuarto no son complementarios. En algunas realizaciones, el primer sitio de cebador de la cadena sentido del ácido nucleico molde y el tercer sitio de cebador de la cadena sentido antisentido del ácido nucleico molde comprenden la misma secuencia. En algunas realizaciones, cada uno de los códigos de barras es diferente. En algunas realizaciones, el primer adaptador de la secuenciación comprende una horquilla monocatenaria que acopla el primer sitio de cebador y el segundo sitio de cebador.

En otra realización, cada uno de los extremos de un ácido nucleico diana está asociado con un adaptador que comprende un código de barras diferente de modo que los productos de extensión de la cadena sentido y antisentido de un ácido nucleico pueden distinguirse entre sí. En algunas realizaciones, las secuencias del sitio del cebador y los códigos de barras se seleccionan de tal manera que la extensión desde un cebador reasociado a la cadena sentido proporciona productos que pueden distinguirse de los productos de extensión de un cebador reasociado a la cadena antisentido. En un ejemplo, el sitio de cebador 3' sentido es el mismo que el sitio de cebador antisentido 3', pero diferente de los sitios de cebador sentido 5' y antisentido 5'. La extensión de los cebadores reasociados al sitio de cebador sentido 3' y el sitio de cebador antisentido 3' produciría los siguientes productos de cada una de las cadenas:

Cadena sentido: (5') código de barras 2 - [secuencia diana] - código de barras 1 (3')

Cadena antisentido: (5') código de barras 1 - [secuencia diana] - código de barras 2 (3')

Por lo tanto, los productos de extensión de la cadena sentido y antisentido de un ácido nucleico se pueden distinguir entre sí. Un método a modo de ejemplo se ilustra en Schmitt M.W., *et al.*, PNAS (2012) 109: 14508-13. En algunos de estos métodos, los sitios de códigos de barras y cebadores pueden estar asociados con el ácido nucleico diana mediante una diversidad de métodos que incluyen la ligadura de adaptadores y la inserción de secuencias de transposones. En algunas realizaciones, las secuencias de transposones pueden diseñarse para proporcionar adaptadores con horquillas. Las horquillas proporcionan la capacidad de mantener la contigüidad física de las cadenas sentido y antisentido de un ácido nucleico diana. Se puede preparar un ácido nucleico molde que comprende horquillas usando secuencias de transposones que comprenden los enlazadores descritos en el presente documento. Ejemplos de enlazadores incluyen ácidos nucleicos monocatenarios.

Algunas realizaciones de la preparación de una colección de ácidos nucleicos molde para obtener información de la secuencia de cada una de las cadenas de un ácido nucleico diana bicatenario incluyen (a) proporcionar una población de transposomas que comprende una transposasa y una primera secuencia de transposón que comprende: (i) un primer sitio de reconocimiento de transposasa, un primer sitio de cebador y un primer código de barras, y (ii) una segunda secuencia de transposón que comprende un segundo sitio de reconocimiento de transposasa, un segundo sitio de cebador y un segundo código de barras, en donde la primera secuencia de transposón no es contigua a la segunda secuencia de transposón; y (b) poner en contacto los transposomas con un ácido nucleico bicatenario bajo condiciones tales que dichas secuencias de transposón primera y segunda se insertan en el ácido nucleico diana bicatenario, preparando así una colección de ácidos nucleicos molde para

obtener información de la secuencia de cada una de las cadenas del ácido nucleico diana bicatenario. En algunas realizaciones, la población de transposomas comprende, además, transposomas que comprenden una transposasa y una secuencia de transposón que comprende un tercer sitio de reconocimiento de transposasa y un cuarto sitio de reconocimiento de transposasa que tiene una secuencia de código de barras dispuesta entre ellos, comprendiendo dicha secuencia de código de barras un tercer código de barras y un cuarto código de barras que tiene un adaptador de la secuenciación dispuesto entre ellos, comprendiendo dicho adaptador de la secuenciación un tercer sitio de cebador y un cuarto sitio de cebador que tiene un enlazador dispuesto entre ellos. En algunas realizaciones, el primer sitio de cebador de la cadena sentido del ácido nucleico molde y el tercer sitio de cebador de la cadena sentido antisentido del ácido nucleico molde comprenden la misma secuencia. Algunas realizaciones también incluyen una etapa (c) seleccionar ácidos nucleicos molde que comprenden secuencias de transposón en las que la primera secuencia de transposón no es contigua con la segunda secuencia de transposón y las secuencias de transposón comprenden un enlazador. En algunas realizaciones, el enlazador comprende una etiqueta de afinidad adaptada para unirse con una sonda de captura. En algunas realizaciones, la etiqueta de afinidad se selecciona del grupo que consiste en His, biotina y estreptavidina. En algunas realizaciones, cada uno de los códigos de barras es diferente. En algunas realizaciones, el enlazador comprende un ácido nucleico monocatenario. En algunas realizaciones, el ácido nucleico diana comprende ADN genómico.

#### Métodos para obtener información de haplotipos

Algunas realizaciones de los métodos y las composiciones proporcionadas en el presente documento incluyen métodos para obtener información de haplotipo a partir de un ácido nucleico diana. La información del haplotipo puede incluir la determinación de la presencia o ausencia de diferentes secuencias en loci especificados en un ácido nucleico diana, tal como un genoma. Por ejemplo, se puede obtener información de la secuencia para copias maternas y paternas de un alelo. En un organismo poliploide, se puede obtener información de la secuencia para al menos un haplotipo. Métodos de este tipo también son útiles para reducir la tasa de error en la obtención de información de la secuencia del ácido nucleico diana.

En general, los métodos para obtener información sobre el haplotipo incluyen la distribución de un ácido nucleico en uno o más compartimentos de manera que cada uno de los compartimentos comprenda una cantidad de ácido nucleico equivalente a aproximadamente un haplotipo del ácido nucleico, o equivalente a menos de aproximadamente un haplotipo del ácido nucleico. La información de la secuencia se puede obtener de cada uno de los compartimentos, obteniendo así información de haplotipo. La distribución del ácido nucleico molde en una pluralidad de recipientes aumenta la probabilidad de que un recipiente individual incluya una copia individual de un alelo o SNP, o que la información de secuencia consenso obtenida de un recipiente individual refleje la información de la secuencia de un alelo o SNP. Como se entenderá, en algunas de tales realizaciones, un ácido nucleico molde puede diluirse antes de compartimentar el ácido nucleico molde en una pluralidad de recipientes. Por ejemplo, cada uno de los recipientes puede contener una cantidad de ácidos nucleicos diana igual a aproximadamente un haplotipo equivalente del ácido nucleico diana. En algunas realizaciones, un recipiente puede incluir menos de aproximadamente un haplotipo equivalente de un ácido nucleico diana.

Métodos para determinar la información del haplotipo, el método de haplotipar con compartimentos virtuales, métodos de preparación de ácidos nucleicos diana para el haplotipado se describen en la publicación de la OMPI. WO/2014/142850.

#### 40 Ejemplos

##### Ejemplo 1 Mantenimiento de la contigüidad del molde

Este ejemplo ilustra un método para mantener la información de contigüidad de un ácido nucleico molde dentro de un CE. El ácido nucleico molde se prepara usando transposomas que comprenden secuencias de transposón no contiguas en las que la transposasa Tn5 permanece unida al ADN molde después de la transposición. El ácido nucleico diana se pone en contacto con transposomas que comprenden transposasa Tn5 y secuencias de transposón no contiguas. Las muestras que se tratan adicionalmente con SDS pueden aparecer como una muestra de diversos fragmentos de ácido nucleico molde; muestras no tratadas con SDS pueden mostrar retención de supuesto ácido nucleico molde de alto peso molecular. Por lo tanto, aunque un ácido nucleico puede estar fragmentado, la transposasa aún puede asociar secuencias adyacentes entre sí.

En aún otro método ilustrativo, se prepara una colección de ácidos nucleicos molde utilizando transposomas que comprenden secuencias de transposones no contiguas con ácido nucleico diana que comprende cromosoma humano. El CE comprende el ácido nucleico diana. Se pueden observar bloques de ADN de haplotipos para muestras en las que la transposasa se separa mediante SDS post-dilución. Por lo tanto, al poner en práctica métodos como los descritos en el presente documento, los ácidos nucleicos diana pueden mantener la integridad diana cuando se transponen, se diluyen y se transforman en colecciones de secuenciación.

La expresión "que comprende", tal como se usa en el presente documento, es sinónimo de "que incluye", "que contiene" o "se caracteriza por", e incluye o es abierta y no excluye elementos adicionales no reseñados o etapas del método.

5 Debe entenderse que todos los números que expresan cantidades de ingredientes, condiciones de reacción, etcétera utilizados en la memoria descriptiva están modificados en todos los casos por el término "aproximadamente". En consecuencia, a menos que se indique lo contrario, los parámetros numéricos establecidos en el presente documento son aproximaciones que pueden variar dependiendo de las propiedades deseadas que se desean obtener. Como mínimo, y no como un intento de limitar la solicitud a la doctrina de los equivalentes en el alcance de cualquier reivindicación en cualquier solicitud que reivindique prioridad a la presente solicitud, cada uno de los parámetros numéricos debe interpretarse a la vista del número de dígitos significativos y enfoques de redondeo ordinarios.

