

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 990 227**

51 Int. Cl.:

**C12Q 1/68** (2008.01)

**C12Q 1/6806** (2008.01)

**C12Q 1/6837** (2008.01)

**C12Q 1/6844** (2008.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **15.01.2009** **E 19214834 (4)**

97 Fecha y número de publicación de la concesión europea: **07.08.2024** **EP 3699291**

54 Título: **Composiciones y procedimientos de análisis de ácido nucleico de secuencia única**

30 Prioridad:

**17.01.2008 US 2187108 P**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:  
**29.11.2024**

73 Titular/es:

**SEQUENOM, INC. (100.0%)**  
**3595 John Hopkins Court**  
**San Diego, CA 92121, US**

72 Inventor/es:

**CANTOR, CHARLES R.**

74 Agente/Representante:

**ILLESCAS TABOADA, Manuel**

ES 2 990 227 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

Composiciones y procedimientos de análisis de ácido nucleico de secuencia única

## 5 Campo de la invención

La presente invención se define por las reivindicaciones. La presente divulgación se refiere de manera general al campo del análisis de secuencias de ácido nucleico y a la metodología y a los componentes para su uso en tal análisis.

## 10 Sumario

La presente invención se define por las reivindicaciones.

15 Por consiguiente, la presente invención se refiere en un primer aspecto a una colección de soportes sólidos que comprende especies de ácido nucleico monocatenario, (a) en la que los soportes sólidos son perlas o partículas, (b) en la que cada soporte sólido en la colección comprende más de una especie de ácido nucleico monocatenario covalentemente unida al soporte sólido, en la que cada especie de ácido nucleico monocatenario unida a un soporte sólido común comprende una secuencia identificadora y una secuencia de sonda, (c) en la que todas las especies de ácido nucleico monocatenario unidas al soporte sólido común comparten una secuencia identificadora común, y en la que la secuencia identificadora de las más de una especies de ácido nucleico monocatenario unidas al soporte sólido común en la colección es diferente de las secuencias identificadoras de las especies de ácido nucleico monocatenario unidas a los demás soportes sólidos en la colección, mediante lo cual cada soporte sólido en la colección está asociado con una secuencia identificadora singular que difiere, cada una, de todas las demás secuencias identificadoras en uno o más nucleótidos, cuando las secuencias identificadoras se alinean entre sí y en la que sólo está presente una especie de secuencia de identificación singular por especie de soporte sólido, (d) en la que la secuencia de sonda de cada una de las más de una especies de ácido nucleico monocatenario unidas al soporte sólido permite que los más de un ácidos nucleicos monocatenarios se hibriden con una subsecuencia en un ácido nucleico de muestra, y (e) en la que las secuencias de sonda de las más de una especies de ácido nucleico monocatenario en cada soporte sólido comprenden al menos 10 especies de secuencia de sonda diferentes, en la que cada especie de secuencia de sonda difiere de todas las demás especies de secuencia de sonda en uno o más nucleótidos, cuando las especies de secuencia de sonda se alinean entre sí.

35 Según una realización preferida, la especie de ácido nucleico monocatenario de cada soporte sólido en la colección comprende además una secuencia de cebador.

Según una realización más preferida, la secuencia de cebador, la secuencia identificadora y la secuencia de sonda están orientadas 5'-(secuencia de cebador)-(secuencia identificadora)-(secuencia de sonda)-3'.

40 Según otra realización más preferida, las especies de ácido nucleico monocatenario unidas a un soporte sólido común en la colección comparten una secuencia de cebador común, o en la que las especies de ácido nucleico monocatenario de todos los soportes sólidos en la colección comparten una secuencia de cebador común.

45 Según una realización preferida, los soportes sólidos son microperlas, nanoperlas, micropartículas o nanopartículas.

Según una realización preferida, las perlas o partículas comprenden un material magnético.

Según una realización preferida, las más de una secuencias de sonda son complementarias a una secuencia de ADN, una secuencia de ADN fragmentado o un ácido nucleico monocatenario.

50 La presente invención se refiere en un segundo aspecto a un sustrato que comprende una colección de soportes sólidos según el primer aspecto, en el que los soportes sólidos son microperlas, nanoperlas, micropartículas o nanopartículas, en el que: (a) los soportes sólidos son perlas; y (b) las perlas están orientadas en una matriz.

55 Según una realización preferida, la especie de ácido nucleico monocatenario de cada perla en la colección comprende además una secuencia de cebador.

Según una realización más preferida, la secuencia de cebador, la secuencia identificadora y la secuencia de sonda están orientadas 5'-(secuencia de cebador)-(secuencia identificadora)-(secuencia de sonda)-3'.

60 Según otra realización preferida, las especies de ácido nucleico monocatenario unidas a una perla común en la colección comparten una secuencia de cebador común, o

65 en el que las especies de ácido nucleico monocatenario de todas las perlas en la colección comparten una secuencia de cebador común.

Según una realización preferida, el soporte sólido es una microperla o una nanopera, y/o en el que la colección de soportes sólidos está comprendida en una emulsión.

Según otra realización preferida, la perla comprende un gel o un material magnético.

Según una realización preferida adicional, las más de una secuencias de sonda son complementarias a una secuencia de ADN o secuencia de ADN fragmentado.

La presente invención se refiere en un tercer aspecto a un método de fabricación de la colección de soportes sólidos según la reivindicación 1, que comprende (a) unir secuencialmente nucleótidos a un nucleótido covalentemente unido a cada soporte sólido en la colección, mediante lo cual se prepara cada una de las especies de ácido nucleico monocatenario y está en asociación con cada soporte sólido; o (b) unir cada especie de ácido nucleico monocatenario en fase de disolución a cada soporte sólido en la colección, mediante lo cual las más de una especies de ácido nucleico monocatenario en fase sólida están en asociación con cada uno de los soportes sólidos; en el que: (i) cada especie de ácido nucleico monocatenario en fase sólida comprende una secuencia identificadora y una secuencia de sonda, en el que las especies de ácido nucleico en fase sólida del soporte sólido comparten una secuencia de sonda común o no comparten una secuencia de sonda común, (ii) en el que todas las especies de ácido nucleico monocatenario unidas al soporte sólido común comparten una secuencia identificadora común, y en el que la secuencia identificadora de las más de una especies de ácido nucleico monocatenario unidas al soporte sólido común en la colección es diferente de la secuencia identificadora de las especies de ácido nucleico monocatenario unidas a los demás soportes sólidos en la colección, mediante lo cual cada soporte sólido en la colección está asociado con una secuencia identificadora singular que difiere, cada una, de todas las demás secuencias identificadoras en uno o más nucleótidos, cuando las secuencias identificadoras se alinean entre sí y en la que sólo está presente una especie de secuencia de identificación singular por especie de soporte sólido, y (iii) en el que la secuencia de sonda de cada una de las más de una especies de ácido nucleico monocatenario unidas al soporte sólido permite que los más de un ácidos nucleicos monocatenarios se hibriden con una subsecuencia en un ácido nucleico de muestra, y (c) en el que las secuencias de sonda de las más de una especies de ácido nucleico monocatenario en cada soporte sólido comprenden al menos 10 especies de secuencia de sonda diferentes, en el que cada especie de secuencia de sonda difiere de todas las demás especies de secuencia de sonda en uno o más nucleótidos, cuando las especies de secuencia de sonda se alinean entre sí.

### Breve descripción de los dibujos

Los dibujos ilustran características de determinadas realizaciones de la invención. Las figuras 1A-1C muestran ejemplos de realizaciones de soporte sólido. La figura 2 muestra un procedimiento representativo para generar ácidos nucleicos que tienen una secuencia de nucleótidos complementaria a una secuencia de nucleótidos diana.

### Descripción detallada

Los procedimientos de análisis de secuencias de ácido nucleico y los soportes sólidos mejorados descritos en el presente documento encuentran múltiples usos en investigación y aplicaciones clínicas. Tales procedimientos y soportes sólidos pueden utilizarse, por ejemplo, para: (a) determinar si una secuencia diana particular está presente en una muestra; (b) realizar análisis de mezclas, por ejemplo, identificar una mezcla y/o su composición o determinar la frecuencia de una secuencia diana en una mezcla (por ejemplo, comunidades mixtas, cuasiespecies); (c) detectar variaciones de secuencia (por ejemplo, mutaciones, polimorfismos de un solo nucleótido) en una muestra; (d) realizar determinaciones de haplotipado; (e) realizar tipado de microorganismos (por ejemplo, patógenos); (f) detectar la presencia o ausencia de una secuencia diana de microorganismo en una muestra; (g) identificar marcadores de enfermedad; (h) detectar microsatélites; (i) identificar repeticiones en tándem cortas; (j) identificar un organismo u organismos; (k) detectar variaciones alélicas; (l) determinar la frecuencia alélica; (m) determinar patrones de metilación; (n) realizar determinaciones epigenéticas; (o) resecuenciar una región de una biomolécula; (p) medicina e investigación clínicas en seres humanos (por ejemplo, detección de marcadores de cáncer, detección de variaciones de secuencia; detección de firmas de secuencia favorables o desfavorables para la administración de un fármaco particular), (q) tipado de HLA; (r) criminalística; (s) control de calidad de vacunas; (t) monitorización de tratamientos; (u) identidad de vectores; (v) realizar control de calidad de cepas de producción o vacunas, (w) detectar la identidad de cepas de prueba, (x) identificar una secuencia o secuencias de ácido nucleico virales específicas en una mezcla o población viral (por ejemplo, mezclas de hepatitis, mezclas de VIH, poblaciones virales mixtas como las que podrían encontrarse en un organismo inmunodeficiente o inmunodeprimido). A continuación se describen determinados aspectos de la invención.

### Ácido nucleico de muestra

Un ácido nucleico de muestra puede derivarse de una o más muestras o fuentes. Tal como se usa en el presente documento, "ácido nucleico" se refiere a polinucleótidos tales como ácido desoxirribonucleico (ADN) y ácido ribonucleico (ARN). Debe entenderse que el término incluye, como equivalentes, derivados, variantes y análogos de ARN o ADN preparados a partir de análogos de nucleótidos, polinucleótidos monocatenarios (sentido o antisentido) y bicatenarios. Debe entenderse que el término "ácido nucleico" no se refiere a ni infiere una longitud específica de

la cadena de polinucleótido, por tanto en la definición también se incluyen nucleótidos, polinucleótidos y oligonucleótidos. Los desoxirribonucleótidos incluyen desoxiadenosina, desoxicitidina, desoxiguanosina y desoxitimidina. Para el ARN, la base uracilo es uridina. Una fuente o muestra que contiene ácido(s) nucleico(s) de muestra puede contener una o una pluralidad de ácidos nucleicos de muestra. Una pluralidad de ácidos nucleicos de muestra tal como se describe en el presente documento se refiere a al menos 2 ácidos nucleicos de muestra e incluye secuencias de ácido nucleico que pueden ser idénticas o diferentes. Es decir, los ácidos nucleicos de muestra pueden ser todos ellos representativos de la misma secuencia de ácido nucleico, o pueden ser representativos de dos o más secuencias de ácido nucleico diferentes (por ejemplo, de 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 50, 100, 1000 o más secuencias).

Una muestra puede recogerse a partir de un organismo, un sitio mineral o geológico (por ejemplo, suelo, roca, depósito mineral, escenario de combate), un sitio de criminalística (por ejemplo, escena de crimen, contrabando o contrabando sospechoso) o un sitio paleontológico o arqueológico (por ejemplo, fósil o hueso), por ejemplo. Una muestra puede ser una "muestra biológica", que se refiere a cualquier material obtenido a partir de una fuente viva o fuente anteriormente viva, por ejemplo, un animal tal como un ser humano u otro mamífero, una planta, una bacteria, un hongo, un protista o un virus. La muestra biológica puede estar en cualquier forma, incluyendo sin limitación un material sólido tal como un tejido, células, un sedimento celular, un extracto celular o una biopsia, o un líquido biológico tal como orina, sangre, saliva, líquido amniótico, exudado de una región de infección o inflamación o un enjuague bucal que contiene células bucales, orina, líquido cefalorraquídeo y líquido sinovial y órganos. Una muestra también puede aislarse en un punto de tiempo diferente en comparación con otra muestra, donde cada una de las muestras procede de la misma fuente o una fuente diferente. Un ácido nucleico de muestra puede ser de una biblioteca de ácidos nucleicos, tal como una biblioteca de ADNc o ARN, por ejemplo. Un ácido nucleico de muestra puede ser un resultado de la purificación o el aislamiento de ácidos nucleicos y/o la amplificación de moléculas de ácido nucleico a partir de la muestra. Un ácido nucleico de muestra proporcionado para los procedimientos de análisis de secuencias descritos en el presente documento puede contener ácido nucleico de una muestra o de dos o más muestras (por ejemplo, de 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 ó 20 muestras). Una muestra de ácido nucleico de muestra puede contener ácido nucleico de muestra huésped y no huésped, y en algunas implementaciones, una muestra puede contener dos o más especies diferentes de ácido nucleico de muestra (por ejemplo, mutante frente a tipo natural, trasplantes, criminalística, madre frente a feto).

Un ácido nucleico de muestra puede comprender o consistir esencialmente en cualquier tipo de ácido nucleico adecuado para su uso con los procedimientos de la presente divulgación, tal como ácido nucleico de muestra que puede hibridarse con ácido nucleico en fase sólida (descrito más adelante), por ejemplo. En determinadas implementaciones, un ácido nucleico de muestra puede comprender o consistir esencialmente en ADN (por ejemplo, ADN complementario (ADNc), ADN genómico (ADNg), y similares), ARN (por ejemplo, ARN mensajero (ARNm), ARN de interferencia pequeño (ARNip), ARN ribosómico (ARNr), ARNt, y similares) y/o análogos de ADN o ARN (por ejemplo, que contienen análogos de bases, análogos de azúcares y/o una estructura principal no nativa, y similares). Un ácido nucleico puede estar en cualquier forma útil para llevar a cabo los procedimientos en el presente documento (por ejemplo, lineal, circular, superenrollado, monocatenario, bicatenario, y similares). En determinadas implementaciones, un ácido nucleico puede ser, o puede proceder de, un plásmido, un fago, una secuencia de replicación autónoma (ARS), un centrómero, un cromosoma artificial, un cromosoma, una célula, un núcleo celular o un citoplasma de una célula. En determinadas implementaciones, un ácido nucleico de muestra procede de un único cromosoma (por ejemplo, una muestra de ácido nucleico puede proceder de un cromosoma de una muestra obtenida a partir de un organismo diploide).

En determinadas implementaciones, puede proporcionarse un ácido nucleico de muestra para llevar a cabo los métodos descritos en el presente documento sin procesamiento de la(s) muestra(s) que contiene(n) el ácido nucleico. En algunas implementaciones, se proporciona un ácido nucleico de muestra para llevar a cabo los métodos descritos en el presente documento después del procesamiento de la(s) muestra(s) que contiene(n) el ácido nucleico. Por ejemplo, un ácido nucleico de muestra puede extraerse, aislarse, purificarse o amplificarse a partir de la(s) muestra(s). El término "aislado" tal como se usa en el presente documento se refiere a un ácido nucleico retirado de su entorno original (por ejemplo, el entorno natural si se produce de manera natural o una célula huésped si se expresa de manera exógena) y, por tanto, se altera "por la mano del hombre" de su entorno original. Un ácido nucleico aislado generalmente está dotado de menos componentes distintos de ácido nucleico (por ejemplo, proteína, lípido) que la cantidad de componentes presentes en una muestra fuente. Una composición que comprende ácido nucleico de muestra aislado puede estar sustancialmente aislada (por ejemplo, aproximadamente el 90 %, el 91 %, el 92 %, el 93 %, el 94 %, el 95 %, el 96 %, el 97 %, el 98 %, el 99 % o más del 99 % libre de componentes distintos de ácido nucleico). El término "purificado" tal como se usa en el presente documento se refiere a un ácido nucleico de muestra siempre que contenga menos especies de ácido nucleico que en la fuente de muestra a partir de la cual se deriva el ácido nucleico de muestra. Una composición que comprende ácido nucleico de muestra puede estar sustancialmente purificada (por ejemplo, aproximadamente el 90 %, el 91 %, el 92 %, el 93 %, el 94 %, el 95 %, el 96 %, el 97 %, el 98 %, el 99 % o más del 99 % libre de otras especies de ácido nucleico). El término "amplificado" tal como se usa en el presente documento se refiere a someter el ácido nucleico de una muestra a un procedimiento que genera de manera lineal o exponencial ácidos nucleicos de amplificación que tienen la misma o sustancialmente la misma secuencia de nucleótidos que la secuencia de nucleótidos del ácido nucleico en la muestra, o una porción de la misma.

En determinadas implementaciones, un ácido nucleico de muestra también puede procesarse sometiendo el ácido nucleico a un método que genera fragmentos de ácido nucleico antes de proporcionar el ácido nucleico de muestra para un procedimiento descrito en el presente documento. En algunas implementaciones, el ácido nucleico de muestra sometido a fragmentación o escisión puede tener una longitud nominal, promedio o media de aproximadamente 5 a aproximadamente 10.000 pares de bases, de aproximadamente 100 a aproximadamente 1.000 pares de bases, de aproximadamente 100 a aproximadamente 500 pares de bases, o de aproximadamente 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000 ó 10000 pares de bases. Pueden generarse fragmentos mediante cualquier método adecuado conocido en la técnica, y la longitud promedio, media o nominal de los fragmentos de ácido nucleico puede controlarse seleccionando un procedimiento de generación de fragmentos apropiado. En determinadas implementaciones, puede utilizarse un ácido nucleico de muestra de una longitud relativamente más corta para analizar secuencias que contienen poca variación de secuencia y/o contienen cantidades relativamente grandes de información de secuencia de nucleótidos conocida. En algunas implementaciones, puede utilizarse un ácido nucleico de muestra de una longitud relativamente más larga para analizar secuencias que contienen una mayor variación de secuencia y/o contienen cantidades relativamente pequeñas de información de secuencia de nucleótidos desconocida. A menudo, los fragmentos de ácido nucleico de muestra contienen secuencias de nucleótidos solapantes, y tales secuencias solapantes pueden facilitar la construcción de una secuencia de nucleótidos del ácido nucleico de muestra previamente no fragmentado, o una porción de la misma. Por ejemplo, un fragmento puede tener las subsecuencias x e y y otro fragmento puede tener las subsecuencias y y z, donde x, y y z son secuencias de nucleótidos que pueden ser de 5 nucleótidos de longitud o más. La secuencia solapante y puede utilizarse para facilitar la construcción de la secuencia de nucleótidos x-y-z en un ácido nucleico de una muestra. En determinadas implementaciones, un ácido nucleico de muestra puede estar parcialmente fragmentado (por ejemplo, a partir de una reacción de escisión específica incompleta o terminada) o totalmente fragmentado.

Un ácido nucleico de muestra puede fragmentarse mediante diversos métodos, que incluyen, sin limitación, procedimientos físicos, químicos y enzimáticos. Se describen ejemplos de tales procedimientos en la publicación de solicitud de patente estadounidense n.º 20050112590 (publicada el 26 de mayo de 2005, titulada "Fragmentation-based methods and systems for sequence variation detection and discovery", a nombre de Van Den Boom *et al.*). Pueden seleccionarse determinados procedimientos para generar fragmentos escindidos inespecíficamente o fragmentos escindidos específicamente. Los ejemplos de procedimientos que pueden generar ácido nucleico de muestra de fragmentos escindidos inespecíficamente incluyen, sin limitación, poner en contacto un ácido nucleico de muestra con un aparato que expone el ácido nucleico a una fuerza de cizalladura (por ejemplo, hacer pasar el ácido nucleico a través de una aguja de jeringa; usar una prensa francesa); exponer el ácido nucleico de muestra a irradiación (por ejemplo, gamma, rayos X, irradiación UV); los tamaños de fragmento pueden controlarse por la intensidad de irradiación; someter a ebullición el ácido nucleico en agua (por ejemplo, produce fragmentos de aproximadamente 500 pares de bases) y exponer el ácido nucleico a un procedimiento de hidrólisis con ácido y base.