**Ejemplo 2- Secuenciación de Transcriptoma Completo de Célula Individual**

10 Este ejemplo describe un método para proporcionar códigos de barras uniformes a lo largo de toda la longitud de un ADNc y el uso de los códigos de barras para determinar la información de contigüidad del ADNc, así como para identificar la fuente celular, es decir, identificar la célula individual asociada con el ARNm.

15 Este ejemplo ilustra un método para secuenciar el transcriptoma de una célula individual. En este ejemplo, se utiliza microfluídica de gotitas para capturar el transcriptoma de múltiples células individuales en perlas de captura individuales y luego se utiliza la transposición para preservar la transposición y la indexación combinatoria (CPT-seq) para codificar con código de barras el ADNc derivado del transcriptoma de cada una de las células individuales. En una realización, el método de la invención usa un proceso de proporcionar múltiples códigos de barras para indexar ADNc de una célula individual en el que se agrega un primer código de barras en una reacción de tagmentación y se agrega un segundo código de barras en una reacción de amplificación por PCR.

20 En un ejemplo, el ARN poli-A + se captura de células individuales y el ARN poli-A + capturado se procesa en masa para la generación de una colección de secuenciación multiplexada.

25 El método puede incluir las siguientes etapas. En la etapa 1, el ARN de una célula individual se captura en una perla de captura. Por ejemplo, múltiples células individuales (p. ej., aproximadamente 1000 células individuales) se encapsulan en gotitas individuales (es decir, en promedio, una célula y una perla por gotita) que comprenden un tampón de lisis y una perla de captura. Inmovilizado en la superficie de la perla de captura hay una pluralidad de sondas de captura que incluyen una secuencia de captura de poli-dT y una secuencia de cebador de PCR. La composición del tampón de lisis de la gotita disocia la membrana citoplasmática de la célula individual, liberando ARN citoplasmático. El ARN poli-A + liberado se captura por hibridación de las secuencias poli-A + en el ARN con las secuencias de captura oligo-dT inmovilizadas en la superficie de la perla de captura co-encapsulada. Cada una de las perlas de captura incluye ahora ARN poli-A + del transcriptoma de una célula individual. Todo el ARN poli-A + de una célula individual se mantiene cerca uno del otro en la perla de captura.

35 En una etapa 2, las perlas de captura con ARN poli-A + de una célula individual se agrupan a partir de múltiples gotitas (p. ej., aproximadamente 1000 perlas de captura) y se sintetiza ADNc bicatenario. Por ejemplo, las perlas de captura se agrupan, se lavan y el ADNc de la primera cadena se sintetiza usando una RNasa H menos transcriptasa inversa que es capaz de cambiar de cadena. Se incluye un cebador de cambio de cadena durante la síntesis de ADNc de la primera cadena que permite la colocación de un sitio de cebador universal en el extremo 3' del ADNc. Luego se prepara ADNc bicatenario usando un cebador universal y una ADN polimerasa de alta fidelidad en una reacción PCR (p. ej., 1 a 2 ciclos de PCR). Cada una de las perlas de captura incluye ahora ADNc transcrito inversamente del ARN poli-A + de una célula individual.

40 En una etapa 3, las perlas de captura con ADNc bicatenario se distribuyen en los pocillos de una placa de 96 pocillos de modo que haya aproximadamente 10 perlas de captura por pocillo.

45 En una etapa 4, el ADNc bicatenario en cada uno de los pocillos se tagmenta usando 96 transposomas indexados de forma única. La tagmentación se usa para modificar el ADNc con secuencias de índice y adaptador al tiempo que se preserva la contigüidad de la célula individual. El ensamblaje de los 96 complejos de transposomas indexados de forma única utilizados en la reacción de tagmentación se describe con más detalle a continuación. La reacción de tagmentación agrega la primera parte de un código de barras bipartito a cada uno de los fragmentos de ADNc futuro. Cada una de las perlas de captura incluye ahora ADNc tagmentado de una célula individual.

50 En una etapa 5, las perlas de captura en todos los pocillos se recogen, se agrupan, se lavan y se redistribuyen en los pocillos de otra placa de 96 pocillos de modo que haya aproximadamente 10 perlas de captura por pocillo. El ARNm/ADNc de una célula individual permanece en la superficie de una perla individual y la transposasa permanece unida al ADNc fragmentado y evita que los fragmentos se disocien.

En una etapa 6, la transposasa y el ADNc tagmentado se liberan de las perlas de captura. Por ejemplo, se agrega una parte alícuota de una solución SDS (SDS al 1%) a cada uno de los pocillos para liberar la transposasa unida y el ADNc tagmentado de las perlas de captura.

55 En una etapa 7, el ADNc etiquetado en cada uno de los pocillos se amplifica usando cebadores de PCR que incluyen una secuencia P5 o P7 y una secuencia de código de barras única. Por ejemplo, una de cada 96 combinaciones únicas de cebadores de PCR P5 y P7 con código de barras se agrega a cada uno de los pocillos y

5 los fragmentos de ADNc tagmentados se amplifican. La reacción de PCR agrega la porción restante del código de barras bipartito a cada uno de los fragmentos de ADNc. Cada uno de los fragmentos de ADNc incluye ahora 4 secuencias de código de barras: dos secuencias agregadas en la reacción de tagmentación y 2 secuencias agregadas durante la amplificación por PCR. Por lo tanto, el ARNm/ADNc de una célula individual se identifica mediante la combinación del índice de tagmentación y el índice de PCR agregado a través de la etapa de amplificación.

En una etapa 8, los fragmentos de ADNc con código de barras de cada uno de los pocillos se agrupan y secuencian.

10 En este ejemplo, la indexación combinatoria de  $96 \times 96$  se usa para codificar con código de barras aproximadamente 1000 células individuales, con aproximadamente un 5% de posibilidades de que dos células tengan los mismos códigos de barras. El rendimiento puede aumentarse fácilmente incrementando el número de "compartimentos". Por ejemplo, mediante el uso de códigos de barras combinatorios  $384 \times 384$  (aproximadamente 147.456 compartimentos virtuales), se pueden crear códigos de barras individuales de aproximadamente 10.000 células en paralelo con aproximadamente un 3% de posibilidades de que dos células tengan el mismo código de barras.

15 Este ejemplo también describe un proceso de ensamblaje de 96 complejos de transposomas con código de barras únicos para agregar la primera parte de un código de barras bipartito en un protocolo de código de barras combinatorio. El proceso incluye, pero no se limita a las siguientes etapas.

20 En la etapa A se forman 20 transposones indexados de manera única mediante la reasociación de oligonucleótidos indexados individuales, cada uno de los cuales contiene la secuencia de Tn5 Mosaic End (ME) en su extremo 3', a un oligonucleótido complementario de ME fosforilado universal 5' (pMENTS). Por ejemplo, un oligonucleótido indexado 1110 que incluye secuencias P5, una secuencia de índice "i5" única de 8 bases, una secuencia de conector universal Universal connector A-C15, y una secuencia de ME se reasocia a una secuencia complementaria de ME 1115. La secuencia complementaria de ME 1115 es un oligonucleótido fosforilado 5' universal (pMENTS) que es complementario a las secuencias ME en el oligonucleótido indexado 1110. La secuencia de conector universal A-C15 se usa más tarde para reasociar el cebador de secuenciación de índice 2 personalizado.

25 Se realiza un segundo conjunto de reacciones de reasociación (es decir, 12 reacciones de reasociación individuales) para formar un segundo conjunto de 12 transposones que incluyen una secuencia de índice "i7" única de 8 bases adyacente a una secuencia P7. Por ejemplo, un oligonucleótido indexado 1120 que incluye secuencias P7, una secuencia única de índice i7 de 8 bases, una secuencia de conector universal B-D15 y una secuencia de ME se reasocian a la secuencia complementaria de ME 1115. La secuencia de conector universal B-D15 se usa más tarde para reasociar el cebador de secuenciación de índice 1 personalizado.

30 En la etapa B, los transposones P5\_i5 1125 reasociados (es decir, 8 transposones P5\_i5 1125 cada uno con una secuencia de índice i5 única de 8 bases) y los transposones de P7\_i7 1130 reasociados (es decir, 12 transposones 1130 cada uno con una secuencia índice i7 única de 8 bases) se ensamblan en reacciones individuales con transposasa Tn5 para formar complejos de transposomas. Por ejemplo, cada uno de los transposones P5\_i5 1125 reasociados se incuba con la transposasa Tn5 1135 a aproximadamente 37°C durante aproximadamente 1 hora para formar un complejo de transposoma P5\_i5 1140. De forma similar, cada transposón P7\_i7 1130 reasociado se incuba con la transposasa Tn5 1135 a aproximadamente 37°C durante aproximadamente 1 hora para formar un complejo de transposoma P7\_i7 1145.

35 En la etapa C se forman 96 complejos de transposomas únicos combinando partes alícuotas de complejos de transposomas P5\_i5 1140 con partes alícuotas de complejos de transposomas P7\_i7 1145. Por ejemplo, los complejos de transposomas P5\_i5 1140 se dividen en partes alícuotas en las filas A a H de una placa de 96 pocillos y complejos de transposomas P7\_i7 1145 se dividen en partes alícuotas en las columnas 1 a 12 de la misma placa de 96 pocillos. La combinación de 8 complejos de transposomas P5\_i5 1140 y 12 complejos de transposomas P7\_i7 1145 crea 96 combinaciones de índice diferentes.

40 Para evaluar los complejos de transposomas ensamblados, se preparó una colección de secuenciación a partir de 10 células individuales usando una reacción de tagmentación única y una reacción PCR única. Diez perlas de captura que comprenden ADNc de 10 células individuales se agruparon y se tagmentaron usando la mezcla de transposomas P5\_i5\_1 más P7\_i7\_1. El ADNc tagmentado se liberó luego de las perlas de captura y se amplificó por PCR usando cebadores P5 y P7 con código de barras para generar una colección de secuenciación. La distribución del tamaño del fragmento en la colección de secuenciación se analizó utilizando un Bioanalizador. En algunas realizaciones, la limpieza se realiza después de la PCR. En algunas realizaciones, la segunda limpieza de SPRI se realiza después de la primera limpieza de SPRI. En algunas realizaciones, la muestra se diluye 10 veces antes de analizarla en un Bioanalizador.

45 En otro ejemplo, se usaron dos mezclas diferentes de complejos de transposomas para preparar una colección de secuenciación a partir de 100 células individuales. En este ejemplo, se utilizó un protocolo de división y agrupación para evaluar los complejos de transposomas. Cien perlas de captura que comprenden ADNc de 100 células individuales se distribuyeron en dos reacciones de tagmentación, una reacción de tagmentación se realizó usando la

- 5 mezcla de transposomas P5\_i5\_2 más P7\_i7\_2 y una segunda reacción de tagmentación se realizó con la mezcla de transposomas P5\_i5\_3 más P7\_i7\_3. Después de las reacciones de tagmentación, las perlas de captura de cada una de las reacciones se agruparon y redistribuyeron para la amplificación por PCR utilizando dos combinaciones únicas de cebadores de PCR P5 y P7 con código de barras (es decir, una primera combinación de cebadores de PCR P5 y P7 y una segunda combinación de cebadores de PCR P5 y P7) para generar dos colecciones de secuenciación. La distribución del tamaño del fragmento en cada una de las colecciones de secuenciación se analizó utilizando un Bioanalizador. La colección con código de barras se analizó después de una sola etapa de limpieza con 0,7x SPRI.

**REIVINDICACIONES**

1. Un método para analizar al menos dos o más analitos de una pluralidad de células individuales, comprendiendo dicho método:

- 5 (a) proporcionar una pluralidad de elementos de preservación de contigüidad (CE), en donde cada uno de los CE comprende una célula individual;
- (b) lisar dicha célula individual dentro de cada uno de los CE, en donde los analitos dentro de dicha célula individual se liberan dentro de cada uno de los CE;
- (c) proporcionar un primer resto informador a un primer analito dentro de dicha célula individual de cada uno de los CE;
- 10 (d) proporcionar un segundo resto indicador a un segundo analito dentro de dicha célula individual de cada uno de los CE;
- (e) modificar dichos analitos de modo que al menos algunos de dichos primer y segundo analitos de dicho CE comprenda dichos restos informadores primero y segundo, respectivamente;
- (f) combinar el CE que comprende dichos analitos que comprenden dichos restos informadores;
- 15 (g) compartimentar el CE que comprende dichos primer y segundo analitos que comprenden dichos primer y segundo restos informadores, respectivamente en una pluralidad de compartimentos;
- (h) proporcionar un tercer resto informador a dicho primer analito que comprende dicho primer resto informador de cada uno de los CE dentro de dichos compartimentos;
- 20 (i) proporcionar un cuarto resto informador a dicho segundo analito que comprende dicho segundo resto informador de cada uno de los CE dentro de dichos compartimentos;
- (j) modificar adicionalmente dichos analitos de modo que al menos algunos primeros analitos comprendan dichos restos informadores primero y tercero y de modo que al menos algunos segundos analitos comprendan dichos restos informadores segundo y cuarto;
- 25 (k) analizar dichos analitos que comprenden dichos restos informadores de cada uno de los compartimentos, en el que dicho análisis detecta una célula individual que es la fuente de cada uno de los analitos.

2. El método de la reivindicación 1, en el que dichos primer y segundo restos informadores identifican la fuente de dichos analitos; o en el que la combinación de dichos restos informadores identifica la fuente de dichos analitos.3. El método de una cualquiera de las reivindicaciones 1-2, en el que la detección de dichos analitos se realiza simultáneamente.

- 30 4. El método de una cualquiera de las reivindicaciones 1-3, en el que dicho primer analito es ADN genómico y el segundo analito es ADNc o ARN; y en el que modificar al menos parte del ADN genómico, ADNc o ARN para que comprenda el primer y segundo restos informadores comprende poner en contacto el ADN genómico, ADNc o ARN con una pluralidad de transposomas, comprendiendo cada uno de los transposomas una transposasa y una secuencia de transposón que comprende el primer resto informador o un segundo resto informador en condiciones tales que al menos algunas de las secuencias de transposón se insertan en el ADN genómico, ADNc o ARN.

5. El método de la reivindicación 4, en el que

- (I) la etapa (g) comprende, además, separar la transposasa del ADN genómico, ADNc o ARN; opcionalmente en el que la transposasa se separa después de modificar dichos analitos en la etapa (e); opcionalmente, en donde la separación de la transposasa comprende un método seleccionado del grupo que consiste en agregar un detergente, cambiar la temperatura, cambiar el pH, agregar una proteasa, agregar una chaperona, cambiar la concentración de sal y agregar una polimerasa de desplazamiento de cadena; o
- 40 (II) la etapa (e) comprende poner en contacto el ácido nucleico con una pluralidad de transposomas, comprendiendo cada uno de los transposomas una primera secuencia de transposón que comprende un primer resto indicador, una segunda secuencia de transposón no contigua a dicha primera secuencia de transposón y una transposasa asociada con la primera secuencia de transposón y la segunda secuencia de transposón.
- 45

6. El método de la reivindicación 5, en el que la primera secuencia de transposón comprende un primer sitio de cebador y la segunda secuencia de transposón comprende un segundo sitio de cebador; opcionalmente en el que el primer sitio de cebador comprende, además, un primer código de barras y el segundo sitio de cebador comprende, además, un segundo código de barras.

7. El método de la reivindicación 1, en el que un analito es proteína, y opcionalmente en el que la proteína está marcada con un resto informador de ácido nucleico y además, opcionalmente en el que el resto informador de ácido nucleico comprende un conjunto de códigos de barras derivados combinatoriamente; en donde dicha proteína es de una célula individual.
- 5 8. El método de cualquiera de las reivindicaciones 1-7, en el que las etapas (c) - (j) se repiten al menos una vez más; opcionalmente en el que los restos informadores adicionales en cada una de las etapas adicionales son diferentes de dichos restos informadores primero, segundo, tercero y cuarto.
9. Un método para analizar analitos de una pluralidad de células individuales, comprendiendo dicho método:
- 10 (a) proporcionar elementos de preservación de contigüidad (CE), en donde cada uno de los CE comprende al menos un analito; en donde dicho al menos un analito de cada uno de los CE son componentes de una célula individual,
- (b) compartimentar dicho CE que comprende dicho al menos un analito en una pluralidad de primeros compartimentos;
- 15 (c) proporcionar un primer conjunto de restos informadores al analito de cada uno de los CE en dicha pluralidad de primeros compartimentos, en donde dicho primer conjunto de restos informadores identifica el CE; en donde un primer resto informador de dicho primer conjunto de restos informadores proporcionados al analito de cada uno de los CE es diferente del primer resto informador proporcionado al analito de cada uno de los otros CE y es único para cada una de las células individuales;
- (d) modificar dicho analito de modo que al menos algunos analitos de dicho CE comprendan un primer resto informador;
- 20 (e) combinar el CE que comprende dicho analito que comprende dicho primer resto informador;
- (f) compartimentar el CE que comprende dicho analito que comprende dicho primer resto informador en una pluralidad de segundos compartimentos;
- 25 (g) proporcionar un segundo conjunto de restos informadores a dicho analito que comprende dicho primer resto informador de cada uno de los CE;
- (h) modificar adicionalmente dicho analito de manera que al menos algunos analitos comprendan un segundo resto informador de dicho segundo conjunto de restos informadores;
- (i) analizar dicho analito que comprende dichos primer y segundo restos informadores de cada segundo compartimento;
- 30 y en el que dichos restos informadores primero y opcionalmente dichos segundos restos identificadores identifican cada una de las células individuales.
10. El método de la reivindicación 9, en el que dicho analito comprende un ácido nucleico; y en el que:
- 35 (I) la etapa (d) comprende poner en contacto el ácido nucleico con una pluralidad de transposomas, cada uno de los cuales comprende una transposasa y una secuencia de transposón que comprende el primer resto informador en condiciones tales que al menos algunas de las secuencias de transposón se insertan en el ácido nucleico; o
- (II) la etapa (d) comprende poner en contacto el ácido nucleico con una pluralidad de transposomas, comprendiendo cada uno de los transposomas una primera secuencia de transposón que comprende un primer resto informador, una segunda secuencia de transposón no contigua a dicha primera secuencia de transposón y una transposasa asociada con la primera secuencia de transposón y la segunda secuencia de transposón.
- 40
11. El método de la reivindicación 10, en el que dicho analito comprende un ácido nucleico, y en el que la etapa (f) comprende separar la transposasa del CE que comprende dicho ácido nucleico que comprende dicho primer resto indicador; y opcionalmente en donde:
- (A) la transposasa se separa después de la etapa (d);
- 45 (B) la transposasa se separa antes de la etapa (i); o
- (C) la separación de la transposasa comprende un método seleccionado del grupo que consiste en agregar un detergente, cambiar la temperatura, cambiar el pH, agregar una proteasa, agregar una chaperona, cambiar la concentración de sal y agregar una polimerasa de desplazamiento de cadena.
12. El método de una cualquiera de las reivindicaciones 10-11, en el que la primera secuencia de transposón