Un ácido nucleico de muestra puede escindirse específicamente poniendo en contacto el ácido nucleico con uno o más agentes de escisión específicos. El término "agente de escisión específico" tal como se usa en el presente documento se refiere a un agente, a veces un producto químico o una enzima, que puede escindir un ácido nucleico en uno o más sitios específicos. A menudo, los agentes de escisión específicos escindirán específicamente según una secuencia de nucleótidos particular en un sitio particular.

Los ejemplos de agentes de escisión específicos enzimáticos incluyen sin limitación endonucleasas (por ejemplo, ADNasa (por ejemplo, ADNasa I, II); ARNasa (por ejemplo, ARNasa E, F, H, P); enzima Cleavase™; ADN polimerasa Taq; ADN polimerasa I de *E. coli* y endonucleasas específicas de estructura eucariotas; endonucleasas FEN-1 murinas; endonucleasas de restricción de tipo I, II o III tales como Acc I, Afl III, Alu I, Alw44 I, Apa I, Asn I, Ava I, BamH I, Ban II, Bcl I, Bgl I, Bgl II, Bln I, Bsm I, BssH II, BstE II, Cfo I, Cla I, Dde I, Dpn I, Dra I, EclX I, EcoR I, EcoR II, EcoR V, Hae II, Hind II, Hind III, Hpa I, Hpa II, Kpn I, Ksp I, Mlu I, MluN I, Msp I, Nci I, Nco I, Nde I, Nde II, Nhe I, Not I, Nru I, Nsi I, Pst I, Pvu I, Pvu II, Rsa I, Sac I, Sal I, Sau3A I, Sca I, ScrF I, Sfi I, Sma I, Spe I, Sph I, Ssp I, Stu I, Sty I, Swa I, Taq I, Xba I, Xho I); glicosilasas (por ejemplo, uracilo-ADN glicosilasa (UDG), 3-metiladenina-ADN glicosilasa, 3-metiladenina-ADN glicosilasa II, pirimidina-hidrato-ADN glicosilasa, FaPy-ADN glicosilasa, timina-ADN glicosilasa de emparejamiento erróneo, hipoxantina-ADN glicosilasa, 5-hidroximetiluracilo-ADN glicosilasa (HmUDG), 5-hidroximetilcitosina-ADN glicosilasa o 1,N6-eteno-adenina-ADN glicosilasa); exonucleasas (por ejemplo, exonucleasa III); ribozimas y ADNzimas. El ácido nucleico de muestra puede tratarse con un agente químico, o sintetizarse usando nucleótidos modificados, y puede escindirse el ácido nucleico modificado. En ejemplos no limitativos, el ácido nucleico de muestra puede tratarse con (i) agentes alquilantes tales como metilnitrosourea que generan varias bases alquiladas, incluyendo N3-metiladenina y N3-metilguanina, que son reconocidas y escindidas por alquil-purina-ADN-glicosilasa; (ii) bisulfito de sodio, que provoca la desaminación de los residuos de citosina en el ADN para formar residuos de uracilo que pueden ser escindidos por uracilo-N-glicosilasa; y (iii) un agente químico que convierte la guanina en su forma oxidada, 8-hidroxiguanina, que puede ser escindida por formamidopirimidina-ADN-N-glicosilasa. Los ejemplos de procedimientos de escisión química incluyen sin limitación alquilación (por ejemplo, alquilación de ácido nucleico modificado con fosforotioato); escisión por labilidad a los ácidos del ácido

nucleico que contiene P3'-N5'-fosforoamidato; y tratamiento de ácido nucleico con tetróxido de osmio y piperidina.

Tal como se usa en el presente documento, el término "reacciones de escisión complementarias" se refiere a reacciones de escisión que se llevan a cabo en el mismo ácido nucleico de muestra usando reactivos de escisión diferentes o alterando la especificidad de escisión del mismo reactivo de escisión de manera que se generen patrones de escisión alternos del mismo ácido nucleico o proteína diana o de referencia. En determinadas implementaciones, el ácido nucleico de muestra puede tratarse con uno o más agentes de escisión específicos (por ejemplo, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 o más agentes de escisión específicos) en uno o más recipientes de reacción (por ejemplo, el ácido nucleico de muestra se trata con cada agente de escisión específico en una recipiente independiente).

Un ácido nucleico de muestra también puede exponerse a un procedimiento que modifique determinados nucleótidos en el ácido nucleico antes de proporcionar el ácido nucleico de muestra para un método descrito en el presente documento. Un procedimiento que modifica selectivamente el ácido nucleico basándose en el estado de metilación de los nucleótidos en el mismo puede aplicarse al ácido nucleico de muestra. El término "estado de metilación" tal como se usa en el presente documento se refiere a si un nucleótido particular en una secuencia de polinucleótido está metilado o no metilado. En la técnica se conocen métodos para modificar una molécula de ácido nucleico diana de una manera que refleje el patrón de metilación de la molécula de ácido nucleico diana, tal como se ejemplifica en la patente estadounidense n.º 5.786.146 y las publicaciones de patente estadounidense 20030180779 y 20030082600. Por ejemplo, los nucleótidos de citosina no metilada en un ácido nucleico pueden convertirse en uracilo mediante tratamiento con bisulfito, que no modifica la citosina metilada. Los ejemplos no limitativos de agentes que pueden modificar una secuencia de nucleótidos de un ácido nucleico incluyen metanosulfonato de metilo, metanosulfonato de etilo, sulfato de dietilo, nitrosoguanidina (N-metil-N'-nitro-N-nitrosoguanidina), ácido nitroso, sulfuro de di-(2-cloroetilo), di-(2-cloroetil)metilamina, 2-aminopurina, t-bromouracilo, hidroxilamina, bisulfito de sodio, hidrazina, ácido fórmico, nitrito de sodio y 5-metilcitosina-ADN glicosilasa. Además, condiciones tales como alta temperatura, radiación ultravioleta, rayos X, pueden inducir cambios en la secuencia de una molécula de ácido nucleico.

Un ácido nucleico de muestra puede proporcionarse en cualquier forma útil para llevar a cabo un procedimiento de fabricación o análisis de secuencias descrito en el presente documento, tal como forma sólida o líquida, por ejemplo. En determinadas implementaciones, un ácido nucleico de muestra puede proporcionarse en forma líquida que comprende opcionalmente uno o más de otros componentes, incluyendo sin limitación uno o más tampones o sales.

#### Soportes sólidos y ácido nucleico en fase sólida

El término "soporte sólido" o "fase sólida" tal como se usa en el presente documento se refiere a una amplia variedad de materiales incluyendo sólidos, semisólidos, geles, películas, membranas, mallas, fieltros, materiales compuestos, partículas, y similares, usados normalmente para secuestrar moléculas, y más específicamente se refiere a un material insoluble con el que puede asociarse un ácido nucleico. Un soporte sólido para su uso con los procedimientos descritos en el presente documento a veces se selecciona en parte según el tamaño: a veces se seleccionan soportes sólidos que tienen un tamaño más pequeño que el tamaño de un microrreactor (definido más adelante). Los ejemplos de soportes sólidos para su uso con los procedimientos descritos en el presente documento incluyen, sin limitación, perlas (por ejemplo, micropérlas, nanopérlas) y partículas (por ejemplo, micropartículas, nanopartículas).

Los términos "perlas" y "partículas" tal como se usan en el presente documento se refieren a soportes sólidos adecuados para asociarse con biomoléculas, y más específicamente con ácidos nucleicos. Las perlas pueden tener una forma regular (por ejemplo, esferoide, ovoide) o irregular (por ejemplo, rugosa, dentada), y a veces no son esféricas (por ejemplo, angulares, de múltiples lados). Pueden utilizarse partículas o perlas que tienen un diámetro nominal, promedio o medio menor que el diámetro nominal, promedio, medio o mínimo de un microrreactor. Pueden utilizarse partículas o perlas que tienen un diámetro nominal, promedio o medio de aproximadamente 1 nanómetro a aproximadamente 500 micrómetros, tales como las que tienen un diámetro nominal, medio o promedio, por ejemplo, de aproximadamente 10 nanómetros a aproximadamente 100 micrómetros; de aproximadamente 100 nanómetros a aproximadamente 100 micrómetros; de aproximadamente 1 micrómetro a aproximadamente 100 micrómetros; de aproximadamente 10 micrómetros a aproximadamente 50 micrómetros; de aproximadamente 1, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 200, 300, 400, 500, 600, 700, 800 ó 900 nanómetros; o de aproximadamente 1, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 200, 300, 400, 500 micrómetros.

Una perla o partícula puede fabricarse prácticamente de cualquier material insoluble o sólido. Por ejemplo, la perla o partícula puede comprender o consistir esencialmente en gel de sílice, vidrio (por ejemplo, vidrio de poro controlado (CPG)), nailon, Sephadex®, Sepharose®, celulosa, una superficie metálica (por ejemplo, acero, oro, plata, aluminio, silicio y cobre), un material magnético, un material plástico (por ejemplo, polietileno, polipropileno, poliamida, poliéster, poli(difluoruro de vinilideno) (PVDF)), y similares. Las perlas o partículas pueden ser hinchables (por ejemplo, perlas poliméricas tales como resina de Wang) o no hinchables (por ejemplo, CPG). Los ejemplos de perlas disponibles comercialmente incluyen sin limitación resina de Wang, resina de Merrifield y Dynabeads®. Las perlas

también pueden fabricarse como partículas sólidas o partículas que contienen huecos internos.

A menudo, los soportes sólidos adecuados para su uso con los procedimientos de análisis de secuencias descritos en el presente documento están en asociación con un ácido nucleico denominado en el presente documento "ácido nucleico en fase sólida". El término "ácido nucleico en fase sólida" tal como se usa en el presente documento generalmente se refiere a una o más especies de ácido nucleico diferentes en asociación con un soporte sólido. Una "especie de ácido nucleico" en fase sólida tal como se usa en el presente documento se refiere a un primer ácido nucleico que tiene una secuencia de nucleótidos que difiere en una base nucleotídica o más de la secuencia de nucleótidos de un segundo ácido nucleico cuando se alinean las secuencias de nucleótidos de los ácidos nucleicos primero y segundo. Por tanto, una especie de ácido nucleico puede diferir de una segunda especie de ácido nucleico en uno o más nucleótidos cuando las secuencias de nucleótidos de los ácidos nucleicos primero y segundo se alinean entre sí (por ejemplo, aproximadamente 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100 o más de 100 diferencias de nucleótidos).

Un soporte sólido puede proporcionarse en una colección de soportes sólidos. Una colección de soportes sólidos puede comprender dos o más especies de soporte sólido diferentes. El término "especie de soporte sólido" tal como se usa en el presente documento se refiere a un soporte sólido en asociación con una especie de ácido nucleico en fase sólida particular o una combinación particular de especies de ácido nucleico en fase sólida diferentes. En determinadas implementaciones, una colección de soportes sólidos comprende de 2 a 10.000 especies de soporte sólido, de 10 a 1.000 especies de soporte sólido o aproximadamente 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, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000 ó 10000 especies de soporte sólido singulares. Los soportes sólidos (por ejemplo, perlas) en la colección de soportes sólidos pueden ser homogéneos (por ejemplo, todos son perlas de resina de Wang) o heterogéneos (por ejemplo, algunos son perlas de resina de Wang y algunos son perlas magnéticas).

Un ácido nucleico en fase sólida generalmente es monocatenario y es de cualquier tipo adecuado para hibridarse con un ácido nucleico de muestra (por ejemplo, ADN, ARN, análogos de los mismos (por ejemplo, ácido nucleico peptídico (ANP)), quimeras de los mismos (por ejemplo, una única cadena comprende bases de ARN y bases de ADN), y similares). Un ácido nucleico en fase sólida está asociado con el soporte sólido de cualquier manera adecuada para la hibridación del ácido nucleico en fase sólida con el ácido nucleico de muestra. Un ácido nucleico en fase sólida puede estar en asociación con un soporte sólido mediante un enlace covalente o una interacción no covalente. Los ejemplos no limitativos de interacciones no covalentes incluyen interacciones hidrófobas (por ejemplo, soporte sólido recubierto con C18 y ácido nucleico tritilado), interacciones polares (por ejemplo, asociación de "humectación" entre ácido nucleico/polietilenglicol), interacciones de pares incluyendo, sin limitación, anticuerpo/antígeno, anticuerpo/anticuerpo, anticuerpo/fragmento de anticuerpo, anticuerpo/receptor de anticuerpo, anticuerpo/proteína A o proteína G, hapteno/anti-hapteno, biotina/avidina, biotina/estreptavidina, ácido fólico/proteína de unión a folato, vitamina B12/factor intrínseco, ácido nucleico/ácido nucleico complementario (por ejemplo, ADN, ARN, ANP), y similares.

Un ácido nucleico en fase sólida puede estar asociado con un soporte sólido mediante una metodología diferente, que incluye, sin limitación (i) sintetizar secuencialmente el ácido nucleico directamente sobre un soporte sólido, y (ii) sintetizar el ácido nucleico, proporcionar el ácido nucleico en fase de disolución y unir el ácido nucleico a un soporte sólido. Un ácido nucleico en fase sólida puede unirse covalentemente en diversos sitios en el ácido nucleico al soporte sólido, tal como (i) en la posición 1', 2', 3', 4' ó 5' de un resto de azúcar o (ii) un resto de base de pirimidina o purina de un nucleótido terminal o no terminal del ácido nucleico, por ejemplo. En determinadas implementaciones, el nucleótido 5'-terminal del ácido nucleico en fase sólida está unido al soporte sólido.

Se conocen métodos para sintetizar secuencialmente un ácido nucleico directamente sobre un soporte sólido. Por ejemplo, puede unirse el extremo 3' del ácido nucleico al soporte sólido (por ejemplo, método de fosforamidito descrito en Caruthers, Science 230: 281-286 (1985)) o puede unirse el extremo 5' del ácido nucleico al soporte sólido (por ejemplo, Claeboe *et al.*, Nucleic Acids Res. 31 (19): 5685-5691 (2003)).

También se conocen métodos para unir un ácido nucleico en fase de disolución a un soporte sólido (por ejemplo, la patente estadounidense n.º 6.133.436, a nombre de Koster *et al.* y titulada "Beads bound to a solid support and to nucleic acids" y el documento WO 91/08307, a nombre de Van Ness y titulado "Enhanced capture of target nucleic acid by the use of oligonucleotides covalently attached to polymers"). Los ejemplos incluyen, sin limitación, enlaces de tioéter (por ejemplo, ácido nucleico tiolado); enlaces de disulfuro (por ejemplo, perlas de tiol, ácido nucleico tiolado); enlaces de amida (por ejemplo, resina de Wang, ácido nucleico unido a amino); enlaces lábiles a los ácidos (por ejemplo, perlas de vidrio, ácido nucleico tritilado), y similares. Un ácido nucleico puede unirse a un soporte sólido sin ningún ligador o con un ligador (por ejemplo, S. S. Wong, "Chemistry of Protein Conjugation and Cross-Linking", CRC Press (1991), y G. T. Hermanson, "Bioconjugate Techniques", Academic Press (1995)). Puede seleccionarse un reactivo ligador homobiofuncional o heterobiofuncional, y los ejemplos de ligadores incluyen sin limitación N-succinimidil(4-yodoacetil)aminobenzoato (SIAB), dimaleimida, ácido ditio-bis-nitrobenzoico (DTNB), N-succinimidil-S-acetil-tioacetato (SATA), N-succinimidil-3-(2-piridilditio)propionato (SPDP), 4-(N-maleimidometil)ciclohexano-1-carboxilato de succinimidilo (SMCC), 6-hidrazinonicotimida (HYNIC), ácido 3-amino-(2-nitrofenil)propiónico, y similares.

Un ácido nucleico puede sintetizarse usando métodos y equipos convencionales, tales como el sintetizador de ADN de alto rendimiento ABI®3900 y el sintetizador de ácidos nucleicos EXPEDITE®8909, ambos de los cuales están disponibles de Applied Biosystems (Foster City, CA). Se describen análogos y derivados en las patentes estadounidenses n.ºs 4.469.863; 5.536.821; 5.541.306; 5.637.683; 5.637.684; 5.700.922; 5.717.083; 5.719.262; 5.739.308; 5.773.601; 5.886.165; 5.929.226; 5.977.296; 6.140.482; en el documento WO 00/56746; en el documento WO 01/14398, y en publicaciones relacionadas. Se dan a conocer métodos para sintetizar ácidos nucleicos que comprenden tales análogos o derivados, por ejemplo, en las publicaciones de patente citadas anteriormente y en las patentes estadounidenses n.ºs 5.614.622; 5.739.314; 5.955.599; 5.962.674; 6.117.992; en el documento WO 00/75372, y en publicaciones relacionadas. En determinadas implementaciones, los ácidos nucleicos análogos incluyen inosinas, sitios abásicos, ácidos nucleicos bloqueados, agentes de unión al surco menor, estabilizadores de dúplex (por ejemplo, acridina, espermidina) y/u otros modificadores de la temperatura de fusión (por ejemplo, un ácido nucleico diana, un ácido nucleico en fase sólida y/o un ácido nucleico de cebador pueden comprender un análogo).

Puede seleccionarse la densidad de las moléculas de ácido nucleico en fase sólida por unidad de soporte sólido (por ejemplo, una perla). Puede seleccionarse una densidad máxima que permita la hibridación del ácido nucleico de muestra con el ácido nucleico en fase sólida. En determinadas implementaciones, la densidad de ácido nucleico en fase sólida por unidad de soporte sólido (por ejemplo, moléculas de ácido nucleico por perla) es de aproximadamente 5 ácidos nucleicos a aproximadamente 10.000 ácidos nucleicos por soporte sólido. En algunas implementaciones, la densidad del ácido nucleico en fase sólida por unidad de soporte sólido es de aproximadamente 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000 ó 10000 ácidos nucleicos por soporte sólido. En determinadas implementaciones, la densidad del ácido nucleico en fase sólida por unidad de soporte sólido es de aproximadamente 1 a 1 (por ejemplo, una molécula de ácido nucleico en fase sólida por una perla).

En algunas implementaciones, un ácido nucleico en fase sólida comprende determinadas subsecuencias. Una subsecuencia puede ser complementaria a o sustancialmente complementaria a una subsecuencia de nucleótidos de ácido nucleico de muestra y permite la hibridación de un ácido nucleico en fase sólida con un ácido nucleico de muestra. Una subsecuencia de este tipo (por ejemplo, ilustrada en las figuras 1A-1C) se denomina en el presente documento secuencia de "sonda", y un soporte sólido puede contener una o más especies de secuencia de sonda. Una "especie de secuencia de sonda" tal como se usa en el presente documento se refiere a una primera secuencia de nucleótidos de sonda que difiere en una base nucleotídica o más de una segunda secuencia de nucleótidos de sonda cuando se alinean las secuencias de nucleótidos de sonda primera y segunda. Por tanto, una especie de secuencia de sonda puede diferir de una segunda especie de secuencia de sonda en uno o más nucleótidos cuando las secuencias de sonda primera y segunda se alinean entre sí (por ejemplo, aproximadamente 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100 o más de 100 nucleótidos que no son idénticos tras la alineación). Se conocen técnicas de alineación y metodologías de evaluación de identidad de secuencia (por ejemplo, algoritmo de Meyers y Miller, CABIOS 4: 11-17 (1989), que se ha incorporado en el programa ALIGN (versión 2.0)).