comprende un primer sitio de cebador y la segunda secuencia de transposón comprende un segundo sitio de cebador, opcionalmente en donde el primer sitio de cebador comprende, además, un primer código de barras y el segundo sitio de cebador comprende, además, un segundo código de barras.

5 13. El método de la reivindicación 9, en el que dicho analito comprende un ácido nucleico, en donde dicho primer resto informador comprende un cebador, y en donde la etapa (d) comprende amplificar dicho ácido nucleico con al menos un cebador;

o en el que dicho primer resto informador comprende un cebador, y en donde la etapa (d) comprende ligar dicho ácido nucleico con al menos un cebador.

14. El método de una cualquiera de las reivindicaciones 9-13,

10 que comprende, además, proporcionar un plásmido a dicho CE en la etapa (c), en el que dicho plásmido se modifica para comprender dichos primeros y/o segundos restos informadores en las etapas (d) - (h), en el que dicho plásmido que comprende dichos primeros o dichos segundos restos informadores identifica dicho CE.

15. El método de la reivindicación 9, en el que dicho analito es proteína, opcionalmente en el que dicha proteína es de una célula individual.

15 16. El método de una cualquiera de las reivindicaciones 9-15, en el que las etapas (c) - (h) se repiten al menos una vez más, opcionalmente en el que los conjuntos adicionales de restos informadores en cada una de las etapas adicionales son diferentes a dichos primer y segundo conjuntos de restos informadores.

20 17. El método de la reivindicación 9, en el que dicho analito es ARNm o ADNc, opcionalmente en el que dichos restos informadores primero y segundo identifican la fuente de dichos analitos; o en donde la combinación de dichos restos informadores identifica la fuente de dichos analitos.

18. El método de la reivindicación 17, en el que

(I) dicho primer conjunto de restos informadores comprende la primera secuencia de código de barras y/o en el que dicho segundo conjunto de restos informadores comprende la segunda secuencia de código de barras; o

25 (II) la etapa (d) comprende poner en contacto el ácido nucleico con una pluralidad de transposomas, cada uno de los cuales comprende una transposasa y una secuencia de transposón que comprende el primer resto informador en condiciones tales que al menos algunas de las secuencias de transposón se insertan en el ácido nucleico; o

30 (III) la etapa (d) comprende poner en contacto el ácido nucleico con una pluralidad de transposomas, comprendiendo cada uno de los transposomas una primera secuencia de transposón que comprende un primer resto informador, una segunda secuencia de transposón no contigua a dicha primera secuencia de transposón y una transposasa asociada con la primera secuencia de transposón y la segunda secuencia de transposón;

en donde en la opción (II) u opción (III), la etapa (f) comprende opcionalmente separar la transposasa del CE compartimentado que comprende el ácido nucleico que comprende dicho primer resto informador, opcionalmente en donde:

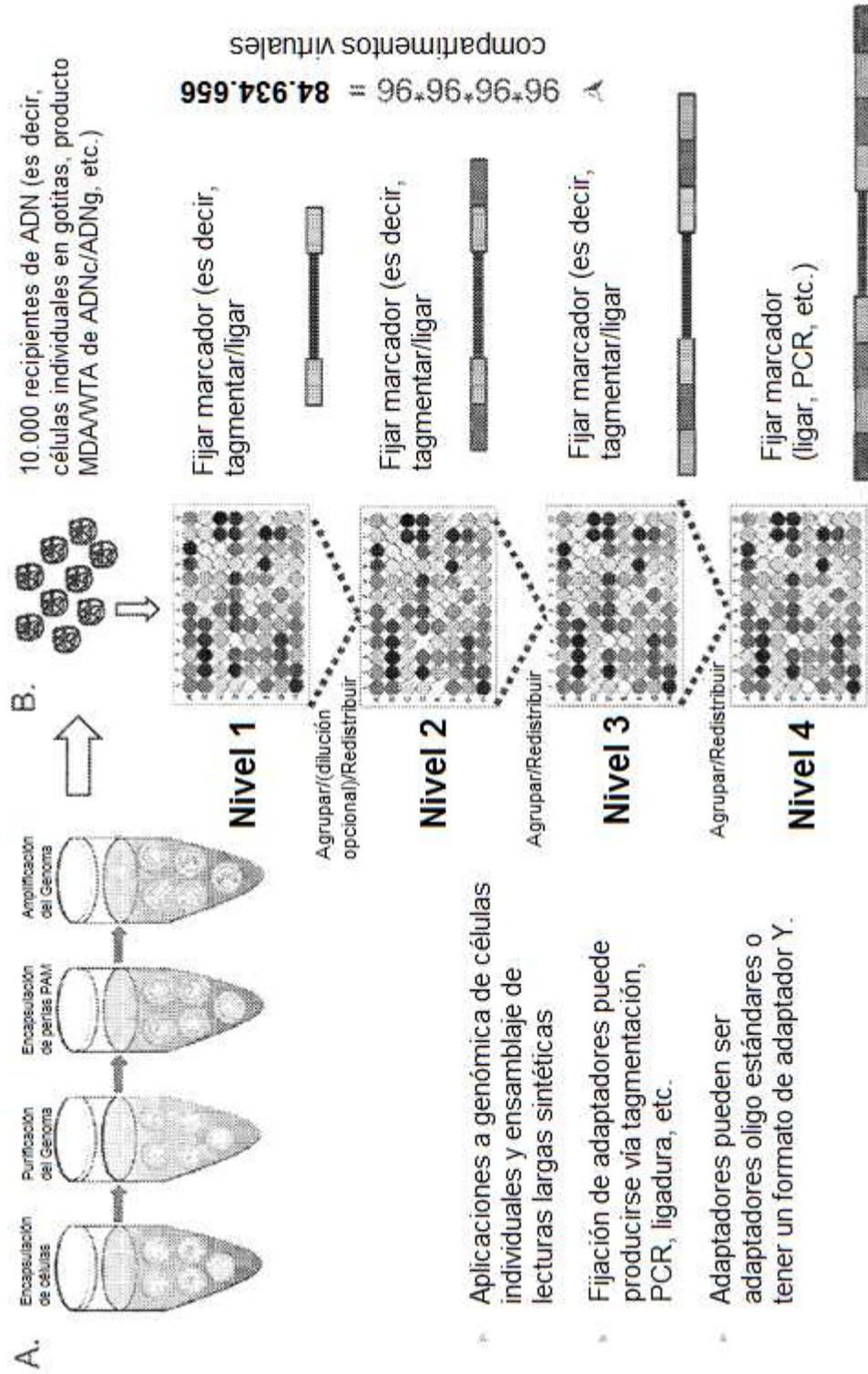
35 (A) la transposasa se separa después de la etapa (d); o

(B) la transposasa se separa antes de la etapa (i); o

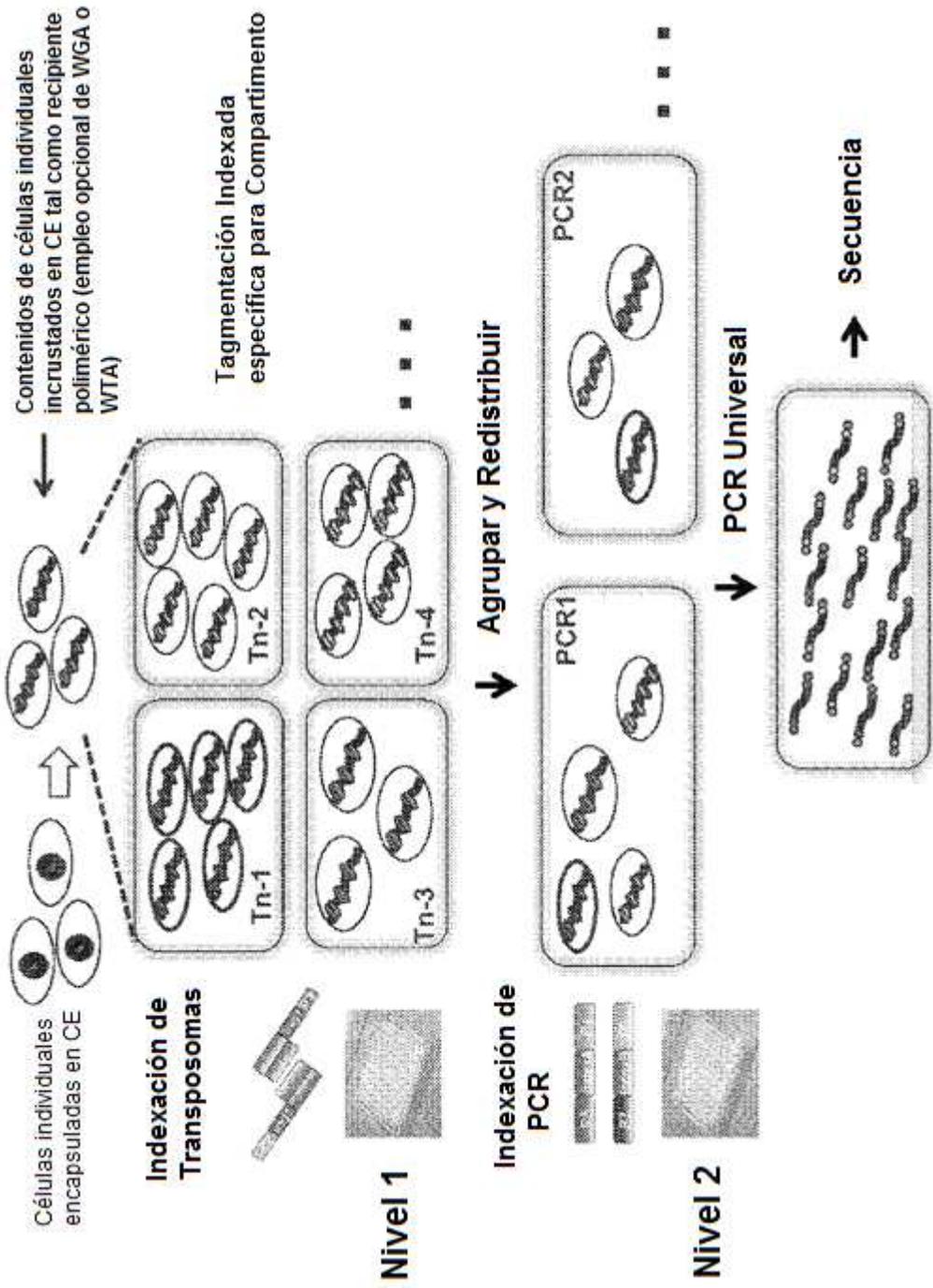
(C) la separación de la transposasa comprende un método seleccionado del grupo que consiste en agregar un detergente, cambiar la temperatura, cambiar el pH, agregar una proteasa, agregar una chaperona, cambiar la concentración de sal y agregar una polimerasa de desplazamiento de cadena.

40

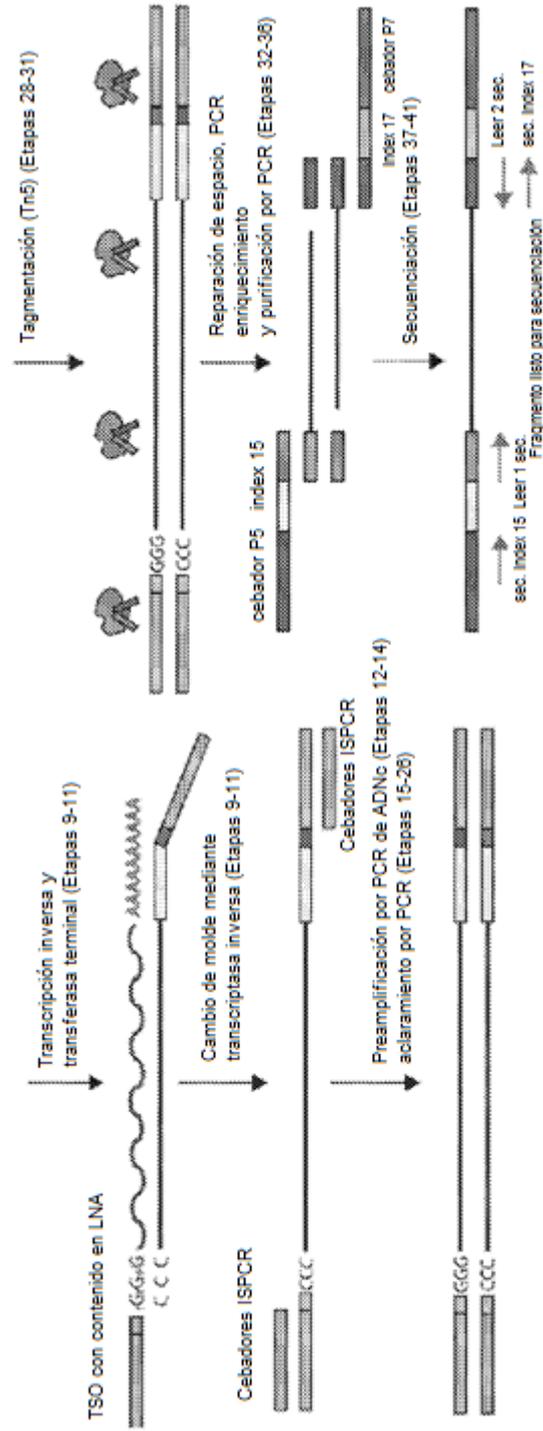
**Fig. 1: Indexación combinatoria de cuatro niveles**