Una secuencia de nucleótidos de sonda tiene una longitud suficiente para hibridarse específicamente con una secuencia de nucleótidos de ácido nucleico de muestra. En determinadas implementaciones, una secuencia de sonda tiene una longitud de aproximadamente 5 a aproximadamente 100 nucleótidos, y a veces tiene una longitud de aproximadamente 5 a aproximadamente 40 nucleótidos. Generalmente, se selecciona una secuencia de sonda más corta para aplicaciones en las que la secuencia de nucleótidos diana es conocida o parcialmente conocida y se selecciona una secuencia de sonda más larga para aplicaciones en las que la secuencia de nucleótidos diana o porciones de la misma son desconocidas. En algunas implementaciones, una secuencia de sonda tiene una longitud de aproximadamente 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 2000, 3000, 4000 ó 5000 nucleótidos.

En algunas implementaciones, un ácido nucleico en fase sólida de una especie de soporte sólido, o una colección de especies de soporte sólido, puede incluir cualquier número de especies de secuencia de sonda útil para llevar a cabo los procedimientos de análisis de secuencias proporcionados en el presente documento. En determinadas implementaciones, un soporte sólido comprende de aproximadamente 10 a aproximadamente 10.000 secuencias de sonda singulares (por ejemplo, aproximadamente 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000 o aproximadamente 10.000 especies de secuencia de sonda diferentes); un soporte sólido comprende de aproximadamente 10 a aproximadamente 1.000 secuencias de sonda singulares (por ejemplo, aproximadamente 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900 o aproximadamente 1.000 especies de secuencia de sonda diferentes); una colección de soportes sólidos comprende de aproximadamente 10 a aproximadamente 10.000 secuencias de sonda singulares (por ejemplo, aproximadamente 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000 o aproximadamente 10.000 especies de secuencia de sonda diferentes); y una colección de soportes sólidos comprende de aproximadamente 10 a aproximadamente 1.000 secuencias de sonda singulares (por ejemplo, aproximadamente 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900 o aproximadamente 1.000 especies de secuencia de sonda diferentes). En



algunas implementaciones, se utilizan menos especies de secuencia de sonda por soporte sólido o por colección de soportes sólidos (por ejemplo, para aplicaciones de haplotipado) y a veces se utilizan mayores números de especies de secuencia de sonda por soporte sólido o por colección de soportes sólidos (por ejemplo, para aplicaciones de secuenciación). En determinadas implementaciones, un soporte sólido, o una colección de soportes sólidos, incluye aproximadamente 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000 ó 10000 especies de secuencia de sonda singulares. En determinadas implementaciones, un ácido nucleico en fase sólida de una especie de soporte sólido comparte únicamente una especie de secuencia de sonda, y en implementaciones de colección relacionadas, un ácido nucleico en fase sólida de cada especie de soporte sólido en la colección comparte únicamente una especie de secuencia de sonda (es decir, una especie de secuencia de sonda por especie de soporte sólido).

En determinadas implementaciones, un ácido nucleico en fase sólida también puede contener una secuencia de identificación (por ejemplo, ilustrada en las figuras 1A-1C), que puede ser útil en parte para construir lecturas parciales de secuencia en construcciones de secuencias más grandes en determinadas implementaciones. Una secuencia de identificación puede ser "singular" para cada especie de soporte sólido, en la que el término "singular" tal como se usa en el presente documento se refiere a que hay una especie de secuencia de identificación para cada especie de soporte sólido. Una "especie de secuencia de identificación" tal como se usa en el presente documento se refiere a una primera secuencia de nucleótidos de identificación que difiere en una base nucleotídica o más de una segunda secuencia de nucleótidos de identificación cuando se alinean las secuencias de nucleótidos de identificación primera y segunda. Por tanto, una especie de secuencia de identificación puede diferir de una segunda especie de secuencia de identificación en uno o más nucleótidos cuando las secuencias de identificación primera y segunda se alinean entre sí (por ejemplo, aproximadamente 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100 o más de 100 nucleótidos que no son idénticos tras la alineación). En algunas implementaciones, una secuencia de identificación puede detectarse mediante una propiedad seleccionada del grupo que consiste en tamaño, forma, propiedades eléctricas, propiedades magnéticas, propiedades ópticas, propiedades químicas, y similares.

Una secuencia de identificación puede tener cualquier longitud adecuada para analizar la secuencia de nucleótidos o una secuencia de nucleótidos parcial de un ácido nucleico de muestra. En algunas implementaciones, una secuencia identificadora tiene una longitud de aproximadamente 5 a aproximadamente 50 nucleótidos contiguos, a veces una longitud de aproximadamente 5 a aproximadamente 20 nucleótidos y a veces una longitud de aproximadamente 10 nucleótidos. En determinadas implementaciones, una secuencia identificadora tiene una longitud de aproximadamente 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 35, 40, 45 ó 50 nucleótidos.

En algunas implementaciones, un ácido nucleico en fase sólida a menudo incluye una secuencia de cebador (Pr), que también se denomina en el presente documento "secuencia de hibridación de cebador". La secuencia de cebador (Pr) puede hibridarse con una secuencia de nucleótidos complementaria en un ácido nucleico de cebador que puede utilizarse para amplificar un ácido nucleico en fase sólida extendido previamente hibridado con un ácido nucleico de muestra. Tal como se usa en el presente documento, el término "ácido nucleico de cebador" se refiere a un ácido nucleico (por ejemplo, que se produce de manera natural o sintético) que tiene una secuencia de ácido nucleico complementaria a una secuencia de hibridación de cebador, y puede hibridarse con la secuencia de hibridación de cebador en condiciones de hibridación y puede extenderse en un procedimiento de amplificación (por ejemplo, extensión de cebador, amplificación por PCR, y similares). Los ácidos nucleicos de cebador pueden tener cualquier longitud adecuada para una hibridación optimizada y pueden estar en el intervalo de aproximadamente 5 nucleótidos a aproximadamente 5000 nucleótidos de longitud (por ejemplo, aproximadamente 5 nucleótidos, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1500, 2000, 2500, 3000, 3500, 4000, 4500 o aproximadamente 5000 nucleótidos de longitud). En algunas implementaciones, los ácidos nucleicos de cebador pueden estar modificados, lo que puede efectuarse mediante un procedimiento de modificación incluyendo, sin limitación, modificación de codones, síntesis usando análogos de nucleótidos, modificación posterior a la síntesis, y similares.

En una colección de soportes sólidos, cada ácido nucleico en fase sólida en cada especie de soporte sólido puede tener una secuencia de cebador común (por ejemplo, todas las especies de soporte sólido tienen la misma especie de secuencia de cebador), en cuyo caso la secuencia de cebador se denomina secuencia de cebador "universal" o "común". En determinadas implementaciones, un ácido nucleico en fase sólida de una primera especie de soporte sólido en una colección puede tener una primera especie de secuencia de cebador y un ácido nucleico en fase sólida de una segunda especie de soporte sólido en la colección puede tener una segunda especie de secuencia de cebador. Una "especie de secuencia de cebador" tal como se usa en el presente documento se refiere a una primera secuencia de nucleótidos de cebador que difiere en una base nucleotídica o más de una segunda secuencia de nucleótidos de cebador cuando se alinean las secuencias de nucleótidos de cebador primera y segunda. Por tanto, una especie de secuencia de cebador puede diferir de una segunda especie de secuencia de cebador en uno o más nucleótidos cuando las secuencias de cebador primera y segunda se alinean entre sí (por ejemplo, aproximadamente 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100 o más de 100 nucleótidos que no son idénticos tras la alineación).

En algunas implementaciones, una secuencia de hibridación de cebador (Pr) puede tener una longitud que permita

la hibridación específica de un cebador en las condiciones para la hibridación de cebador. En determinadas implementaciones, la longitud de la secuencia de hibridación de cebador (Pr) es de aproximadamente 10 a aproximadamente 100 nucleótidos, de aproximadamente 10 a aproximadamente 50 nucleótidos o de aproximadamente 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95 ó 100 nucleótidos de longitud. En determinadas implementaciones, un ácido nucleico de muestra de una especie de soporte sólido, o colección de soportes sólidos, incluye una o más especies de secuencia de cebador (por ejemplo, aproximadamente 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45 ó 50 especies de secuencia de hibridación de cebador singulares). Por tanto, en determinadas implementaciones, un ácido nucleico en fase sólida de una especie de soporte sólido incluye una especie de secuencia de hibridación de cebador, y un ácido nucleico de una colección de soportes sólidos comparte una especie de secuencia de hibridación de cebador común.

En las figuras 1A, 1B y 1C se muestran ejemplos no limitativos de especies de soporte sólido diferentes que tienen, cada una, una especie de secuencia de sonda, una especie de secuencia de identificación y una especie de secuencia de hibridación de cebador. La figura 1A muestra una especie de soporte sólido que tiene una combinación particular de ácido nucleico en fase sólida que difiere en la especie de secuencia de sonda  $P_1, P_2, P_3, \dots P_n$ . En la figura 1A, las especies de secuencia de sonda son complementarias a las subsecuencias en un ácido nucleico de muestra (por ejemplo, las especies de secuencia de sonda  $P_1, P_2, P_3, \dots P_n$  son complementarias a las subsecuencias de ácido nucleico de muestra 1, 2, 3, ..., N, respectivamente). La figura 1B muestra una colección de tres especies de soporte sólido, en la que cada especie de soporte sólido incluye un ácido nucleico en fase sólida que tiene una secuencia de identificación singular y especies de secuencia de sonda diferentes. En la figura 1B, un ácido nucleico en fase sólida de la especie de soporte sólido X tiene las especies de secuencia de sonda  $P_{X1}, P_{X2}, P_{X3}, \dots P_{Xn_i}$ ; un ácido nucleico en fase sólida de la especie de soporte sólido Y tiene las especies de secuencia de sonda  $P_{Y1}, P_{Y2}, P_{Y3}, \dots P_{Yn_i}$ ; y un ácido nucleico en fase sólida de la especie de soporte sólido Z tiene las especies de secuencia de sonda  $P_{Z1}, P_{Z2}, P_{Z3}, \dots P_{Zn}$ . La figura 1C muestra una colección de tres especies de soporte sólido, en la que cada especie de soporte sólido incluye un ácido nucleico en fase sólida que tiene una secuencia de identificación singular y la misma especie de secuencia de sonda.

Las secuencias de sonda, de identificación y de hibridación de cebador en un ácido nucleico en fase sólida pueden estar dispuestas en cualquier orientación adecuada unas con respecto a otras para realizar los métodos descritos en el presente documento. Cualesquiera dos de estas secuencias pueden ser contiguas o pueden estar separadas por una secuencia intermedia de una longitud adecuada (por ejemplo, una secuencia intermedia de aproximadamente 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100 o más de 100 nucleótidos). En determinadas implementaciones, las secuencias son contiguas y están en la siguiente orientación: 5' - (secuencia de cebador) - (secuencia de identificación) - (secuencia de sonda) - 3'.

Los soportes sólidos que tienen ácidos nucleicos en fase sólida pueden proporcionarse en cualquier forma conveniente para ponerse en contacto con un ácido nucleico de muestra, tal como forma sólida o líquida, por ejemplo. En determinadas implementaciones, un soporte sólido puede proporcionarse en forma líquida que comprende opcionalmente uno o más de otros componentes, que incluyen sin limitación uno o más tampones o sales. Los soportes sólidos de una colección pueden proporcionarse en un recipiente, pueden distribuirse en múltiples recipientes.

En determinadas implementaciones, los soportes sólidos pueden proporcionarse en una matriz, o pueden proporcionarse instrucciones para disponer los soportes sólidos en una matriz sobre un sustrato. El término "matriz" tal como se usa en el presente documento puede referirse a una disposición de ubicaciones de muestra en un único soporte sólido bidimensional o a una disposición de soportes sólidos en una superficie bidimensional. Una matriz puede tener cualquier forma general conveniente (por ejemplo, circular, ovalada, cuadrada, rectangular). Una matriz puede denominarse "matriz de X por Y" para matrices cuadradas o rectangulares, en la que la matriz incluye un número X de soportes sólidos o ubicaciones de muestra en una dimensión y un número Y de soportes sólidos o ubicaciones de muestra en una dimensión perpendicular. Una matriz puede ser simétrica (por ejemplo, una matriz de 16 por 16) o asimétrica (por ejemplo, una matriz de 8 por 16). Una matriz puede incluir cualquier número conveniente de soportes sólidos o ubicaciones de muestra en cualquier disposición adecuada. Por ejemplo, en algunas implementaciones, X o Y puede ser independientemente 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29 ó 30.

Una matriz puede contener una especie de soporte sólido o múltiples especies de soporte sólido a partir de una colección. La matriz puede disponerse sobre cualquier sustrato adecuado para los procedimientos de fabricación o análisis de secuencias descritos en el presente documento. Los ejemplos de sustratos incluyen, sin limitación, sustratos planos, sustratos de filtro, sustratos de oblea, sustratos grabados, sustratos que tienen múltiples pocillos u orificios (por ejemplo, pocillos u orificios de microlitros (volumen de aproximadamente 1 microlitro, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 300, 400, 500, 600, 700, 800, 900 y hasta aproximadamente 999 microlitros), de nanolitros (volumen de 1 nanolitro, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 300, 400, 500, 600, 700, 800, 900 y hasta aproximadamente 999 nanolitros), de picolitros (volumen de 1 picolitro, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 300, 400, 500, 600, 700, 800, 900 y hasta aproximadamente 999 picolitros); pocillos que tienen

fondos de filtro), sustratos que tienen uno o más canales, sustratos que tienen uno o más electrodos, y similares, y combinaciones de los mismos. Los pocillos u orificios de sustratos de múltiples pocillos y orificios pueden contener una o más unidades de soporte sólido (por ejemplo, siendo cada unidad una única perla o partícula). Los sustratos pueden comprender o consistir esencialmente en un material adecuado para llevar a cabo los procedimientos de fabricación o análisis de secuencias descritos en el presente documento, incluyendo sin limitación fibra (por ejemplo, filtros de fibra), vidrio (por ejemplo, superficies de vidrio, superficies de fibra óptica), metal (por ejemplo, acero, oro, plata, aluminio, silicio y cobre; recubrimiento metálico), plástico (por ejemplo, polietileno, polipropileno, poliamida, poli(difluoruro de vinilideno)), silicio, y similares. En determinadas implementaciones, la matriz puede ser una micromatriz o una nanomatriz. Una "nanomatriz" a menudo es una matriz en la que las unidades de soporte sólido están separadas por de aproximadamente 0,1 nanómetros a aproximadamente 10 micrómetros, por ejemplo, desde aproximadamente 1 nanómetro hasta aproximadamente 1 micrómetro (por ejemplo, aproximadamente 0,1 nanómetros, 0,5, 1, 2, 3, 4, 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900 nanómetros, 1 micrómetro, 2, 3, 4, 5, 6, 7, 8, 9 y hasta aproximadamente 10 micrómetros). Una "micromatriz" es una matriz en la que las unidades de soporte sólido están separadas por más de 1 micrómetro. La densidad de las unidades de soporte sólido en las matrices a menudo es de al menos 100/cm<sup>2</sup>, y puede ser de 100/cm<sup>2</sup> a aproximadamente 10.000/cm<sup>2</sup>, de 100/cm<sup>2</sup> a aproximadamente 1.000/cm<sup>2</sup> o de aproximadamente 150, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000 ó 10000 unidades de soporte sólido/cm<sup>2</sup>.

#### Moléculas únicas de ácido nucleico de muestra

En determinados métodos descritos en el presente documento, un ácido nucleico de muestra y un soporte sólido se ponen en contacto en condiciones en las que una única molécula de ácido nucleico de muestra se hibrida con una única molécula de un soporte sólido. Es decir, en algunas implementaciones, las condiciones de hibridación pueden optimizarse para permitir una única molécula de ácido nucleico de muestra por soporte sólido (por ejemplo, perla o partícula), o para permitir que se hibride más de una especie de ácido nucleico de muestra por soporte sólido (por ejemplo, las perlas o partículas están configuradas para tener más de una especie de secuencia de cebador, secuencia de identificación, secuencia de sonda, o combinaciones de las mismas). En algunas implementaciones, puede hibridarse una única molécula de ácido nucleico muestra por soporte sólido en condiciones de concentración de ADN diluido en las que se favorece la hibridación de tan sólo una molécula de ácido nucleico de muestra por perla. En algunas implementaciones, las condiciones de hibridación pueden estar configuradas para incluir sólo una molécula de ácido nucleico de muestra en la etapa de hibridación. En determinadas implementaciones, tales condiciones pueden incluir proporcionar las moléculas de soporte sólido y una única molécula de ácido nucleico de muestra en un "microrreactor". Tales condiciones también incluyen proporcionar una mezcla en la que la molécula de ácido nucleico de muestra puede hibridarse con un ácido nucleico en fase sólida sobre el soporte sólido.

Tal como se usa en el presente documento, el término "microrreactor" se refiere a un espacio dividido en el que una única molécula de ácido nucleico de muestra puede hibridarse con una molécula de soporte sólido. En algunas implementaciones, el volumen de microrreactor es lo suficientemente grande como para alojar una perla de soporte sólido en el microrreactor y lo suficientemente pequeño como para excluir la presencia de dos o más perlas en el microrreactor. Los ejemplos de microrreactores incluyen sin limitación un glóbulo en emulsión (descrito más adelante) y un hueco en un sustrato. En determinadas implementaciones, un hueco en un sustrato puede ser un orificio, un poro o un pocillo (por ejemplo, micropocillo, nanopocillo, picopocillo, microporo o nanoporo) en un sustrato construido a partir de material sólido útil para contener fluidos (por ejemplo, plástico (por ejemplo, polipropileno, polietileno, poliestireno) o silicio). Los glóbulos en emulsión están divididos por una fase inmiscible tal como se describe con mayor detalle más adelante. Una única molécula de ácido nucleico de muestra puede proporcionarse en un microrreactor poniendo en contacto moléculas de ácido nucleico de muestra con un exceso (por ejemplo, exceso molar) de moléculas de soporte sólido. En determinadas implementaciones, la cantidad (por ejemplo, cantidad molar) en exceso de soporte sólido es de aproximadamente 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 150, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10000 veces, o más, la cantidad de ácido nucleico de muestra.

El término "emulsión" tal como se usa en el presente documento se refiere a una mezcla de dos sustancias inmiscibles y no mezclables, en la que una sustancia (la fase dispersa) a menudo se dispersa en la otra sustancia (la fase continua). En determinadas implementaciones, la fase dispersa puede ser una disolución acuosa (es decir, una disolución que comprende agua). En algunas implementaciones, la fase dispersa está compuesta predominantemente por agua (por ejemplo, más del 70 %, más del 75 %, más del 80 %, más del 85 %, más del 90 %, más del 95 %, más del 97 %, más del 98 % y más del 99 % de agua (en peso)). Cada porción discreta de una fase dispersa, tal como una fase dispersa acuosa, se denomina en el presente documento "glóbulo" o "microrreactor". En determinadas implementaciones, un glóbulo a veces puede tener una forma esferoide, sustancialmente esferoide o semiesferoide.

Los términos "aparato de emulsión" y "componente(s) de emulsión" tal como se usan en el presente documento se refieren a un aparato y a componentes que pueden usarse para preparar una emulsión. Los ejemplos no limitativos de un aparato de emulsión incluyen sin limitación un aparato de flujo en contracorriente, de corriente cruzada, de tambor giratorio y de membrana adecuado para su uso para preparar una emulsión. En determinadas

implementaciones, un componente de emulsión forma la fase continua de una emulsión, e incluye sin limitación una sustancia inmiscible con agua, tal como un componente que comprende o que consiste esencialmente en un aceite (por ejemplo, un aceite biocompatible termoestable (por ejemplo, aceite mineral ligero)). Puede utilizarse un estabilizador de emulsión biocompatible como componente de emulsión. Los estabilizadores de emulsión incluyen sin limitación Atlox 4912, Span 80 y otros tensioactivos biocompatibles.