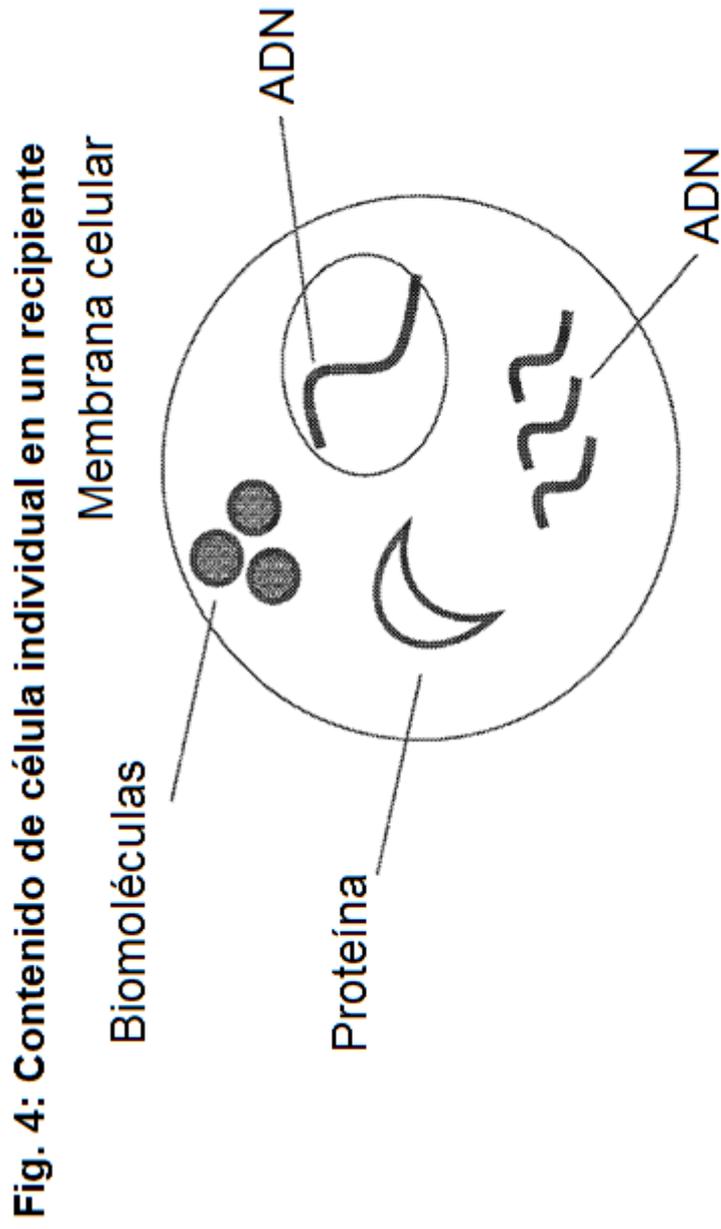


**Fig. 2: Indexación Combinatoria de Dos Niveles para Colecciones de Células Individuales**



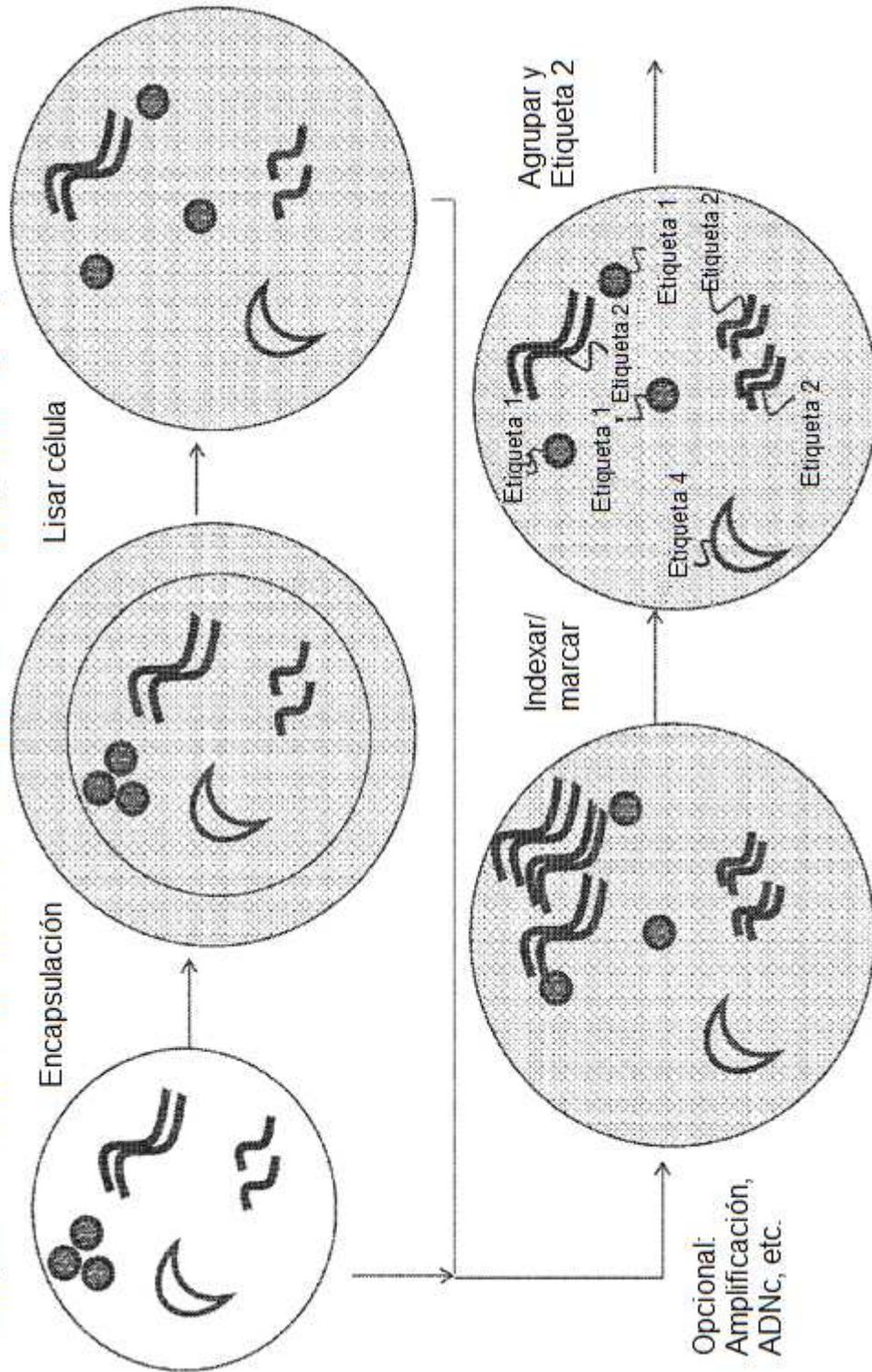
**Fig. 3: Expresión génica de células individuales en CE (p. ej., gotita)**