En algunas implementaciones, pueden incluirse componentes útiles para reacciones biológicas en la fase dispersa. Los glóbulos de la emulsión pueden incluir (i) una unidad de soporte sólido (por ejemplo, una perla o una partícula); (ii) una molécula de ácido nucleico de muestra; y (iii) una cantidad suficiente de agentes de extensión para alargar el ácido nucleico en fase sólida y amplificar el ácido nucleico en fase sólida alargado (por ejemplo, nucleótidos de extensión, polimerasa, cebador). Los glóbulos inactivos en la emulsión pueden incluir un subconjunto de estos componentes (por ejemplo, soporte sólido y reactivos de extensión y ningún ácido nucleico de muestra) y algunos pueden estar vacíos (es decir, algunos glóbulos no incluirán soporte sólido, ácido nucleico de muestra ni agentes de extensión).

Las emulsiones pueden prepararse usando métodos adecuados (por ejemplo, Nakano *et al.* "Single-molecule PCR using water-in-oil emulsion"; Journal of Biotechnology 102 (2003) 117-124). Los métodos de emulsificación incluyen sin limitación métodos de adyuvante, métodos de flujo en contracorriente, métodos de corriente cruzada, métodos de tambor giratorio, métodos de membrana, y similares. En determinadas implementaciones, se prepara una mezcla de reacción acuosa que contiene un soporte sólido (a continuación en el presente documento "mezcla de reacción") y luego se añade a un aceite biocompatible. En determinadas implementaciones, la mezcla de reacción puede contener (i) un soporte sólido o una colección de soportes sólidos; (ii) un ácido nucleico de muestra; (iii) agentes de extensión y (iv) uno o más cebadores. Cada uno de estos componentes puede mezclarse en cualquier orden adecuado para preparar la mezcla de reacción. En determinadas implementaciones, la mezcla de reacción puede añadirse gota a gota a una mezcla con agitación de aceite biocompatible (por ejemplo, aceite mineral ligero (Sigma)) y dejarse emulsionar. En algunas implementaciones, la mezcla de reacción puede añadirse gota a gota a un flujo cruzado de aceite biocompatible. El tamaño de los glóbulos acuosos en la emulsión puede ajustarse, tal como variando el caudal y la velocidad a la que se añaden los componentes entre sí, por ejemplo.

En determinadas implementaciones, el tamaño de los glóbulos en emulsión puede seleccionarse basándose en dos factores competitivos: (i) los glóbulos son lo suficientemente grandes como para alojar una molécula de soporte sólido, una molécula de ácido nucleico de muestra y agentes de extensión suficientes para el grado de alargamiento y amplificación requerido; y (ii) los glóbulos son lo suficientemente pequeños de modo que una población de glóbulos puede amplificarse mediante equipos de laboratorio convencionales (por ejemplo, equipos de termociclado, tubos de ensayo, incubadoras, y similares). En determinadas implementaciones, los glóbulos en la emulsión pueden tener un diámetro nominal, medio o promedio de aproximadamente 5 micras a aproximadamente 500 micras, de aproximadamente 10 micras a aproximadamente 350 micras, de aproximadamente 50 a 250 micras, de aproximadamente 100 micras a aproximadamente 200 micras, o de aproximadamente 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 200, 300, 400 ó 500 micras.

El ácido nucleico de muestra, el/los soporte(s) sólido(s), los agentes de extensión y el/los componente(s) de emulsión pueden mezclarse de cualquier manera adecuada y en cualquier razón adecuada para llevar a cabo los métodos descritos en el presente documento, incluyendo sin limitación medios manuales y automatizados (por ejemplo, estaciones de trabajo biológicas). Puede utilizarse cualquier razón adecuada de soporte sólido con respecto a ácido nucleico de muestra para obtener glóbulos que tienen un ácido nucleico de muestra por unidad de soporte sólido, y en algunas implementaciones, una razón de concentración de soporte sólido con respecto a concentración de ácido nucleico de muestra es igual o superior a 10:1, y en algunas implementaciones, la razón es de aproximadamente 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000 ó 10000, con respecto a 1. En algunas implementaciones, (a) el ácido nucleico de muestra puede ponerse en contacto con un soporte sólido o una colección de soportes sólidos en condiciones en las que el ácido nucleico de muestra puede hibridarse con el ácido nucleico en fase sólida, (b) la mezcla de (a) puede ponerse en contacto con agentes de extensión, y (c) la mezcla de (b) puede emulsionarse con una disolución inmiscible con agua (por ejemplo, un aceite biocompatible). En determinadas implementaciones, una emulsión puede prepararse contemporáneamente con la puesta en contacto de la mezcla con los agentes de extensión.

Se conocen condiciones de hibridación que permiten la hibridación de un ácido nucleico de muestra con un ácido nucleico en fase sólida. Los ejemplos no limitativos de condiciones de hibridación incluyen sin limitación hibridación en cloruro de sodio 6X/citrato de sodio (SSC) a aproximadamente 45 °C, seguido de uno o más lavados en SSC 0,2X, SDS al 0,1 % a 50 °C. Otro ejemplo de condiciones de hibridación rigurosas son hibridación en cloruro de sodio 6X/citrato de sodio (SSC) a aproximadamente 45 °C, seguido de uno o más lavados en SSC 0,2X, SDS al 0,1 % a 55 °C, o cloruro de sodio 6X/citrato de sodio (SSC) a aproximadamente 45 °C, seguido de uno o más lavados en SSC 0,2X, SDS al 0,1 % a 60 °C. Las condiciones de hibridación rigurosas a veces son hibridación en cloruro de sodio 6X/citrato de sodio (SSC) a aproximadamente 45 °C, seguido de uno o más lavados en SSC 0,2X, SDS al 0,1 % a 65 °C. Las condiciones de rigurosidad a veces son fosfato de sodio 0,5 M, SDS al 7 % a 65 °C, seguido de uno o más lavados en SSC 0,2X, SDS al 1 % a 65 °C. Las condiciones de hibridación también se

describen, por ejemplo, en el documento WO 91/08307, titulado "Enhanced capture of target nucleic acid by the use of oligonucleotides covalently attached to polymers", a nombre de Van Ness, y "Nucleic Acid Hybridization, A Practical Approach", Ed. Hames and Higgins, IRL Press, 1985.

## 5 Amplificación

Los términos "agentes de extensión" y "reactivos de extensión" tal como se usan en el presente documento se refieren a componentes útiles para extender un ácido nucleico. Se conocen condiciones en las que los ácidos nucleicos pueden extenderse y/o amplificarse mediante tales agentes. En determinadas implementaciones, los agentes de extensión pueden incluir uno o más de los siguientes: nucleótidos de extensión, una polimerasa y un cebador que puede hibridarse con una secuencia de cebador en un ácido nucleico en fase sólida. Los nucleótidos de extensión incluyen, en algunas implementaciones, desoxinucleótido trifosfatos que se producen de manera natural (dATP, dTTP, dCTP, dGTP, dUTP) y nucleótidos que no se producen de manera natural o análogos de nucleótidos, tales como análogos que contienen un marcador detectable (por ejemplo, marcador fluorescente o colorimétrico), por ejemplo. Las polimerasas incluyen, en algunas implementaciones, polimerasas para amplificación por termociclado (por ejemplo, ADN polimerasa Taq; ADN polimerasa Taq Q-Bio™ (forma truncada recombinante de ADN polimerasa Taq que carece de actividad 5'-3'exo); polimerasa SurePrime™ (ADN polimerasa Taq modificada químicamente para PCR de "inicio en caliente"); ADN polimerasa Taq Arrow™ (amplificación con molde grande y de alta sensibilidad)) y polimerasas para amplificación termoestable (por ejemplo, ARN polimerasa para amplificación mediada por transcripción (TMA) descrita en la URL de la red informática mundial "gen-probe.com/pdfs/tma\_whiteppr.pdf"). Pueden añadirse otros componentes enzimáticos, tales como transcriptasa inversa para reacciones de TMA, por ejemplo.

Un ácido nucleico de cebador puede tener cualquier longitud adecuada para hibridarse con una secuencia de hibridación de cebador en un ácido nucleico en fase sólida y realizar los procedimientos de análisis de secuencias descritos en el presente documento. En algunas implementaciones, un cebador puede tener una longitud de aproximadamente 10 a aproximadamente 100 nucleótidos, de aproximadamente 10 a aproximadamente 70 nucleótidos, de aproximadamente 10 a aproximadamente 50 nucleótidos, de aproximadamente 15 a aproximadamente 30 nucleótidos, o de aproximadamente 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95 ó 100 nucleótidos. Un cebador puede estar compuesto por nucleótidos que se producen de manera natural y/o que no se producen de manera natural, o una mezcla de los mismos. Un cebador a menudo incluye una subsecuencia de nucleótidos que es complementaria a una secuencia de hibridación de cebador de ácido nucleico en fase sólida o sustancialmente complementaria a una secuencia de hibridación de cebador de ácido nucleico en fase sólida (por ejemplo, aproximadamente el 75 %, el 76 %, el 77 %, el 78 %, el 79 %, el 80 %, el 81 %, el 82 %, el 83 %, el 84 %, el 85 %, el 86 %, el 87 %, el 88 %, el 89 %, el 90 %, el 91 %, el 92 %, el 93 %, el 94 %, el 95 %, el 96 %, el 97 %, el 98 %, el 99 % o más del 99 % idéntica al complemento de secuencia de hibridación de cebador cuando se alinean). Un cebador puede contener una subsecuencia de nucleótidos no complementaria a o sustancialmente no complementaria a una secuencia de hibridación de cebador de ácido nucleico en fase sólida (por ejemplo, en el extremo 3' o 5' de la subsecuencia de nucleótidos en el cebador complementario a o sustancialmente complementario a la secuencia de hibridación de cebador en fase sólida). En determinadas implementaciones, un cebador puede contener una molécula detectable (por ejemplo, un fluoróforo, un radioisótopo, un agente colorimétrico, una partícula, una enzima, y similares).

En los procedimientos proporcionados en el presente documento, los componentes de un microrreactor pueden ponerse en contacto con agentes de extensión en condiciones de amplificación. El término "condiciones de amplificación" tal como se usa en el presente documento se refiere a condiciones de termociclado y termoestables que pueden facilitar la amplificación de un ácido nucleico. Pueden mantenerse las condiciones termoestables, y el tipo y la cantidad de amplificación generalmente depende de los agentes de extensión añadidos a la mezcla (por ejemplo, cebadores, ARN polimerasa y componentes de transcriptasa inversa para TMA (descrito anteriormente)). Las condiciones de termociclado generalmente implican repetir ciclos de fluctuación de temperatura, y hay disponibles aparatos para efectuar tales ciclos. Un ejemplo no limitativo de condiciones de termociclado es tratar la muestra a 95 °C durante 5 minutos; repetir cuarenta y cinco ciclos de 95 °C durante 1 minuto, 59 °C durante 1 minuto y 10 segundos, y 72 °C durante 1 minuto y 30 segundos; y luego tratar la muestra a 72 °C durante 5 minutos. Con frecuencia se realizan múltiples ciclos usando un termociclador disponible comercialmente (por ejemplo, aparato termociclador 2720 de Applied Biosystems). En determinadas implementaciones, una mezcla emulsionada puede someterse a condiciones de termociclado para la amplificación lineal usando un cebador que se hibrida con una secuencia de hibridación de cebador de ácido nucleico en fase sólida.

Un producto de amplificación para el análisis de señales puede tener cualquier longitud adecuada para los métodos de análisis de secuencias. En determinadas implementaciones, un producto de amplificación puede tener una longitud de aproximadamente 5 a aproximadamente 10.000 nucleótidos, una longitud de aproximadamente 10 a aproximadamente 1.000 nucleótidos, una longitud de aproximadamente 10 a aproximadamente 100 nucleótidos, una longitud de aproximadamente 10 a aproximadamente 50 nucleótidos, o una longitud de aproximadamente 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000 ó 10000 nucleótidos. Un producto de amplificación puede incluir nucleótidos que se producen de manera natural, nucleótidos que no se

producen de manera natural, análogos de nucleótidos, y similares y combinaciones de los anteriores. Un producto de amplificación a menudo tiene una secuencia de nucleótidos que es idéntica a o sustancialmente idéntica a una secuencia de nucleótidos de ácido nucleico de muestra o un complemento de la misma. Una secuencia de nucleótidos "sustancialmente idéntica" en un producto de amplificación tendrá generalmente un alto grado de identidad de secuencia con la secuencia de nucleótidos de muestra o un complemento de la misma (por ejemplo, aproximadamente el 75 %, el 76 %, el 77 %, el 78 %, el 79 %, el 80 %, el 81 %, el 82 %, el 83 %, el 84 %, el 85 %, el 86 %, el 87 %, el 88 %, el 89 %, el 90 %, el 91 %, el 92 %, el 93 %, el 94 %, el 95 %, el 96 %, el 97 %, el 98 %, el 99 % o más del 99 % de identidad para secuencia), y las variaciones a menudo serán el resultado de la infidelidad de la polimerasa usada para la extensión y/o amplificación.

En determinadas implementaciones, los productos de amplificación (por ejemplo, productos de amplificación de la figura 2) pueden ponerse en contacto con agentes de amplificación adicionales y/o someterse a condiciones de amplificación adicionales, tales como amplificación exponencial que implica más de un cebador y termociclado, por ejemplo. En determinadas implementaciones, puede utilizarse cualquier procedimiento de amplificación adecuado, tal como métodos de amplificación para su uso con la pirosecuenciación y secuenciación mediante las metodologías de ligación descritas más adelante, por ejemplo.

En determinadas implementaciones, los productos de amplificación lineal se analizan directamente sin amplificación adicional mediante otro procedimiento (por ejemplo, los productos de amplificación lineal de la figura 2 no se amplifican adicionalmente mediante un procedimiento de amplificación exponencial). En algunas implementaciones, el ácido nucleico de muestra, el ácido nucleico en fase sólida extendido y/o los productos de amplificación no están ligados a uno o más ácidos nucleicos heterólogos (por ejemplo, a diferencia de los métodos descritos en la publicación de solicitud de patente estadounidense n.º 20040110191, titulada "Comparative analysis of nucleic acids using population tagging", a nombre de Winkler *et al.*; la publicación de solicitud de patente estadounidense n.º 20050214825, titulada "Multiplex sample analysis on universal arrays", a nombre de Stuelpnagel; y Nakano *et al.* "Single-molecule PCR using water-in-oil emulsion"; Journal of Biotechnology 102 (2003) 117-124), tales como ácidos nucleicos heterólogos que se hibridan con un cebador de amplificación. Las secuencias de ácido nucleico de los productos de amplificación que no se amplifican adicionalmente pueden analizarse mediante un método de análisis de secuencias directo (por ejemplo, la metodología de secuenciación de molécula única descrita más adelante).

En determinadas implementaciones, los productos de amplificación pueden desprenderse a partir de un soporte sólido. En algunas implementaciones, puede utilizarse un método adecuado para desprender un producto de amplificación a partir de un soporte sólido, tal como calentando el soporte sólido (por ejemplo, calentando hasta aproximadamente 95 °C), exponiendo el soporte sólido a una cantidad de un caótropeo (por ejemplo, HCl de guanidinio) suficiente para desprender el ácido nucleico amplificado, y similares, por ejemplo.

La figura 2 ilustra una implementación de procedimiento en la que una molécula de ácido nucleico de muestra (S), proporcionada como fragmentos de ácido nucleico en algunas implementaciones, puede hibridarse con un soporte sólido descrito en el presente documento. La molécula de ácido nucleico de muestra (S) puede incluir una secuencia de nucleótidos (es decir, subsec. N) que se hibrida con una secuencia de sonda complementaria (P<sub>N</sub>) del ácido nucleico en fase sólida. El ácido nucleico en fase sólida hibridado con el ácido nucleico de muestra (S) puede extenderse (por ejemplo, ilustrado en la figura 2 como A) y se genera un "ácido nucleico en fase sólida extendido". En algunas implementaciones, el ácido nucleico en fase sólida extendido puede incluir una subsecuencia de nucleótidos complementaria a una secuencia de nucleótidos diana en la molécula de ácido nucleico de muestra. El ácido nucleico en fase sólida extendido puede amplificarse mediante la hibridación de un cebador (Pr'), que es complementario a la secuencia de cebador (Pr) del ácido nucleico en fase sólida extendido, en presencia de reactivos de amplificación/extensión (ilustrado en la figura 2 como B).

El cebador hibridado se extiende, generando de ese modo un producto de amplificación. El producto de amplificación contiene la secuencia de nucleótidos de cebador (Pr'), una secuencia de nucleótidos complementaria a la secuencia de identificación en el ácido nucleico en fase sólida, y una secuencia de nucleótidos idéntica a o sustancialmente idéntica a una secuencia de nucleótidos diana en la molécula de ácido nucleico de muestra. Este producto de amplificación lineal puede desprenderse a partir del soporte sólido para el análisis de secuencias (por ejemplo, ilustrado en la figura 2 como C).

#### Análisis de secuencias

Los productos de amplificación generados mediante los procedimientos descritos en el presente documento pueden someterse a análisis de secuencias. El término "análisis de secuencias" tal como se usa en el presente documento se refiere a determinar una secuencia de nucleótidos de un producto de amplificación. Puede determinarse toda la secuencia o una secuencia parcial de un producto de amplificación, y la secuencia de nucleótidos determinada se denomina en el presente documento "lectura". Una lectura puede obtenerse con o sin amplificación adicionales de los productos de amplificación resultantes de la extensión de un cebador que se hibrida con la secuencia de hibridación de cebador en un ácido nucleico en fase sólida. Por ejemplo, en algunas implementaciones, los productos de amplificación lineal pueden analizarse directamente sin amplificación adicional (por ejemplo, usando la metodología de secuenciación de molécula única (descrita con mayor detalle más adelante)). En determinadas

implementaciones, los productos de amplificación lineal pueden someterse a amplificación adicional y luego analizarse (por ejemplo, usando la metodología de secuenciación por ligación o pirosecuenciación (descrita con mayor detalle más adelante)). Las lecturas pueden someterse a diferentes tipos de análisis de secuencias.

En determinadas implementaciones de análisis de secuencias, las lecturas pueden usarse para construir una secuencia de nucleótidos más grande, que puede facilitarse mediante la identificación de secuencias solapantes en diferentes lecturas y mediante el uso de secuencias de identificación en las lecturas. Se conocen tales software y métodos de análisis de secuencias para construir secuencias más grandes a partir de lecturas (por ejemplo, Venter *et al.*, Science 291: 1304-1351 (2001)). En determinadas implementaciones de análisis de secuencias, las lecturas, los constructos de secuencia de nucleótidos parcial y los constructos de secuencia de nucleótidos completa específicos pueden compararse entre las secuencias de nucleótidos dentro de un ácido nucleico de muestra (es decir, comparación interna) o pueden compararse con una secuencia de referencia (es decir, comparación de referencia). Las comparaciones internas a veces se realizan en situaciones en las que un ácido nucleico de muestra se prepara a partir de múltiples muestras o a partir de una única fuente de muestra que contiene variaciones de secuencia. Las comparaciones de referencia a veces se realizan cuando se conoce una secuencia de nucleótidos de referencia y un objetivo es determinar si un ácido nucleico de muestra contiene una secuencia de nucleótidos que es sustancialmente similar o igual a, o diferente de, una secuencia de nucleótidos de referencia. El análisis de secuencias se facilita mediante el aparato y los componentes de análisis de secuencias.

En algunas implementaciones, una especie de ácido nucleico diana puede analizarse adicionalmente mediante secuenciación de nucleótidos. Puede utilizarse cualquier método de secuenciación adecuado. En algunas implementaciones, la secuenciación de nucleótidos puede ser mediante procedimientos y métodos de secuenciación de nucleótido único. Los métodos de secuenciación de nucleótido único implican poner en contacto un ácido nucleico de muestra y un soporte sólido en condiciones en las que una única molécula de ácido nucleico de muestra se hibrida con una única molécula de un soporte sólido. Tales condiciones pueden incluir proporcionar las moléculas de soporte sólido y una única molécula de ácido nucleico de muestra en un "microrreactor". Tales condiciones también pueden incluir proporcionar una mezcla en la que la molécula de ácido nucleico de muestra puede hibridarse con un ácido nucleico en fase sólida sobre el soporte sólido.

Los términos "aparato de análisis de secuencias" y "componente(s) de análisis de secuencias" usados en el presente documento se refieren a un aparato, y a uno o más componentes usados junto con tal aparato, que pueden usarse para determinar una secuencia de nucleótidos a partir de productos de amplificación resultantes de los procedimientos descritos en el presente documento (por ejemplo, productos de amplificación lineal y/o exponencial). Los ejemplos no limitativos de un aparato y componentes de análisis de secuencias actuales incluyen, sin limitación, sistemas que implican (i) secuenciación por ligación de sondas modificadas con colorante (por ejemplo, incluyendo ligación y escisión cíclicas), (ii) pirosecuenciación, y (iii) secuenciación de molécula única. Un producto de amplificación generado mediante un procedimiento descrito en el presente documento (por ejemplo, producto de amplificación lineal desprendido en la figura 2) puede considerarse como un "ácido nucleico de estudio" con el propósito de analizar una secuencia de nucleótidos mediante tal aparato y tales componentes de análisis de secuencias. Los ejemplos de plataformas de secuenciación incluyen, sin limitación, la plataforma 454 (Roche) (Margulies, M. *et al.* 2005 Nature 437, 376-380), el analizador genómico Illumina (o plataforma Solexa) o el sistema SOLID (Applied Biosystems) o la tecnología de secuenciación de ADN de molécula única Helicos True (Harris TD *et al.* 2008 Science, 320, 106-109), la tecnología en tiempo real de molécula única (SMRTM) de Pacific Biosciences, y secuenciación de nanoporos (Soni GV y Meller A. 2007 Clin Chem 53: 1996-2001). Tales plataformas permiten la secuenciación de muchas moléculas de ácido nucleico aisladas a partir de una muestra en órdenes de multiplexación altos de manera paralela (Dear Brief Funct Genomic Proteomic 2003; 1: 397-416). Cada una de estas plataformas permite la secuenciación de moléculas únicas clonalmente expandidas o no amplificadas de fragmentos de ácido nucleico.

La secuenciación por ligación es otro método de secuenciación de ácidos nucleicos. La secuenciación por ligación se basa en la sensibilidad de la ADN ligasa al apareamiento erróneo de bases. La ADN ligasa une entre sí los extremos del ADN que tienen un apareamiento correcto de bases. La combinación de la capacidad de la ADN ligasa para unir entre sí sólo los extremos del ADN que tienen un apareamiento correcto de bases, con agrupaciones mixtas de cebadores u oligonucleótidos marcados de manera fluorescente, permite la determinación de secuencias mediante detección por fluorescencia. Pueden obtenerse lecturas de secuencias más largas incluyendo cebadores que contienen enlaces escindibles que pueden escindirse después de la identificación del marcador. La escisión en el ligador elimina el marcador y regenera el fosfato 5' en el extremo del cebador ligado, preparando al cebador para otra tanda de ligación. En algunas implementaciones, los cebadores pueden ser uno o más marcadores fluorescentes (por ejemplo, un marcador fluorescente; 2, 3 ó 4 marcadores fluorescentes).

Un ejemplo de un sistema que puede usarse basado en secuenciación por ligación generalmente implica las siguientes etapas. Pueden prepararse poblaciones de perlas clonales en microrreactores en emulsión que contienen ácido nucleico de estudio ("molde"), componentes de reacción de amplificación, perlas y cebadores. Después de la amplificación, los moldes se desnaturalizan y se realiza enriquecimiento de perlas para separar las perlas con moldes extendidos a partir de las perlas no deseadas (por ejemplo, perlas sin moldes extendidos). El molde sobre las perlas seleccionadas experimenta una modificación en 3' para permitir la unión covalente al portaobjetos, y las



perlas modificadas pueden depositarse sobre un portaobjetos de vidrio. Las cámaras de deposición ofrecen la capacidad de segmentar un portaobjetos en una, cuatro u ocho cámaras durante el procedimiento de carga de perlas. Para el análisis de secuencias, los cebadores se hibridan con la secuencia adaptadora. Un conjunto de cuatro sondas marcadas con colorante de color compiten por la ligación al cebador de secuenciación. Se logra especificidad de ligación de sonda interrogando cada 4ª y 5ª base durante la serie de ligaciones. De cinco a siete tandas de ligación, detección y escisión registran el color en cada 5ª posición con el número de tandas determinado por el tipo de biblioteca usadas. Después de cada tanda de ligación, se establece un nuevo cebador complementario desplazado en una base en el sentido 5' para otra serie de ligaciones. Las tandas de restablecimiento y ligación del cebador (5-7 ciclos de ligación por tanda) se repiten secuencialmente cinco veces para generar 25-35 pares de bases de secuencia para una única etiqueta. Con la secuenciación por apareamiento de parejas, este procedimiento se repite para una segunda etiqueta. Un sistema de este tipo puede usarse para amplificar exponencialmente los productos de amplificación generados mediante un procedimiento descrito en el presente documento, por ejemplo, ligando un ácido nucleico heterólogo al primer producto de amplificación generado mediante un procedimiento descrito en el presente documento (por ejemplo, un producto de amplificación lineal de la figura 2) y realizando amplificación en emulsión usando el mismo soporte sólido o uno diferente originalmente usado para generar el primer producto de amplificación. Un sistema de este tipo también puede usarse para analizar los productos de amplificación directamente generados mediante un procedimiento descrito en el presente documento (por ejemplo, un producto de amplificación lineal de la figura 2) evitando un procedimiento de amplificación exponencial y clasificando directamente los soportes sólidos descritos en el presente documento sobre el portaobjetos de vidrio.

La pirosecuenciación es un método de secuenciación de ácidos nucleicos basado en secuenciación por síntesis, que se basa en la detección de un pirofosfato desprendido durante la incorporación de nucleótidos. Generalmente, la secuenciación por síntesis implica sintetizar, un nucleótido cada vez, una cadena de ADN complementaria a la cadena cuya secuencia está buscándose. Los ácidos nucleicos de estudio pueden inmovilizarse en un soporte sólido, hibridarse con un cebador de secuenciación, incubarse con ADN polimerasa, ATP sulfurilasa, luciferasa, apirasa, adenosina 5'-fosfosulfato y luciferina. Las disoluciones de nucleótidos se añaden y retiran secuencialmente. La correcta incorporación de un nucleótido desprende un pirofosfato, que interacciona con ATP sulfurilasa y produce ATP en presencia de adenosina 5'-fosfosulfato, desencadenando la reacción de luciferina, que produce una señal quimioluminiscente que permite la determinación de secuencia.

Un ejemplo de un sistema que puede usarse basado en pirosecuenciación generalmente implica las siguientes etapas: ligar un ácido nucleico adaptador a un ácido nucleico de estudio e hibridar el ácido nucleico de estudio con una perla; amplificar una secuencia de nucleótidos en el ácido nucleico de estudio en una emulsión; clasificar las perlas usando un soporte sólido de múltiples pocillos de picolitros; y secuenciar secuencias de nucleótidos amplificadas mediante metodología de pirosecuenciación (por ejemplo, Nakano *et al.* "Single-molecule PCR using water-in-oil emulsion"; Journal of Biotechnology 102 (2003) 117-124). Un sistema de este tipo puede usarse para amplificar exponencialmente los productos de amplificación generados mediante un procedimiento descrito en el presente documento, por ejemplo, ligando un ácido nucleico heterólogo al primer producto de amplificación generado mediante un procedimiento descrito en el presente documento (por ejemplo, un producto de amplificación lineal de la figura 2) y realizando amplificación en emulsión usando el mismo soporte sólido o uno diferente originalmente usado para generar el primer producto de amplificación. Un sistema de este tipo también puede usarse para analizar los productos de amplificación directamente generados mediante un procedimiento descrito en el presente documento (por ejemplo, un producto de amplificación lineal de la figura 2) evitando un procedimiento de amplificación exponencial y clasificando directamente los soportes sólidos descritos en el presente documento sobre el soporte de múltiples pocillos de picolitros.

Determinadas implementaciones de secuenciación de molécula única se basan en el principio de secuenciación por síntesis, y utilizan transferencia de energía por resonancia de fluorescencia de par único (FRET de par único) como mecanismo mediante el cual se emiten fotones como resultado de una incorporación de nucleótidos exitosa. Los fotones emitidos a menudo se detectan usando dispositivos de carga acoplada enfriados de alta sensibilidad o intensificados junto con microscopía de reflexión interna total (TIRM). Los fotones se emiten únicamente cuando la disolución de reacción introducida contiene el nucleótido correcto para su incorporación en la cadena de ácido nucleico creciente que se sintetiza como resultado del procedimiento de secuenciación. En la secuenciación de molécula única basada en FRET, se transfiere energía entre dos colorantes fluorescentes, a veces colorantes de polimetina-cianina Cy3 y Cy5, a través de interacciones dipolares de largo alcance. El donador se excita a su longitud de onda de excitación específica y la energía en estado excitado se transfiere, de manera no radiativa, al colorante aceptor, que a su vez se excita. El colorante aceptor eventualmente regresa al estado inicial mediante emisión radiativa de un fotón. Los dos colorantes usados en el procedimiento de transferencia de energía representan el "par único" en FRET de par único. Cy3 a menudo se usa como fluoróforo donador y a menudo se incorpora como primer nucleótido marcado. Cy5 a menudo se usa como fluoróforo aceptor y se usa como marcador nucleotídico para las sucesivas adiciones de nucleótido después de la incorporación de un primer nucleótido marcado con Cy3. Los fluoróforos generalmente están dentro de 10 nanómetros entre sí para que se produzca con éxito la transferencia de energía.

Un ejemplo de un sistema que puede usarse basado en secuenciación de molécula única generalmente implica hibridar un cebador con un ácido nucleico de estudio para generar un complejo; asociar el complejo con una fase



sólida; extender iterativamente el cebador mediante un nucleótido etiquetado con una molécula fluorescente; y capturar una imagen de las señales de transferencia de energía por resonancia de fluorescencia después de cada iteración (por ejemplo, patente estadounidense n.º 7.169.314; Braslavsky *et al.*, PNAS 100(7): 3960-3964 (2003)). Un sistema de este tipo puede usarse para secuenciar directamente los productos de amplificación generados mediante los procedimientos descritos en el presente documento (por ejemplo, producto de amplificación lineal desprendido en la figura 2). En algunas implementaciones, el producto de amplificación lineal desprendido puede hibridarse con un cebador que contiene secuencias complementarias a las secuencias de captura inmovilizadas presentes sobre un soporte sólido, una perla o un portaobjetos de vidrio, por ejemplo. La hibridación de los complejos cebador-producto de amplificación lineal desprendido con las secuencias de captura inmovilizadas inmoviliza los productos de amplificación lineal desprendidos a los soportes sólidos para la secuenciación por síntesis basada en FRET de par único. El cebador a menudo es fluorescente, de modo que puede generarse una imagen de referencia inicial de la superficie del portaobjetos con los ácidos nucleicos inmovilizados. La imagen de referencia inicial es útil para determinar ubicaciones en las que está produciéndose una verdadera incorporación de nucleótidos. Las señales de fluorescencia detectadas en ubicaciones de matriz no identificadas inicialmente en la imagen de referencia de "cebador sólo" se descartan como fluorescencia inespecífica. Después de la inmovilización de los complejos cebador-producto de amplificación lineal desprendido, los ácidos nucleicos unidos a menudo se secuencian en paralelo mediante las etapas iterativas de a) extensión con polimerasa en presencia de un nucleótido marcado de manera fluorescente, b) detección de la fluorescencia usando una microscopía apropiada, TIRM, por ejemplo, c) retirada del nucleótido fluorescente, y d) regreso a la etapa a) con un nucleótido marcado de manera fluorescente diferente.

El análisis de secuencias de nucleótidos puede incluir, en algunas implementaciones, fijar la información de secuencia de nucleótidos en forma tangible o electrónica. La información de secuencia de nucleótidos incluye sin limitación una o más secuencias de nucleótidos (por ejemplo, cadena(s) de bases nucleotídicas, secuencias completas, secuencias parciales), información relativa a procedimiento(s) usado(s) para obtener una secuencia de nucleótidos de muestra, información relativa a procedimiento(s) usado(s) para obtener un ácido nucleico de muestra a partir de una muestra, e información relativa a la(s) muestra(s) a partir de la(s) cual(es) se obtuvo el ácido nucleico de muestra (por ejemplo, información de paciente, información de población, ubicación de una fuente de muestra). La información de secuencia de nucleótidos puede fijarse en cualquier forma tangible o electrónica, incluyendo sin limitación un medio físico (por ejemplo, papel, y similares) o un medio legible por ordenador (por ejemplo, medio de transmisión o almacenamiento óptico y/o magnético, disco flexible, disco duro, memoria de acceso aleatorio, unidad central de procesamiento, señal de facsímil, señal de satélite, Internet, red informática mundial, y similares). La información de secuencia de nucleótidos puede fijarse en un registro oficial o no oficial (por ejemplo, registro de paciente, registro de seguro, cuaderno de laboratorio, registro gubernamental (por ejemplo, registro del Centro de Control de Enfermedades), y similares). La información de secuencia a veces se almacena y organiza en una base de datos. En determinadas implementaciones, la información de secuencia puede transferirse desde una ubicación hasta otra usando un medio físico o un medio electrónico (por ejemplo, transmisión desde un sitio en China hasta un sitio en Estado Unidos o un territorio del mismo).

#### Kits

Los kits a menudo comprenden uno o más recipientes que contienen uno o más componentes descritos en el presente documento. Un kit comprende uno o más componentes en cualquier número de recipientes, paquetes, tubos, viales, placas de múltiples pocillos, y similares independientes, o los componentes pueden combinarse en diversas combinaciones en tales recipientes. En un kit pueden incluirse, por ejemplo, uno o más de los siguientes componentes: (i) un soporte sólido que tiene un ácido nucleico en fase sólida; (ii) una colección de soportes sólidos que tiene un ácido nucleico en fase sólida; (iii) un ácido nucleico que puede asociarse con un soporte sólido para generar un soporte sólido que tiene un ácido nucleico en fase sólida; (iv) uno o más agentes que pueden usarse para asociar un ácido nucleico con un soporte sólido para generar un soporte sólido que tiene un ácido nucleico en fase sólida; (v) soporte(s) sólido(s) libre(s) de ácido nucleico; (vi) uno o más agentes de extensión; (vii) uno o más componentes; (viii) un aparato de emulsión y/o componente(s) de emulsión; (ix) un aparato de amplificación de ácidos nucleicos y/o componente(s) de amplificación de ácidos nucleicos; (x) un aparato de análisis de secuencias y/o componente(s) de análisis de secuencias; (xi) un sustrato que contiene pocillos u orificios de microrreactor; y (xii) software de análisis de secuencias de nucleótidos.

Un kit a veces se utiliza junto con un procedimiento, y puede incluir instrucciones para realizar uno o más procedimientos y/o una descripción de una o más composiciones. Un kit puede utilizarse para llevar a cabo un procedimiento (por ejemplo, usar un soporte sólido) descrito en el presente documento. Las instrucciones y/o descripciones pueden estar en forma tangible (por ejemplo, papel, y similares) o forma electrónica (por ejemplo, archivo legible por ordenador en un medio tangible (por ejemplo, disco compacto), y similares) y pueden incluirse en un prospecto del kit. Un kit también puede incluir una descripción escrita de una ubicación en Internet que proporciona tales instrucciones o descripciones.

#### Aplicaciones

Los procedimientos y soportes sólidos proporcionados en el presente documento son útiles para varios tipos de

análisis, cuyos ejemplos no limitativos se describen más adelante.

### 1. Identificación microbiana

Una cepa o cepas de microorganismos pueden identificarse usando los procedimientos y soportes sólidos descritos en el presente documento. El/los microorganismo(s) se selecciona(n) de una variedad de organismos incluyendo, pero sin limitarse a, bacterias, hongos, protozoos, ciliados y virus. Los microorganismos no están limitados a un género, una especie, una cepa, un subtipo o un serotipo particular. Los microorganismos pueden identificarse determinando variaciones de secuencia en una secuencia de microorganismo diana con respecto a una o más muestras o secuencias de referencia. La(s) secuencia(s) de referencia puede(n) obtenerse a partir de, por ejemplo, otros microorganismos del mismo o diferente género, especie, cepa o serotipo, o a partir de un organismo procariota o eucariota huésped.

La identificación y el tipado de patógenos (por ejemplo, bacterianos o virales) son críticos en la gestión clínica de enfermedades infecciosas. La identidad precisa de un microbio se usa no sólo para diferenciar un estado patológico de un estado sano, sino que también es fundamental para determinar la fuente de la infección y su propagación, y si y qué antibióticos u otras terapias antimicrobianas son más adecuados para el tratamiento. Además, puede monitorizarse el tratamiento. Los métodos tradicionales de tipado de patógenos han usado una variedad de características fenotípicas, incluyendo características de crecimiento, color, morfología de células o colonias, susceptibilidad a los antibióticos, tinción, olor, serotipado y reactividad con anticuerpos específicos para identificar microbios (por ejemplo, bacterias). Todos estos métodos requieren el cultivo del patógeno sospechoso, que se ve afectado por varios inconvenientes graves, incluyendo altos costes de materiales y mano de obra, peligro de exposición de los trabajadores, falsos positivos debidos a la incorrecta manipulación y falsos negativos debidos a los bajos números de células viables o debidos a los fastidiosos requisitos de cultivo de muchos patógenos. Además, los métodos de cultivo requieren un tiempo relativamente largo para lograr un diagnóstico, y debido a la naturaleza potencialmente mortal de tales infecciones, a menudo se inicia una terapia antimicrobiana antes de que puedan obtenerse los resultados. Algunos organismos no pueden mantenerse en cultivo o muestran tasas de crecimiento prohibitivamente lentas (por ejemplo, hasta 6-8 semanas para *Mycobacterium tuberculosis*).

En muchos casos, los patógenos están presentes en pequeñas cantidades y/o son muy similares a los organismos que constituyen la flora normal, y pueden ser indistinguibles de las cepas inocuas mediante los métodos citados anteriormente. En estos casos, la determinación de la presencia de la cepa patógena puede requerir la mayor resolución proporcionada por los métodos de tipado molecular proporcionados en el presente documento. Por ejemplo, en algunas implementaciones, la amplificación por PCR de una secuencia de ácido nucleico diana, seguido de escisión específica (por ejemplo, específica de base), seguido de espectrometría de masas con tiempo de vuelo de desorción/ionización láser asistida por matriz, seguido de cribado para las variaciones de secuencia, tal como se proporciona en el presente documento, permite una discriminación fiable de secuencias que difieren en tan sólo un nucleótido y combina el poder discriminatorio de la información de secuencia generada con la velocidad de MALDI-TOF MS.

Por tanto, en el presente documento se proporciona un método para detectar una secuencia de nucleótidos microbiana en una muestra, que comprende (a) proporcionar un ácido nucleico de muestra (por ejemplo, tomado de un sujeto); (b) preparar una mezcla del ácido nucleico de muestra con un soporte sólido descrito en el presente documento que tiene un ácido nucleico en fase sólida en condiciones en las que una única molécula del ácido nucleico de muestra se hibrida con una molécula de soporte sólido; (c) poner en contacto la mezcla con agentes de extensión en condiciones en las que se extiende el ácido nucleico en fase sólida hibridado con el ácido nucleico de muestra; (d) amplificar el ácido nucleico en fase sólida extendido; (e) analizar la secuencia del ácido nucleico amplificado de (d); y (f) basándose en una secuencia determinada en (e), identificar la presencia o ausencia de la secuencia de nucleótidos microbiana en el ácido nucleico de muestra. La parte (d) es opcional en determinadas implementaciones: el ácido nucleico en fase sólida extendido de (c) puede analizarse mediante secuenciación (e), sin amplificación (d), usando los métodos de secuenciación por síntesis descritos anteriormente, por ejemplo.