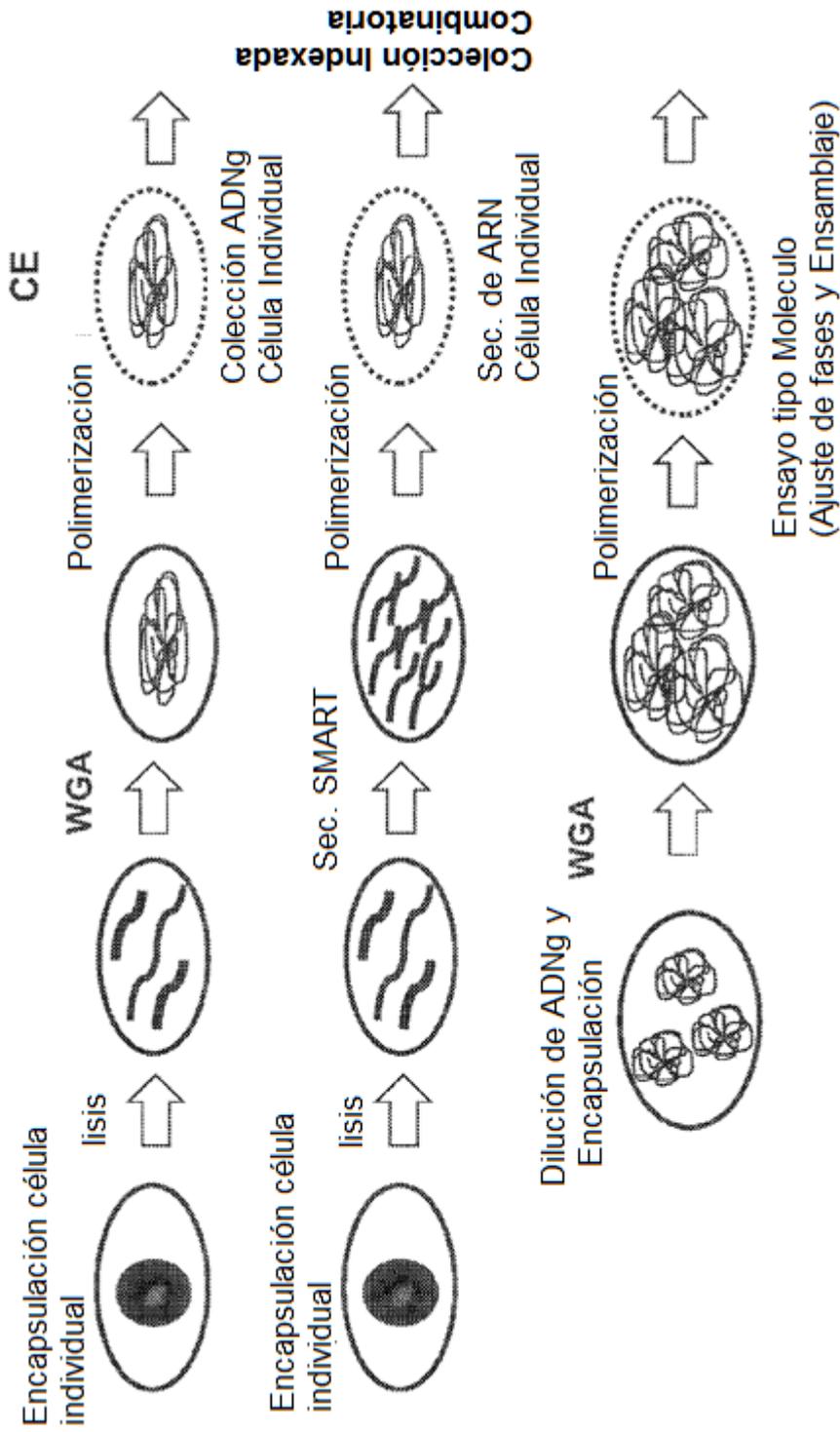


Sección transversal de una célula mostrando las diversas biomoléculas

**Fig. 5A: Formación de un Elemento de Contigüidad (CE)**

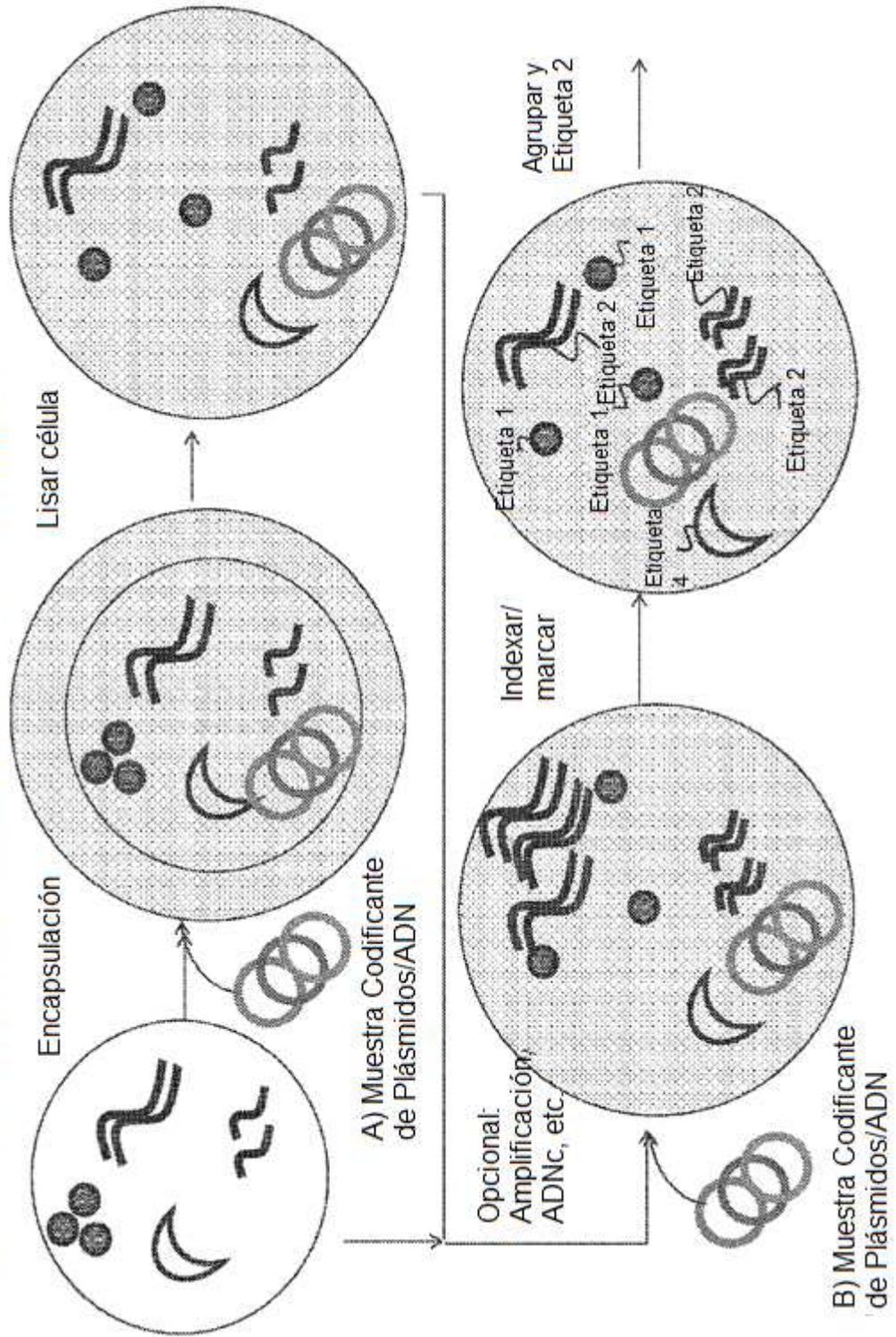


**Fig. 5B: Ensayos personalizados mediante creación de Elementos de Contigüidad (CE)**

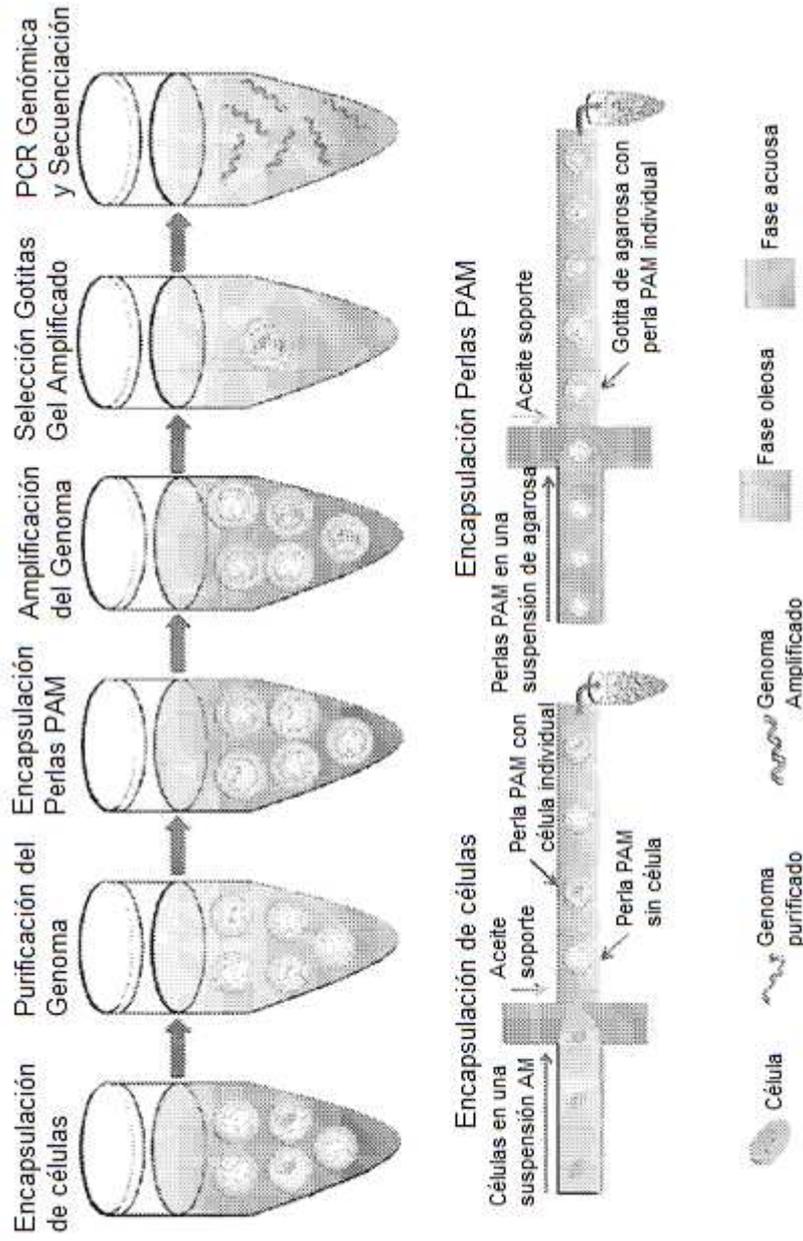


➤ Mismo marco de ensayo permite colecciones de células, ajuste de fases y ensamblaje.

**Fig. 5C: Formación de un Elemento de Contigüidad (CE)**



**Fig. 6. Encapsulación y Amplificación de Células Individuales**



Esquema 1. Diagrama de flujo de trabajo de estrategia para secuenciación genómica de célula individual de alto

Fig. 7. Análisis de alto rendimiento de componentes celulares por captura en superficie directa.

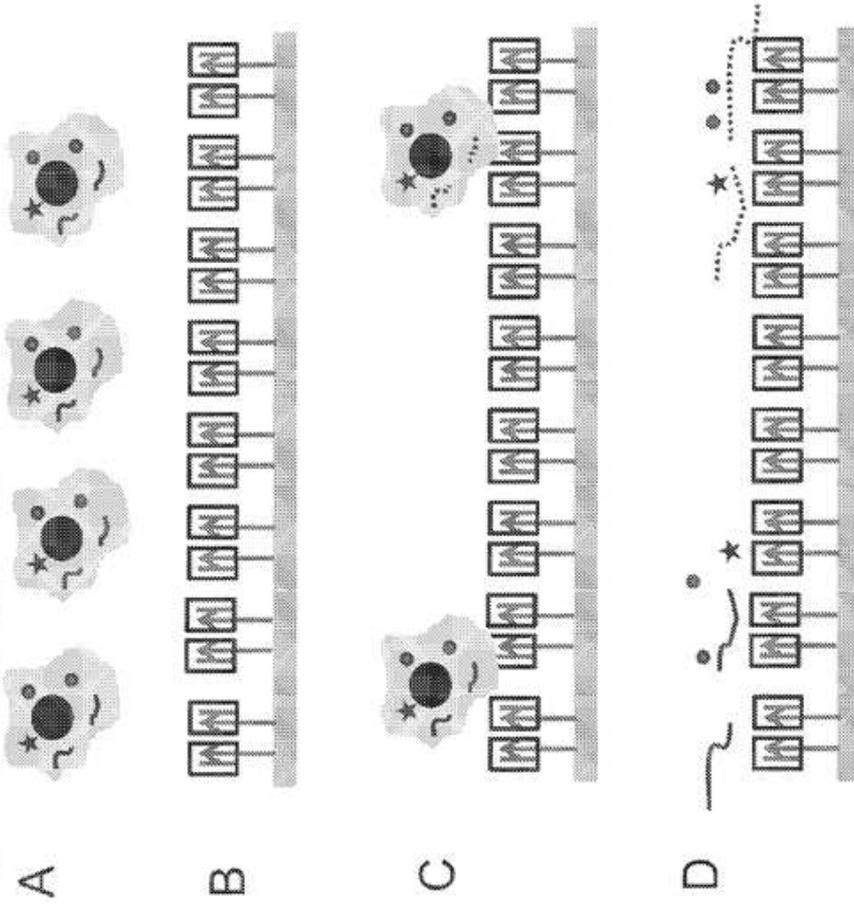
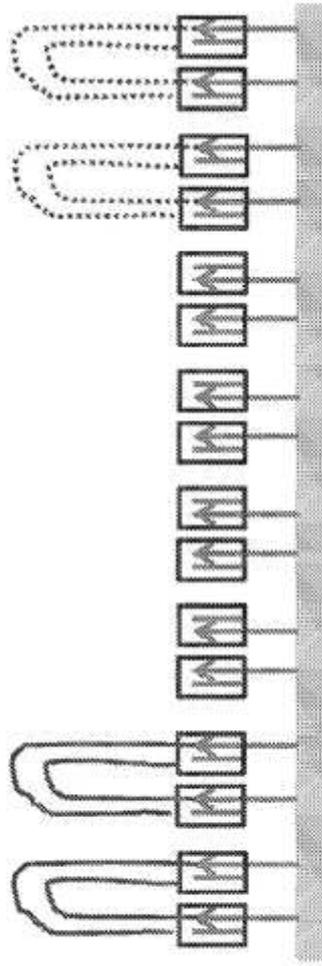


Fig. 7 (Continúa).