### 2. Detección de variaciones de secuencia

Las bases genómicas de una enfermedad y los marcadores de la misma pueden detectarse usando los procedimientos y soportes sólidos descritos en el presente documento. Los candidatos a variación de secuencia identificados mediante los métodos proporcionados en el presente documento incluyen secuencias que contienen variaciones de secuencia que son polimorfismos. Los polimorfismos incluyen tanto variaciones de secuencia somáticas que se producen de manera natural como aquellas que surgen a partir de una mutación. Los polimorfismos incluyen pero no se limitan a: microvariantes de secuencia en las que uno o más nucleótidos en una región localizada varían de individuo a individuo, inserciones y deleciones que pueden variar en cuanto a tamaño desde un nucleótidos hasta millones de bases, y repeticiones de nucleótidos o microsatélites que varían en los números de repeticiones. Las repeticiones de nucleótidos incluyen repeticiones homogéneas tales como repeticiones de dinucleótidos, trinucleótidos, tetranucleótidos o más grandes, en las que la misma secuencia se repite múltiples veces, y también repeticiones de heteronucleótidos en las que se encuentra que se repiten motivos de secuencia. Para un locus dado, el número de repeticiones de nucleótidos puede variar dependiendo del individuo.

Un sitio o marcador polimórfico es el locus en el que se produce divergencia. Un sitio de este tipo puede ser tan pequeño como un par de bases (un SNP). Los marcadores polimórficos incluyen, pero no se limitan a, polimorfismos de longitud de fragmentos de restricción (RFLP), repeticiones en tándem de número variable (VNTR), regiones hipervariables, minisatélites, repeticiones de dinucleótidos, repeticiones de trinucleótidos, repeticiones de tetranucleótidos y otros patrones de repetición, repeticiones de secuencia sencillas y elementos de inserción, tales como Alu. Las formas polimórficas también se manifiestan como diferentes alelos mendelianos para un gen. Los polimorfismos pueden observarse por diferencias en proteínas, modificaciones de proteínas, modificación de la expresión de ARN, metilación de ADN y ARN, factores reguladores que alteran la expresión génica y la replicación de ADN, y cualquier otra manifestación de alteraciones en ácidos nucleicos genómicos o ácidos nucleicos de orgánulos.

Además, numerosos genes tienen regiones polimórficas. Puesto que los individuos tienen una cualquiera de las varias variantes alélicas de una región polimórfica, los individuos pueden identificarse basándose en el tipo de variantes alélicas de las regiones polimórficas de los genes. Esto puede usarse, por ejemplo, con propósitos forenses. En otras situaciones, es crucial conocer la identidad de las variantes alélicas que tiene un individuo. Por ejemplo, diferencias alélicas en determinados genes, por ejemplo, genes de complejo mayor de histocompatibilidad (CMH), están implicadas en el rechazo de injerto o la enfermedad de injerto contra huésped en el trasplante de médula ósea. Por consiguiente, es muy deseable desarrollar métodos rápidos, sensibles y precisos para determinar la identidad de las variantes alélicas de las regiones polimórficas de los genes o las lesiones genéticas. Pueden usarse implementaciones de método o kit, tal como se proporciona en el presente documento, para someter a genotipado un sujeto determinando la identidad de una o más variantes alélicas de una o más regiones polimórficas en uno o más genes o cromosomas del sujeto. El genotipado de un sujeto usando un método tal como se proporciona en el presente documento puede usarse con propósitos de pruebas de identidad o forenses y las regiones polimórficas pueden estar presente en genes mitocondriales o pueden ser repeticiones en tándem cortas.

Los polimorfismos de un solo nucleótido (SNP) son generalmente sistemas bialélicos, es decir, hay dos alelos que puede tener un individuo para cualquier marcador particular. Esto significa que el contenido de información por marcador de SNP es relativamente bajo en comparación con marcadores de microsatélites, que pueden tener más de 10 alelos. Los SNP también tienden a ser muy específicos de población; un marcador que es polimórfico en una población a veces no es muy polimórfico en otra. Los SNP, encontrados aproximadamente cada kilobase (véase Wang *et al.* (1998) Science 280:1077-1082), ofrecen la posibilidad de generar mapas genéticos de densidad muy alta, que serán extremadamente útiles para desarrollar sistemas de haplotipado para genes o regiones de interés, y debido a la naturaleza de los SNP, de hecho pueden ser los polimorfismos asociados con los fenotipos de la enfermedad en estudio. La baja tasa de mutación de los SNP también los hace excelentes marcadores para estudiar rasgos genéticos complejos.

Gran parte del enfoque de la genómica ha estado en la identificación de los SNP, que son importantes por una variedad de motivos. Los SNP permiten pruebas indirectas (asociación de haplotipos) y pruebas directas (variantes funcionales). Los SNP son los marcadores genéticos más abundantes y estables. Las enfermedades comunes se explican mejor por las alteraciones genéticas comunes, y la variación natural en la población humana ayuda a entender la enfermedad, la terapia y las interacciones ambientales.

Por tanto, en el presente documento se proporciona un método para detectar una variación de secuencia en un ácido nucleico de muestra, que comprende (a) proporcionar un ácido nucleico a partir de una muestra; (b) preparar una mezcla del ácido nucleico de muestra con un soporte sólido descrito en el presente documento que tiene un ácido nucleico en fase sólida en condiciones en las que una única molécula de ácido nucleico de muestra se hibrida con una molécula de soporte sólido; (c) poner en contacto la mezcla con agentes de extensión en condiciones en las que se extiende el ácido nucleico en fase sólida hibridado con el ácido nucleico de muestra; (d) amplificar el ácido nucleico en fase sólida extendido; (e) analizar la secuencia del ácido nucleico amplificado de (d); y (f) basándose en una secuencia determinada en (e), identificar la presencia o ausencia de la secuencia de nucleótidos de marcador de enfermedad en el ácido nucleico de muestra. La muestra puede procesarse antes de la etapa (b), purificando el ácido nucleico en la muestra y/o fragmentando el ácido nucleico de muestra, por ejemplo. La parte (d) es opcional en determinadas implementaciones: el ácido nucleico en fase sólida extendido de (c) puede analizarse mediante secuenciación (e), sin amplificación (d), usando los métodos de secuenciación por síntesis descritos anteriormente, por ejemplo.

### 3. Detección de la presencia de secuencias de ácido nucleico microbianas indicativas de una infección

Los procedimientos y soportes sólidos descritos en el presente documento pueden usarse para determinar la presencia de secuencias de ácido nucleico microbianas indicativas de una infección mediante la identificación de variaciones de secuencia que están presentes en las secuencias de ácido nucleico virales o bacterianas con respecto a una o más secuencias de referencia. La(s) secuencia(s) de referencia puede(n) incluir, pero no se limita(n) a, secuencias obtenidas a partir de organismos no infecciosos relacionados, o secuencias procedentes de organismos huésped.

Los virus, las bacterias, los hongos y otros organismos infecciosos contienen secuencias de ácido nucleico distintas, incluyendo variantes de secuencia, que son diferentes de las secuencias contenidas en la célula huésped, y en algunos casos diferentes de las secuencias de especies, subespecies, serotipos, y similares relacionados, que pueden formar parte de la flora o fauna normal del huésped. Una secuencia de ADN diana puede formar parte de una secuencia genética exógena tal como el genoma de un microorganismo invasor, incluyendo, por ejemplo, bacterias y sus fagos, virus, hongos, protozoos, y similares. Los procedimientos proporcionados en el presente documento son particularmente aplicables para distinguir entre variantes o cepas diferentes de un microorganismo (por ejemplo, patógeno, menos patógeno, resistente frente a no resistente, y similares) con el fin de, por ejemplo, elegir una intervención terapéutica apropiada. Los ejemplos de virus causantes de enfermedad que infectan a seres humanos y animales y que pueden detectarse mediante un procedimiento dado a conocer incluyen pero no se limitan a *Retroviridae* (por ejemplo, virus de la inmunodeficiencia humana tales como VIH-1 (también denominado VLTH-III, VLA o VLTH-III/VLA; Ratner *et al.*, Nature, 313:227-284 (1985); Wain Hobson *et al.*, Cell, 40:9-17 (1985), VIH-2 (Guyader *et al.*, Nature, 328:662-669 (1987); publicación de patente europea n.º 0 269 520; Chakrabarti *et al.*, Nature, 328:543-547 (1987); solicitud de patente europea n.º 0 655 501), y otros aislados tales como VIH-LP (publicación internacional n.º WO 94/00562); *Picornaviridae* (por ejemplo, virus de la poliomieltitis, virus de la hepatitis A (Gust *et al.*, Intervirology, 20:1-7 (1983)); enterovirus, virus de Coxsackie humano, rinovirus, echovirus); *Calciviridae* (por ejemplo, cepas que provocan gastroenteritis); *Togaviridae* (por ejemplo, virus de la encefalitis equina, virus de la rubéola); *Flaviridae* (por ejemplo, virus del dengue, virus de la encefalitis, virus de la fiebre amarilla); *Coronaviridae* (por ejemplo, coronavirus); *Rhabdoviridae* (por ejemplo, virus de la estomatitis vesicular, virus de la rabia); *Filoviridae* (por ejemplo, virus del Ébola); *Paramyxoviridae* (por ejemplo, virus paragripales, virus de la parotiditis, virus del sarampión, virus sincitial respiratorio); *Orthomyxoviridae* (por ejemplo, virus de la gripe); *Bungaviridae* (por ejemplo, virus de Hantaan, virus b unga, flebovirus y virus de Nairo); *Arenaviridae* (virus de la fiebre hemorrágica); *Reoviridae* (por ejemplo, reovirus, orbivirus y rotavirus); *Birnaviridae*; *Hepadnaviridae* (virus de la hepatitis B); *Parvoviridae* (parvovirus); *Papovaviridae*; *Hepadnaviridae* (virus de la hepatitis B); *Parvoviridae* (la mayoría de los adenovirus); *Papovaviridae* (virus del papiloma, virus del polioma); *Adenoviridae* (la mayoría de los adenovirus); *Herpesviridae* (virus del herpes simple de tipo 1 (VHS-1) y VHS-2, virus de la varicela-zóster, citomegalovirus, virus del herpes; *Poxviridae* (virus de la viruela, virus de la vacuna, virus de la sífilis); *Iridoviridae* (por ejemplo, virus de la fiebre porcina africana); y virus no clasificados (por ejemplo, los agentes causales de encefalopatías espongiiformes, el agente de la hepatitis delta (se piensa que es un satélite defectuoso del virus de la hepatitis B), los agentes de la hepatitis distinta de A, distinta de B (clase 1 = transmitida internamente; clase 2 = transmitida parenteralmente, por ejemplo, hepatitis C); virus de Norwalk y relacionados, y astrovirus.

Los ejemplos de bacterias infecciosas incluyen pero no se limitan a *Helicobacter pylori*, *Borelia burgdorferi*, *Legionella pneumophila*, *Mycobacteria* sp. (por ejemplo, *M. tuberculosis*, *M. avium*, *M. intracellulare*, *M. kansaii*, *M. gordonae*), *Salmonella*, *Staphylococcus aureus*, *Neisseria gonorrhoeae*, *Neisseria meningitidis*, *Listeria monocytogenes*, *Streptococcus pyogenes* (*Streptococcus* del grupo A), *Streptococcus agalactiae* (*Streptococcus* del grupo B), *Streptococcus* sp. (grupo viridans), *Streptococcus faecalis*, *Streptococcus bovis*, *Streptococcus* sp. (especies anaerobias), *Streptococcus pneumoniae*, *Campylobacter* sp. patógenas, *Enterococcus* sp., *Haemophilus influenzae*, *Bacillus anthracis*, *Corynebacterium diphtheriae*, *Corynebacterium* sp., *Erysipelothrix rhusiopathiae*, *Clostridium perfringens*, *Clostridium tetani*, *Escherichia coli*, *Enterobacter aerogenes*, *Klebsiella pneumoniae*, *Pasturella multocida*, *Bacteroides* sp., *Fusobacterium nucleatum*, *Streptobacillus moniliformis*, *Treponema pallidum*, *Treponema pertenue*, *Leptospira* y *Actinomyces israelii*.

Los ejemplos de hongos infecciosos incluyen pero no se limitan a *Cryptococcus neoformans*, *Histoplasma capsulatum*, *Coccidioides immitis*, *Blastomyces dermatitidis*, *Chlamydia trachomatis*, *Candida albicans*. Otros organismos infecciosos incluyen protistas tales como *Plasmodium falciparum* y *Toxoplasma gondii*.

Por tanto, en el presente documento se proporciona un método para detectar una secuencia de nucleótidos microbiana infecciosa en una muestra, que comprende (a) proporcionar un ácido nucleico de muestra (por ejemplo, tomado de un sujeto); (b) preparar una mezcla del ácido nucleico de muestra con un soporte sólido descrito en el presente documento que tiene un ácido nucleico en fase sólida en condiciones en las que una única molécula del ácido nucleico de muestra se hibrida con una molécula de soporte sólido; (c) poner en contacto la mezcla con agentes de extensión en condiciones en las que se extiende el ácido nucleico en fase sólida hibridado con el ácido nucleico de muestra; (d) amplificar el ácido nucleico en fase sólida extendido; (e) analizar la secuencia del ácido nucleico amplificado de (d); y (f) basándose en una secuencia determinada mediante la parte (e), identificar la presencia o ausencia de la secuencia de nucleótidos microbiana infecciosa en el ácido nucleico de muestra. En determinadas implementaciones, la secuencia de nucleótidos microbiana puede compararse con una secuencia microbiana de referencia, y puede usarse, a veces junto con otra información, para diagnosticar una infección del sujeto. En algunas implementaciones, la muestra puede procesarse antes de la etapa (b), purificando el ácido nucleico en la muestra y/o fragmentando el ácido nucleico de muestra, por ejemplo. La parte (d) es opcional en determinadas implementaciones: el ácido nucleico en fase sólida extendido de (c) puede analizarse mediante secuenciación (e), sin amplificación (d), usando los métodos de secuenciación por síntesis descritos anteriormente.

4. Detección de la presencia de secuencias de ácido nucleico virales específica en una mezcla viral o población viral mixta

Los procedimientos y soportes sólidos descritos en el presente documento pueden usarse para determinar la presencia de secuencias de ácido nucleico virales específicas en una mezcla de secuencias virales o una población mixta de secuencias virales (por ejemplo, una mezcla "homogénea" que contiene sólo virus del mismo género, o poblaciones "heterogéneas" que podrían encontrarse en una muestra tomada de una fuente ambiental o un organismo inmunodeficiente).

Pruebas recientes sugieren que las "mezclas virales" (por ejemplo, poblaciones mixtas de virus del mismo o diferente género y especie, subespecie, cultivariedad, y similares, por ejemplo, hepatitis A, B y C), pueden conducir a una aparición aumentada de determinadas enfermedades, por ejemplo, cáncer. Esto es particularmente evidente con la infección por virus de la hepatitis B (VHB). Se observaron un aumento en la gravedad del transcurso de la enfermedad y un aumento en la reaparición de carcinoma hepatocelular en individuos coinfectados con dos o tres subgenotipos de VHB, particularmente los subgenotipos C2 y B2 (Yin *et al.*, "Role of hepatitis B virus genotype mixture, subgenotypes C2 and B2 on hepatocellular carcinoma: compared with chronic hepatitis B and asymptomatic carrier state in the same area" *Carcinogenesis*, 29(9): 1685-1691, 2008). En determinados casos también se conoce la variabilidad genética de los virus de ARN. Esta variabilidad genética, por ejemplo, tal como se observa en el virus de la inmunodeficiencia humana (VIH), ha conducido al descubrimiento de "cuasiespecies" o poblaciones virales mixtas, con un aumento en las formas farmacorresistentes del VIH descubiertas como resultado de la recombinación entre diferentes genotipos del VIH, debido a presiones selectivas antirretrovirales.

Las poblaciones virales mixtas se producen de manera natural. Se estima que los océanos del mundo contienen más de 22 toneladas métricas de partículas de fagos y virus (por ejemplo, más de  $10^{31}$  partículas, Rohwer y Edwards, "The phage proteomic tree: a genome based taxonomy for phage", *Journal of Bacteriology*, 184:4529-4535, 2002), algunas de las cuales se sabe que son patógenos humanos (Griffin *et al.*, *Pathogenic human viruses in coastal waters. Clinical Microbiol. Rev.* 16:129-143, 2003). Los océanos del mundo pueden tener un contenido viral promedio en el intervalo inferior de las cargas virales notificadas para el plasma humano de pacientes virémicos. Incluso suponiendo tasas de recombinación y mutación mínimas, diariamente se crea el equivalente de cientos o miles de "genomas humanos" completos válidos para nuevas secuencias genéticas. La identificación temprana de secuencias patógenas potencialmente nuevas usando los métodos descritos en el presente documento para detectar secuencias virales específicas en mezclas virales o poblaciones virales mixtas puede ser crucial para desarrollar tratamientos nuevos y eficaces. Además, las poblaciones virales están presentes en poblaciones humanas y muestras ambientales, y pueden evaluarse mediante las composiciones y los procedimientos descritos en el presente documento.

Los procedimientos proporcionados en el presente documento son particularmente aplicables para distinguir entre variantes o cepas, genotipos o subgenotipos de virus diferentes (por ejemplo, patógeno, menos patógeno, resistente frente a no resistente, y similares) con el fin de, por ejemplo, elegir una intervención terapéutica apropiada. Los ejemplos de virus causantes de enfermedad que infectan a seres humanos y animales y que pueden detectarse mediante un procedimiento dado a conocer incluyen pero no se limitan a *Retroviridae* (por ejemplo, virus de la inmunodeficiencia humana tales como VIH-1 (también denominado VLTH-III, VLA o VLTH-III/VLA; Ratner *et al.*, *Nature*, 313:227-284 (1985); Wain Hobson *et al.*, *Cell*, 40:9-17 (1985), VIH-2 (Guyader *et al.*, *Nature*, 328:662-669 (1987); publicación de patente europea n.º 0 269 520; Chakrabarti *et al.*, *Nature*, 328:543-547 (1987); solicitud de patente europea n.º 0 655 501), y otros aislados tales como VIH-LP (publicación internacional n.º WO 94/00562); *Picornaviridae* (por ejemplo, virus de la poliomielitis, virus de la hepatitis A (Gust *et al.*, *Intervirology*, 20:1-7 (1983)); enterovirus, virus de Coxsackie humano, rinovirus, echovirus); *Calciviridae* (por ejemplo, cepas que provocan gastroenteritis); *Togaviridae* (por ejemplo, virus de la encefalitis equina, virus de la rubéola); *Flaviridae* (por ejemplo, virus del dengue, virus de la encefalitis, virus de la fiebre amarilla); *Coronaviridae* (por ejemplo, coronavirus); *Rhabdoviridae* (por ejemplo, virus de la estomatitis vesicular, virus de la rabia); *Filoviridae* (por ejemplo, virus del Ébola); *Paramyxoviridae* (por ejemplo, virus paragripales, virus de la parotiditis, virus del sarampión, virus sincitial respiratorio); *Orthomyxoviridae* (por ejemplo, virus de la gripe); *Bungaviridae* (por ejemplo, virus de Hantaan, virus bunga, flebovirus y virus de Nairo); *Arenaviridae* (virus de la fiebre hemorrágica); *Reoviridae* (por ejemplo, reovirus, orbivirus y rotavirus); *Birnaviridae*; *Hepadnaviridae* (virus de la hepatitis B); *Parvoviridae* (parvovirus); *Papovaviridae*; *Hepadnaviridae* (virus de la hepatitis B); *Parvoviridae* (la mayoría de los adenovirus); *Papovaviridae* (virus del papiloma, virus del poliooma); *Adenoviridae* (la mayoría de los adenovirus); *Herpesviridae* (virus del herpes simple de tipo 1 (VHS-1) y VHS-2, virus de la varicela-zóster, citomegalovirus, virus del herpes; *Poxviridae* (virus de la viruela, virus de la vacuna, virus de la sífilis); *Iridoviridae* (por ejemplo, virus de la fiebre porcina africana); y virus no clasificados (por ejemplo, los agentes causales de encefalopatías espongiiformes, el agente de la hepatitis delta (se piensa que es un satélite defectuoso del virus de la hepatitis B), los agentes de la hepatitis distinta de A, distinta de B (clase 1 = transmitida internamente; clase 2 = transmitida parenteralmente, por ejemplo, hepatitis C); virus de Norwalk y relacionados, y astrovirus. En determinadas implementaciones, los procedimientos proporcionados en el presente documento pueden usarse para detectar secuencias de ácido nucleico de la hepatitis B en una mezcla de secuencias de *Hepadnaviridae*.

Por tanto, en el presente documento se proporciona un método para detectar secuencias de nucleótidos virales específicas en una muestra de mezcla viral o población viral mixta, que comprende (a) proporcionar un ácido nucleico de muestra (por ejemplo, tomado de un sujeto u océano); (b) preparar una mezcla del ácido nucleico de muestra con un soporte sólido descrito en el presente documento que tiene un ácido nucleico en fase sólida en

condiciones en las que una única molécula del ácido nucleico de muestra se hibrida con una molécula de soporte sólido; (c) poner en contacto la mezcla con agentes de extensión en condiciones en las que se extiende el ácido nucleico en fase sólida hibridado con el ácido nucleico de muestra; (d) amplificar el ácido nucleico en fase sólida extendido; (e) analizar la secuencia del ácido nucleico amplificado de (d); y (f) basándose en una secuencia determinada mediante la parte (e), identificar la presencia o ausencia de la secuencia de nucleótidos viral en el ácido nucleico de muestra. En determinadas implementaciones, la secuencia de nucleótidos viral puede compararse con una secuencia viral de referencia, y puede usarse, a veces junto con otra información, para detectar la presencia de una secuencia de ácido nucleico viral específica, o por ejemplo, para diagnosticar una infección de un sujeto. En algunas implementaciones, la muestra puede procesarse antes de la etapa (b), purificando el ácido nucleico en la muestra y/o fragmentando el ácido nucleico de muestra, por ejemplo. La parte (d) es opcional en determinadas implementaciones: el ácido nucleico en fase sólida extendido de (c) puede analizarse mediante secuenciación (e), sin amplificación (d), usando los métodos de secuenciación por síntesis descritos anteriormente.

#### 5. Obtención de perfiles de antibióticos

Los procedimientos y soportes sólidos descritos en el presente documento pueden utilizarse para identificar cambios de nucleótidos implicados en la farmacorresistencia, incluyendo la resistencia a los antibióticos. Se han identificado los loci genéticos implicados en la resistencia a isoniazida, rifampina, estreptomina, fluoroquinolonas y etionamida [Heym *et al.*, Lancet 344:293 (1994) y Morris *et al.*, J. Infect. Dis. 171:954 (1995)]. Habitualmente se usa una combinación de isoniazida (inh) y rifampina (rif) junto con pirazinamida y etambutol o estreptomina como primera línea de ataque contra casos confirmados de *M. tuberculosis* [Banerjee *et al.*, Science 263:227 (1994)]. La creciente incidencia de tales cepas resistentes necesita el desarrollo de ensayos rápidos para detectarlas y de ese modo reducir el gasto y los riesgos para la salud de buscar tratamientos ineficaces y posiblemente perjudiciales. La identificación de algunos de los loci genéticos implicados en la farmacorresistencia ha facilitado la adopción de tecnologías de detección de mutaciones para el rápido cribado de cambios de nucleótidos que dan como resultado farmacorresistencia. Además, la tecnología facilita la monitorización y el seguimiento del tratamiento o las estructuras de población microbiana.

Por tanto, en algunas implementaciones, la secuencia de nucleótidos diana identificada puede ser (i) un locus genético mutado como resultado de un organismo (por ejemplo, la secuencia estará presente si está presente un organismo farmacorresistente); (ii) un locus genético que no cambia como resultado de la farmacorresistencia (por ejemplo, una secuencia de un patógeno disminuirá a lo largo del tiempo si el fármaco se agota en el organismo); (iii) una secuencia de nucleótidos de una cepa particular no resistente al fármaco. Por consiguiente, en el presente documento se proporciona un método para determinar la presencia de farmacorresistencia, que comprende (a) proporcionar un ácido nucleico a partir de una muestra; (b) preparar una mezcla del ácido nucleico de muestra con un soporte sólido descrito en el presente documento que tiene un ácido nucleico en fase sólida en condiciones en las que una única molécula del ácido nucleico de muestra se hibrida con una molécula de soporte sólido; (c) poner en contacto la mezcla con agentes de extensión en condiciones en las que se extiende el ácido nucleico en fase sólida hibridado con el ácido nucleico de muestra; (d) amplificar el ácido nucleico en fase sólida extendido; (e) analizar la secuencia del ácido nucleico amplificado de (d); y (f) basándose en una secuencia determinada en (e), detectar la presencia o ausencia de una secuencia de ácido nucleico diana indicativa de farmacorresistencia. Puede identificarse la presencia de una secuencia indicativa de resistencia a un primer fármaco, y puede recetarse un fármaco alternativo. En algunas implementaciones, la muestra puede procesarse antes de la etapa (b), purificando el ácido nucleico en la muestra y/o fragmentando el ácido nucleico de muestra, por ejemplo. La parte (d) es opcional en determinadas implementaciones: el ácido nucleico en fase sólida extendido de (c) puede analizarse mediante secuenciación (e), sin amplificación (d), usando los métodos de secuenciación por síntesis descritos anteriormente.

#### 6. Identificación de marcadores de enfermedad

Los procedimientos y soportes sólidos descritos en el presente documento pueden utilizarse para identificar de manera rápida y precisa variaciones de secuencia que son marcadores genéticos de enfermedad, que pueden usarse para diagnosticar o determinar el pronóstico de una enfermedad. Las enfermedades caracterizadas por marcadores genéticos pueden incluir, pero no se limitan a, aterosclerosis, obesidad, diabetes, trastornos autoinmunitarios y cáncer. Las enfermedades en todos los organismos tienen un componente genético, y asea hereditario o resultante de la respuesta del cuerpo a tensiones ambientales, tales como virus y toxinas. El objetivo final de la investigación genómica en curso es usar esta información para desarrollar nuevos modos para identificar, tratar y posiblemente curar estas enfermedades. La primera etapa ha sido examinar el tejido patológico e identificar cambios genómicos a nivel de muestras individuales. La identificación de estos marcadores de "enfermedad" depende de la capacidad para detectar cambios en los marcadores genómicos con el fin de identificar genes o variantes de secuencia errantes. Los marcadores genómicos (todos los loci genéticos incluyendo polimorfismos de un solo nucleótido (SNP), microsatélites y otras regiones genómicas no codificantes, repeticiones en tándem, intrones y exones) pueden usarse para la identificación de todos los organismos, incluyendo seres humanos. Estos marcadores proporcionan un modo no sólo para identificar poblaciones, sino también para permitir la estratificación de las poblaciones según su respuesta a la enfermedad, el tratamiento farmacológico, la resistencia a los agentes ambientales, y otros factores.

Por tanto, en el presente documento se proporciona un método para detectar una secuencia de nucleótidos de marcador de enfermedad, que comprende (a) proporcionar un ácido nucleico a partir de una muestra; (b) preparar una mezcla del ácido nucleico de muestra con un soporte sólido descrito en el presente documento que tiene un ácido nucleico en fase sólida en condiciones en las que una única molécula del ácido nucleico de muestra se hibrida con una molécula de soporte sólido; (c) poner en contacto la mezcla con agentes de extensión en condiciones en las que se extiende el ácido nucleico en fase sólida hibridado con el ácido nucleico de muestra; (d) amplificar el ácido nucleico en fase sólida extendido; (e) analizar la secuencia del ácido nucleico amplificado de (d); y (f) basándose en una secuencia determinada mediante la parte (e), identificar la presencia o ausencia de la secuencia de nucleótidos de marcador de enfermedad en el ácido nucleico de muestra. En algunas implementaciones, la muestra puede procesarse antes de la etapa (b), purificando el ácido nucleico en la muestra y/o fragmentando el ácido nucleico de muestra, por ejemplo. La parte (d) es opcional en determinadas implementaciones: el ácido nucleico en fase sólida extendido de (c) puede analizarse mediante secuenciación (e), sin amplificación (d), usando los métodos de secuenciación por síntesis descritos anteriormente.

## 7. Haplotipado

Los procedimientos y soportes sólidos descritos en el presente documento pueden usarse para detectar haplotipos. En cualquier célula diploide, hay dos haplotipos en cualquier gen u otro segmento cromosómico que contienen al menos una varianza distintiva. En muchos sistemas genéticos bien estudiados, los haplotipos se correlacionan de manera más potente con los fenotipos que las variaciones de un solo nucleótido. Por tanto, la determinación de los haplotipos es valiosa para entender la base genética de una variedad de fenotipos incluyendo la predisposición o susceptibilidad a la enfermedad, la respuesta a las intervenciones terapéuticas, y otros fenotipos de interés en medicina, ganadería y agricultura.

Los procedimientos de haplotipado tal como se proporcionan en el presente documento permiten la selección de una porción de secuencia a partir de uno de los dos cromosomas homólogos de un individuo y el genotipado de los SNP unidos en esa porción de secuencia. La resolución directa de los haplotipos puede producir un mayor contenido de información, mejorando el diagnóstico de cualquier gen patológico vinculado o identificando vínculos asociados con esas enfermedades.

Por tanto, en el presente documento se proporciona un método para identificar un haplotipo que comprende dos o más nucleótidos, que comprende (a) proporcionar un ácido nucleico a partir de una muestra, en la que el ácido nucleico es de un cromosoma de un organismo diploide; (b) preparar una mezcla del ácido nucleico de muestra con un soporte sólido descrito en el presente documento que tiene un ácido nucleico en fase sólida en condiciones en las que una única molécula del ácido nucleico de muestra se hibrida con una molécula de soporte sólido; (c) poner en contacto la mezcla con agentes de extensión en condiciones en las que se extiende el ácido nucleico en fase sólida hibridado con el ácido nucleico de muestra; (d) amplificar el ácido nucleico en fase sólida extendido; (e) analizar la secuencia del ácido nucleico amplificado de (d); y (f) basándose en una secuencia determinada mediante la parte (e), determinar el haplotipo en el ácido nucleico de muestra. En algunas implementaciones, la muestra puede procesarse antes de la etapa (b), purificando el ácido nucleico en la muestra y/o fragmentando el ácido nucleico de muestra, por ejemplo. La parte (d) es opcional en determinadas implementaciones: el ácido nucleico en fase sólida extendido de (c) puede analizarse mediante secuenciación (e), sin amplificación (d), usando los métodos de secuenciación por síntesis descritos anteriormente.

## 8. Microsatélites

Los procedimientos y soportes sólidos descritos en el presente documento permiten una detección rápida e inequívoca de variaciones de secuencia que son microsatélites. Los microsatélites (a veces denominados repeticiones en tándem de número variable o VNTR) son unidades nucleotídicas repetidas en tándem cortas de una a siete o más bases, siendo las más prominentes entre ellas las repeticiones de dinucleótidos, trinucleótidos y tetranucleótidos. Los microsatélites están presentes cada 100.000 pb en el ADN genómico (J. L. Weber y P. E. Can, *Am. J. Hum. Genet.* 44, 388 (1989); J. Weissenbach *et al.*, *Nature* 359, 794 (1992)). Las repeticiones de dinucleótidos CA, por ejemplo, constituyen aproximadamente el 0,5 % del genoma extramitocondrial humano; las repeticiones CT y AG juntas constituyen aproximadamente el 0,2 %. Las repeticiones CG son raras, lo más probablemente debido a la función reguladora de las islas CpG. Los microsatélites son muy polimórficos con respecto a la longitud y se distribuyen ampliamente a lo largo de todo el genoma con una abundancia principal en las secuencias no codificantes, y se desconoce su función dentro del genoma.

Los microsatélites son importantes en aplicaciones forenses, ya que una población mantendrá una variedad de microsatélites característicos de esa población y distintos de otras poblaciones, que no se cruzan. Muchos cambios dentro de los microsatélites pueden ser silenciosos, pero algunos pueden conducir a alteraciones significativas en los niveles de expresión o productos génicos. Por ejemplo, las repeticiones de trinucleótidos halladas en las regiones codificantes de los genes se ven afectadas en algunos tumores (C. T. Caskey *et al.*, *Science* 256, 784 (1992)), y la alteración de los microsatélites puede dar como resultado una inestabilidad genética que da como resultado una predisposición al cáncer (P. J. McKinnen, *Hum. Genet.* 1 75, 197 (1987); J. German *et al.*, *Clin. Genet.* 35, 57 (1989)).

Por tanto, en el presente documento se proporciona un método para detectar una secuencia de microsatélite, que comprende (a) proporcionar un ácido nucleico a partir de una muestra; (b) preparar una mezcla del ácido nucleico de muestra con un soporte sólido descrito en el presente documento que tiene un ácido nucleico en fase sólida en condiciones en las que una única molécula del ácido nucleico de muestra se hibrida con una molécula de soporte sólido; (c) poner en contacto la mezcla con agentes de extensión en condiciones en las que se extiende el ácido nucleico en fase sólida hibridado con el ácido nucleico de muestra; (d) amplificar el ácido nucleico en fase sólida extendido; (e) analizar la secuencia del ácido nucleico amplificado de (e); y (f) basándose en una secuencia determinada mediante la parte (d), determinar si la secuencia de microsatélite está presente en el ácido nucleico de muestra. Una secuencia de microsatélite puede ser una secuencia de microsatélite completa o una porción de una secuencia de microsatélite completa. En algunas implementaciones, la muestra puede procesarse antes de la etapa (b), purificando el ácido nucleico en la muestra y/o fragmentando el ácido nucleico de muestra, por ejemplo. La parte (d) es opcional en determinadas implementaciones: el ácido nucleico en fase sólida extendido de (c) puede analizarse mediante secuenciación (e), sin amplificación (d), usando los métodos de secuenciación por síntesis descritos anteriormente.

#### 9. Repeticiones en tándem cortas

Los procedimientos y soportes sólidos descritos en el presente documento pueden usarse para identificar regiones de repetición en tándem corta (STR) en algunas secuencias diana del genoma humano con respecto a, por ejemplo, secuencias de referencia en el genoma humano que no contiene regiones de STR. Las regiones de STR son regiones polimórficas que no están relacionadas con ninguna enfermedad o afección. Muchos loci en el genoma humano contienen una región de repetición en tándem corta (STR) polimórfica. Los loci de STR contienen elementos de secuencia cortos y repetitivos de 3 a 7 pares de bases de longitud. Se estima que hay 200.000 STR triméricas y tetraméricas esperadas, que están presentes tan frecuentemente como una vez cada 15 kb en el genoma humano (véanse, por ejemplo, la solicitud PCT internacional n.º WO 9213969 A1, Edwards *et al.*, Nucl. Acids Res. 19:4791 (1991); Beckmann *et al.* (1992) Genomics 12:627-631). Prácticamente la mitad de estos loci de STR son polimórficos, proporcionando una rica fuente de marcadores genéticos. La variación en el número de unidades de repetición en un locus particular es responsable de las variaciones de secuencia observadas reminiscentes de los loci de repeticiones en tándem de nucleótidos variables (VNTR) (Nakamura *et al.* (1987) Science 235:1616-1622); y los loci de minisatélites (Jeffreys *et al.* (1985) Nature 314:67-73), que contienen unidades de repetición más largas, y loci de repeticiones de dinucleótidos o microsatélites (Luty *et al.* (1991) Nucleic Acids Res. 19:4308; Litt *et al.* (1990) Nucleic Acids Res. 18:4301; Litt *et al.* (1990) Nucleic Acids Res. 18:5921; Luty *et al.* (1990) Am. J. Hum. Genet. 46:776-783; Tautz (1989) Nucl. Acids Res. 17:6463-6471; Weber *et al.* (1989) Am. J. Hum. Genet. 44:388-396; Beckmann *et al.* (1992) Genomics 12:627-631).

Los ejemplos de loci de STR incluyen, pero no se limitan a, repeticiones de pentanucleótidos en el locus de CD4 humana (Edwards *et al.*, Nucl. Acids Res. 19:4791 (1991)); repeticiones de tetranucleótidos en el gen de aromatasas citocromo P-450 humano (CYP19; Polymeropoulos *et al.*, Nucl. Acids Res. 19:195 (1991)); repeticiones de tetranucleótidos en el gen de la subunidad A del factor de coagulación XIII humano (F13A1; Polymeropoulos *et al.*, Nucl. Acids Res. 19:4306 (1991)); repeticiones de tetranucleótidos en el locus de F13B (Nishimura *et al.*, Nucl. Acids Res. 20:1167 (1992)); repeticiones de tetranucleótidos en protooncogén c-les/fps humano (FES; Polymeropoulos *et al.*, Nucl. Acids Res. 19:4018 (1991)); repeticiones de tetranucleótidos en el gen de LFL (Zuliani *et al.*, Nucl. Acids Res. 18:4958 (1990)); variaciones de secuencia de repeticiones de trinucleótidos en el gen de fosfolipasa pancreática A-2 humano (PLA2; Polymeropoulos *et al.*, Nucl. Acids Res. 18:7468 (1990)); variaciones de secuencia de repeticiones de tetranucleótidos en el gen de VWF (Ploos *et al.*, Nucl. Acids Res. 18:4957 (1990)); y repeticiones de tetranucleótidos en el locus de peroxidasa tiroidea humana (hTPO) (Anker *et al.*, Hum. Mol. Genet. 1:137 (1992)).

Por tanto, en el presente documento se proporciona un método para detectar una secuencia de repetición en tándem corta, que comprende (a) proporcionar un ácido nucleico a partir de una muestra; (b) preparar una mezcla del ácido nucleico de muestra con un soporte sólido descrito en el presente documento que tiene un ácido nucleico en fase sólida en condiciones en las que una única molécula del ácido nucleico de muestra se hibrida con una molécula de soporte sólido; (c) poner en contacto la mezcla con agentes de extensión en condiciones en las que se extiende el ácido nucleico en fase sólida hibridado con el ácido nucleico de muestra; (d) amplificar el ácido nucleico en fase sólida extendido; (e) analizar la secuencia del ácido nucleico amplificado de (d); y (f) basándose en una secuencia determinada en (e), determinar si la secuencia de repetición en tándem corta está presente en el ácido nucleico de muestra. Una repetición en tándem corta secuencia puede ser una secuencia de STR completa o una porción de una secuencia de STR completa. En algunas implementaciones, la muestra puede procesarse antes de la etapa (b), purificando el ácido nucleico en la muestra y/o fragmentando el ácido nucleico de muestra, por ejemplo. La parte (d) es opcional en determinadas implementaciones: el ácido nucleico en fase sólida extendido de (c) puede analizarse mediante secuenciación (e), sin amplificación (d), usando los métodos de secuenciación por síntesis descritos anteriormente.

#### 10. Identificación de organismos

Los procedimientos y soportes sólidos descritos en el presente documento pueden utilizarse para identificar loci de



STR polimórficos y otras regiones polimórficas útiles para discriminar un organismo de otro. Determinados loci de STR polimórficos y otras regiones polimórficas de genes son variaciones de secuencia que son marcadores útiles para identificación de seres humanos, pruebas de paternidad y maternidad, mapeo genético, disputas de inmigración y herencia, pruebas de cigosidad en gemelos, pruebas de endogamia en seres humanos, control de calidad de células cultivadas humanas, identificación de restos humanos, y pruebas de muestras de semen, manchas de sangre, microbios y otros materiales en medicina forense. Tales loci también son marcadores útiles en la cría de animales comerciales y el análisis de pedigrí y en el cultivo de plantas comerciales. Pueden identificarse rasgos de importancia económica en cultivos de plantas y animales a través del análisis de vínculos usando marcadores de ADN polimórficos. En el presente documento se proporcionan métodos eficientes y preciosos para determinar la

identidad de tales loci.