E



**Fig. 8. Optimización de la Sec. CPT: en flujo de trabajo de perlas**

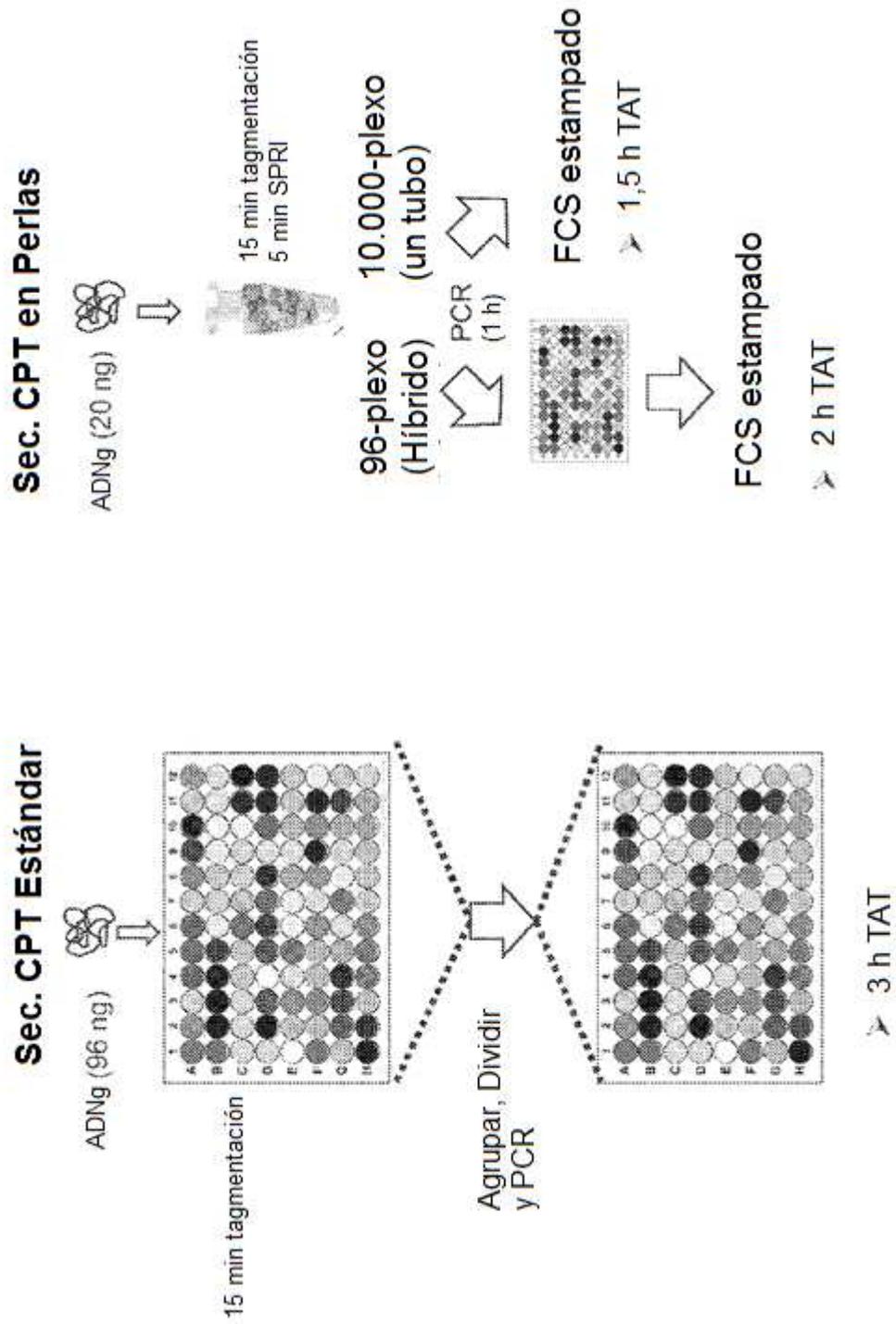


Fig. 9A. Partición y Enfoque de Mutagénesis para Ensamblar Regiones Repetidas en ADN Genómico

**comparación de 96 particiones con 46000 particiones**

Fig. 9B. Partición y Enfoque de Mutagénesis para Ensamblar Regiones Repetidas en ADN Genómico

## **Estipulación experimental**

- Fragmento de 10 kb
- 10 mutaciones en cada caso
- Cobertura larga = 10
- 20% mutación C->T
- Tamaño inserción = 500 pb
- 2x300 pb, sin error de secuenciación
- Genoma de Ecoli -> 46000 fragmentos de 10 kb
- Ensamblador Spade

Fig. 9C. Partición y Enfoque de Mutagénesis para Ensamblar Regiones Repetidas en ADN Genómico

## 96 particiones

- ▶ Distribución del tamaño de cóntigos reseñada por SPADe al analizar en 96 particiones

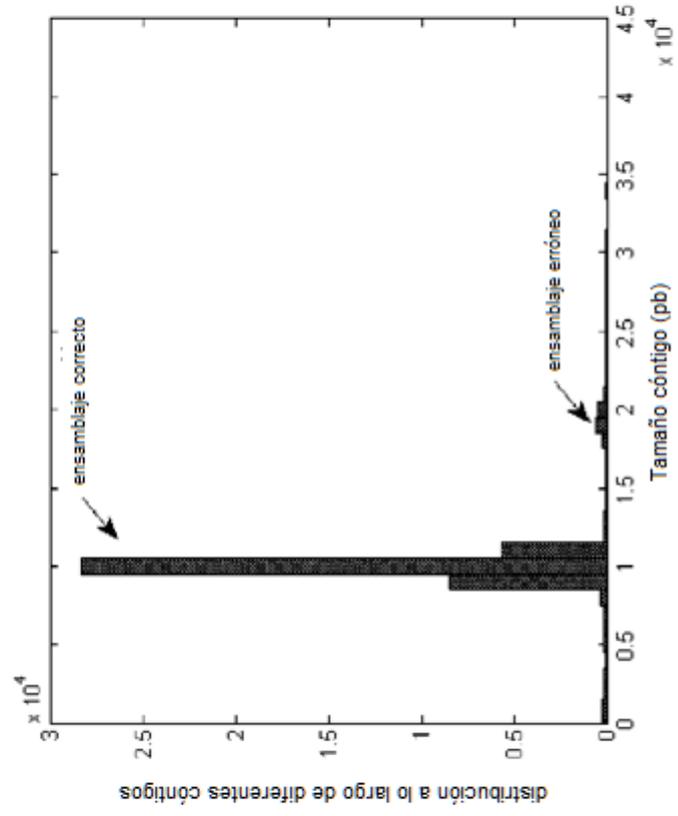


Fig. 9D. Partición y Enfoque de Mutagénesis para Ensamblar Regiones Repetidas en ADN Genómico

### 46000 particiones

- \* Distribución del tamaño de cóntigos reseñados por SPAdE al analizar 46000 particiones (cada fragmento de 10 kb en una partición individual)

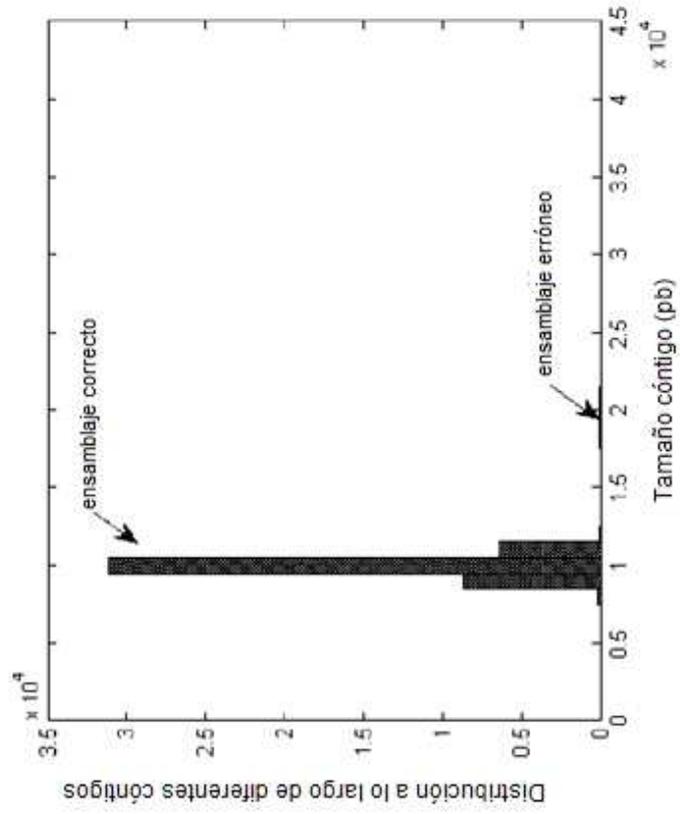


Fig. 10