Por tanto, en el presente documento se proporciona un método para detectar una secuencia de nucleótidos diana de un organismo, que comprende (a) proporcionar un ácido nucleico a partir de una muestra; (b) preparar una mezcla del ácido nucleico de muestra con un soporte sólido descrito en el presente documento que tiene un ácido nucleico en fase sólida en condiciones en las que una única molécula del ácido nucleico de muestra se hibrida con una molécula de soporte sólido; (c) poner en contacto la mezcla con agentes de extensión en condiciones en las que se extiende el ácido nucleico en fase sólida hibridado con el ácido nucleico de muestra; (d) amplificar el ácido nucleico en fase sólida extendido; (e) analizar la secuencia del ácido nucleico amplificado de (d); y (f) basándose en una secuencia determinada en (e), determinar si está presente la secuencia de nucleótidos diana. Si está detectándose la presencia de un primer organismo que reside en un segundo organismo, generalmente se selecciona una secuencia de nucleótidos diana presente en un ácido nucleico del primer organismo que no está presente en el ácido nucleico del segundo organismo (por ejemplo, una secuencia de nucleótidos en un ácido nucleico patógeno que no está presente en un ácido nucleico humano; una secuencia de nucleótidos en un ácido nucleico fetal que no está presente en un ácido nucleico materno). En algunas implementaciones, la muestra puede procesarse antes de la etapa (b), purificando el ácido nucleico en la muestra y/o fragmentando el ácido nucleico de muestra, por ejemplo. La parte (d) es opcional en determinadas implementaciones: el ácido nucleico en fase sólida extendido de (c) puede analizarse mediante secuenciación (e), sin amplificación (d), usando los métodos de secuenciación por síntesis descritos anteriormente.

#### 11. Detección de variaciones alélicas

Los procedimientos y soportes sólidos descritos en el presente documento permiten una detección de alto rendimiento, rápida y precisa de variantes alélicas. Las poblaciones humanas son heterogéneas en cuanto a susceptibilidad a enfermedades o respuestas a intervenciones terapéuticas particulares. Cada vez más pruebas sugieren que una variación alélica en la expresión génica es un fenómeno generalizado, y puede contribuir a una variación fenotípica entre individuos. A medida que se secuencian más genomas, la identificación y la caracterización de las causas de una variación heredable dentro de una especie será cada vez más importante. Puede observarse una variación alélica entre poblaciones étnicas o regionales y dentro de las poblaciones étnicas y regionales. En algunos casos, puede encontrarse una variación intrapoblación dentro de poblaciones relativamente pequeñas. Una variación alélica heredable en la expresión génica puede contribuir a una enfermedad esporádica y familiar, pero es relativamente inexplorada. El entendimiento de la variación alélica puede ayudar a proporcionar información sobre varios fenómenos de heterogeneidad genética, incluyendo pero sin limitarse a sellado genético, susceptibilidad a una enfermedad y respuesta terapéutica.

Algunos estudios de variación alélica implican no sólo la detección de una secuencia específica en un fondo complejo, sino también la discriminación entre secuencias con pocas diferencias de nucleótidos, o una sola diferencia de nucleótido. Los estudios de variaciones alélicas pueden realizarse en ADN o ARN, por tanto pueden realizarse correlaciones entre las variantes alélicas, los SNP y los niveles de expresión. Un método para detectar el grado de variación en la expresión alélica en cualquier locus específico es genotipar cuantitativamente el ARNm de individuos heterocigóticos para un polimorfismo de un solo nucleótido (SNP) exónico en el gen de interés. Si no hay ninguna alélica variación en la expresión génica, entonces los dos alelos del SNP deben expresarse al mismo nivel, pero cuando hay una expresión alélica diferencial, un alelo se encontrará a un nivel mayor que el otro.

Por tanto, en el presente documento se proporciona un método para detectar una variación de secuencia en una secuencia de nucleótidos diana, que comprende (a) proporcionar un ácido nucleico a partir de una muestra; (b) preparar una mezcla del ácido nucleico de muestra con un soporte sólido descrito en el presente documento que tiene un ácido nucleico en fase sólida en condiciones en las que una única molécula del ácido nucleico de muestra se hibrida con una molécula de soporte sólido; (c) poner en contacto la mezcla con agentes de extensión en condiciones en las que se extiende el ácido nucleico en fase sólida hibridado con el ácido nucleico de muestra; (d) amplificar el ácido nucleico en fase sólida extendido; (e) analizar la secuencia del ácido nucleico amplificado de (d); y (f) basándose en una secuencia determinada en (e), determinar si está presente una variación de secuencia en la secuencia de nucleótidos diana. En algunas implementaciones, la muestra puede procesarse antes de la etapa (b), purificando el ácido nucleico en la muestra y/o fragmentando el ácido nucleico de muestra, por ejemplo. La parte (d) es opcional en determinadas implementaciones: el ácido nucleico en fase sólida extendido de (c) puede analizarse mediante secuenciación (e), sin amplificación (d), usando los métodos de secuenciación por síntesis descritos anteriormente.

## 12. Determinación de la frecuencia alélica

Los procedimientos y soportes sólidos descritos en el presente documento son útiles para identificar uno o más marcadores genéticos cuya frecuencia cambia dentro de la población en función de la edad, el grupo étnico, el sexo o algunos otros criterios. Por ejemplo, en la técnica se conoce la distribución dependiente de la edad de los genotipos de ApoE (véase Schechter *et al.* (1994) *Nature Genetics* 6:29-32). Las frecuencias de las variaciones de secuencia que se sabe que están asociadas al mismo nivel con una enfermedad también pueden usarse para detectar o monitorizar la progresión de un estado patológico. Por ejemplo, el polimorfismo N291S (N291S) del gen de la lipoproteína lipasa, que da como resultado una sustitución de una serina por una asparagina en el codón de aminoácido 291, conduce a niveles reducidos de colesterol de lipoproteínas de alta densidad (HDL-C) que está asociado con un mayor riesgo de hombres con arterioesclerosis y en particular infarto de miocardio (véase Reymer *et al.* (1995) *Nature Genetics* 10:28-34). Además, la determinación de cambios en la frecuencia alélica puede permitir la identificación de variaciones de secuencia previamente desconocidas y, en última instancia, un gen o una ruta implicado en el inicio y la progresión de la enfermedad.

Por tanto, en el presente documento se proporciona un método para determinar la frecuencia de una secuencia de nucleótidos diana en una población de individuos, que comprende (a) proporcionar un ácido nucleico de muestra (por ejemplo, tomado de un sujeto); (b) preparar una mezcla del ácido nucleico de muestra con un soporte sólido descrito en el presente documento que tiene un ácido nucleico en fase sólida en condiciones en las que una única molécula del ácido nucleico de muestra se hibrida con una molécula de soporte sólido; (c) poner en contacto la mezcla con agentes de extensión en condiciones en las que se extiende el ácido nucleico en fase sólida hibridado con el ácido nucleico de muestra; (d) amplificar el ácido nucleico extendido; (e) analizar la secuencia del ácido nucleico amplificado; (f) identificar la presencia o ausencia de la secuencia de nucleótidos diana según en (e); y (g) repetir las etapas (a) a (f) para otros individuos de la población y determinar la frecuencia de la secuencia de nucleótidos diana en la población. En algunas implementaciones, la muestra puede procesarse antes de la etapa (b), purificando el ácido nucleico en la muestra y/o fragmentando el ácido nucleico de muestra, por ejemplo. En algunas implementaciones, los nucleótidos metilados en el ácido nucleico de muestra pueden convertirse en otro nucleótido usando métodos conocidos en la técnica, tales como conversión con bisulfito de citosina metilada en uracilo, por ejemplo. La parte (d) es opcional en determinadas implementaciones: el ácido nucleico en fase sólida extendido de (c) puede analizarse mediante secuenciación (e), sin amplificación (d), usando los métodos de secuenciación por síntesis descritos anteriormente.

## 13. Epigenética

Los procedimientos y soportes sólidos descritos en el presente documento pueden usarse para estudiar variaciones en una proteína o un ácido nucleico diana con respecto a una proteína o un ácido nucleico de referencia que no se basan en la secuencia, por ejemplo, la identidad de bases o aminoácidos que son las unidades monoméricas que se producen de manera natural de la proteína o del ácido nucleico. Por ejemplo, los reactivos de escisión específicos empleados en los métodos proporcionados en el presente documento pueden reconocer diferencias en características independientes de la secuencia tales como los patrones de metilación, la presencia de bases o aminoácidos modificados, o diferencias en la estructura de mayor orden entre la molécula diana y la molécula de referencia, para generar fragmentos que se escinden en sitios independientes de la secuencia. La epigenética es el estudio de la herencia de información basada en diferencias en la expresión génica en lugar de diferencias en la secuencia génica. Los cambios epigenéticos se refieren a cambios heredables de manera mitótica y/o meiótica en la función génica o cambios en la estructura de ácido nucleico de mayor orden que no pueden explicarse por los cambios en la secuencia de ácido nucleico. Los ejemplos de características que están sujetas a variación o cambio epigenético incluyen, pero no se limitan a, patrones de metilación del ADN en animales, modificación de histonas y los complejos proteicos de grupos Polycomb-trithorax (Pc-G/trx) (véase, por ejemplo, Bird, A., *Genes Dev.*, 16:6-21 (2002)).

Los cambios epigenéticos habitualmente, aunque no necesariamente, conducen a cambios en la expresión génica que habitualmente, aunque no necesariamente, son heredables. Por ejemplo, tal como se comenta adicionalmente a continuación, los cambios en los patrones de metilación a veces pueden ser un acontecimiento temprano en el desarrollo y la progresión del cáncer y otras enfermedades. En muchos cánceres, determinados genes se inactivan o activan de manera inapropiada debido a una metilación aberrante. Puede heredarse la capacidad de los patrones de metilación para reprimir o activar la transcripción. Los complejos proteicos Pc-G/trx, al igual que la metilación, pueden reprimir la transcripción de manera heredable. El conjunto multiproteico Pc-G/trx está dirigido a regiones específicas del genoma en las que congela eficazmente el estado de expresión génica embrionaria de un gen, ya sea que el gen esté activo o inactivo, y propaga de manera estable ese estado a través del desarrollo. La capacidad del grupo de proteínas Pc-G/trx para dirigirse y unirse a un genoma afecta sólo al nivel de expresión de los genes contenidos en el genoma, y no a las propiedades de los productos génicos. Los métodos proporcionados en el presente documento pueden usarse con reactivos de escisión específicos que identifican variaciones en una secuencia diana con respecto a una secuencia de referencia que se basan en cambios independientes de la secuencia, tales como cambios epigenéticos.

Por tanto, en el presente documento se proporciona un método para el análisis epigenético de una secuencia de nucleótidos diana, que comprende (a) proporcionar un ácido nucleico a partir de una muestra en la que los nucleótidos metilados o los nucleótidos no metilados se han convertido en otro resto de nucleótido; (b) preparar una mezcla del ácido nucleico de muestra con un soporte sólido descrito en el presente documento que tiene un ácido nucleico en fase sólida en condiciones en las que una única molécula del ácido nucleico de muestra se hibrida con una molécula de soporte sólido; (c) poner en contacto la mezcla con agentes de extensión en condiciones en las que el ácido nucleico en fase sólida se hibrida con el ácido nucleico de muestra; (d) preparar una emulsión; (e) amplificar el ácido nucleico extendido; (f) analizar la secuencia del ácido nucleico amplificado; y (g) basándose en una secuencia determinada en (f), comparar el patrón de metilación del ácido nucleico diana con el patrón de metilación de un ácido nucleico de referencia. En algunas implementaciones, la muestra puede procesarse antes de la etapa (b), purificando el ácido nucleico en la muestra y/o fragmentando el ácido nucleico de muestra, por ejemplo. La parte (d) es opcional en determinadas implementaciones: el ácido nucleico en fase sólida extendido de (c) puede analizarse mediante secuenciación (f), sin amplificación (e), usando los métodos de secuenciación por síntesis descritos anteriormente.

El término "otro resto de nucleótido" tal como se usa en el presente documento se refiere a un resto de nucleótido distinto del nucleótido que estaba metilado o no metilado. El "otro resto de nucleótido" puede producirse de manera natural o no producirse de manera natural. Los nucleótidos metilados en el ácido nucleico de muestra pueden convertirse en otro resto de nucleótido usando métodos conocidos en la técnica, tales como conversión con bisulfito de citosina metilada en uracilo, por ejemplo.

#### 14. *Patrones de metilación*

Los procedimientos y soportes sólidos descritos en el presente documento pueden usarse para detectar variaciones de secuencia que son cambios epigenéticos en la secuencia diana, tales como un cambio en los patrones de metilación en la secuencia diana. El análisis de la metilación celular es una disciplina de investigación emergente. La adición covalente de grupos metilo a citosina está presente principalmente en dinucleótidos (microsatélites) CpG. Aunque la función de las islas CpG no ubicadas en regiones promotoras sigue inexplorada, las islas CpG en las regiones promotoras son de especial interés porque su estado de metilación regula la transcripción y la expresión del gen asociado. La metilación de regiones promotoras conduce al silenciamiento de la expresión génica. Este silenciamiento es permanente y continúa a través del procedimiento de mitosis. Debido a su importante papel en la expresión génica, la metilación del ADN tiene un impacto sobre los procesos del desarrollo, el sellado y la inactivación de cromosomas X, así como sobre la génesis tumoral, el envejecimiento y también la supresión de ADN parasitario. Se cree que la metilación está implicada en la carcinogénesis de muchos tumores generalizados, tales como cáncer de pulmón, mama y colon, y en leucemia. También existe una relación entre la metilación y las disfunciones proteicas (síndrome de Q-T largo) o enfermedades metabólicas (diabetes neonatal transitoria, diabetes de tipo 2).

El tratamiento del ADN genómico con bisulfito puede utilizarse para analizar posiciones de residuos de citosina metilada dentro del ADN. El tratamiento de ácidos nucleicos con bisulfito desamina los residuos de citosina para dar residuos de uracilo, mientras que la citosina metilada permanece sin modificar. Por tanto, comparando la secuencia de un ácido nucleico diana que no está tratado con bisulfito con la secuencia del ácido nucleico que está tratado con bisulfito en los métodos proporcionados en el presente documento, puede deducirse el grado de metilación en un ácido nucleico, así como las posiciones en las que la citosina está metilada.

El análisis de la metilación a través de una reacción con endonucleasas de restricción se vuelve posible mediante el uso de enzimas de restricción, que tienen sitios de reconocimiento específicos de metilación, tales como HpaII y MspI. El principio básico es que determinadas enzimas son bloqueadas por la citosina metilada en la secuencia de reconocimiento. Una vez que se ha logrado esta diferenciación, puede realizarse el posterior análisis de los fragmentos resultantes usando los métodos tal como se proporcionan en el presente documento.

Estos métodos pueden usarse juntos en análisis combinado de restricción con bisulfito (COBRA). El tratamiento con bisulfito provoca una pérdida en el sitio de reconocimiento de BstUI en un producto de PCR amplificado, lo que provoca la aparición de un nuevo fragmento detectable durante el análisis en comparación con una muestra sin tratar. Los métodos basados en escisión proporcionados en el presente documento pueden usarse junto con la escisión específica de los sitios de metilación para proporcionar información rápida y fiable sobre los patrones de metilación en una secuencia de ácido nucleico diana.

Por tanto, en el presente documento se proporciona un método para analizar un patrón de metilación de una secuencia de nucleótidos diana, que comprende (a) proporcionar un ácido nucleico a partir de una muestra en la que los nucleótidos metilados o los nucleótidos no metilados se han convertido en otro resto de nucleótido; (b) preparar una mezcla del ácido nucleico de muestra con un soporte sólido descrito en el presente documento que tiene un ácido nucleico en fase sólida en condiciones en las que una única molécula del ácido nucleico de muestra se hibrida con una molécula de soporte sólido; (c) poner en contacto la mezcla con agentes de extensión en condiciones en las que se extiende el ácido nucleico en fase sólida hibridado con el ácido nucleico de muestra; (d) amplificar el ácido nucleico extendido; (e) analizar la secuencia del ácido nucleico amplificado; y (f) determinar el patrón de metilación

basándose en la secuencia en (e). En algunas implementaciones, la muestra puede procesarse antes de la etapa (b), purificando el ácido nucleico en la muestra y/o fragmentando el ácido nucleico de muestra, por ejemplo. La parte (d) es opcional en determinadas implementaciones: el ácido nucleico en fase sólida extendido de (c) puede analizarse mediante secuenciación (e), sin amplificación (d), usando los métodos de secuenciación por síntesis descritos anteriormente.

#### 15. *Resecuenciación*

Los procedimientos y soportes sólidos descritos en el presente documento son útiles para análisis de resecuenciación rápidos. La cantidad drásticamente creciente de información de secuencia genómica disponible a partir de diversos organismos aumenta la necesidad de tecnologías que permitan un análisis de secuencias comparativo a gran escala para correlacionar la información de secuencia con la función, el fenotipo o la identidad. La aplicación de tales tecnologías para el análisis de secuencias comparativo puede estar generalizada, incluyendo descubrimiento de SNP e identificación de patógenos específica de secuencia. Por tanto, las tecnologías de resecuenciación y cribado mutacional de alto rendimiento son críticas para la identificación de mutaciones subyacentes a la enfermedad, así como para la variabilidad genética subyacente a la respuesta farmacológica diferencial.

Se han desarrollado varios enfoques con el fin de satisfacer estas necesidades. Una tecnología actual para la secuenciación de ADN de alto rendimiento incluye secuenciadores de ADN que usan detección por fluorescencia inducida por láser y electroforesis. Los métodos de secuenciación basados en electroforesis tienen limitaciones inherentes para detectar heterocigotos y están comprometidos por compresiones de GC. Por tanto, una plataforma de secuenciación de ADN que produce datos digitales sin usar electroforesis superará estos problemas. La espectrometría de masas con tiempo de vuelo de desorción/ionización láser asistida por matriz (MALDI-TOF MS) mide fragmentos de ADN con la salida de datos digitales. Los métodos de análisis de fragmentación por escisión específica proporcionados en el presente documento permiten alto rendimiento, alta velocidad y alta precisión en la detección de variaciones de secuencia con respecto a una secuencia de referencia. Este enfoque posibilita el uso rutinario de secuenciación con MALDI-TOF MS para una detección de mutaciones precisa, tal como cribado para mutaciones fundadoras en BRCA1 y BRCA2, que están vinculados al desarrollo del cáncer de mama.

Por tanto, la presente divulgación proporciona en parte un método para resecuenciar una secuencia de nucleótidos diana, que comprende (a) proporcionar un ácido nucleico de muestra (por ejemplo, tomado de un sujeto); (b) preparar una mezcla del ácido nucleico de muestra con un soporte sólido descrito en el presente documento que tiene un ácido nucleico en fase sólida en condiciones en las que una única molécula del ácido nucleico de muestra se hibrida con una molécula de soporte sólido; (c) poner en contacto la mezcla con agentes de extensión en condiciones en las que se extiende el ácido nucleico en fase sólida hibridado con el ácido nucleico de muestra; (d) amplificar el ácido nucleico extendido; (e) analizar la secuencia del ácido nucleico amplificado; y (f) comparar una secuencia determinada en la parte (e) con una secuencia de nucleótidos de referencia, mediante lo cual se resecuencia la secuencia de nucleótidos diana. La secuencia de referencia puede ser una secuencia de nucleótidos ya identificada a partir de la muestra. En algunas implementaciones, la muestra puede procesarse antes de la etapa (b), purificando el ácido nucleico en la muestra y/o fragmentando el ácido nucleico de muestra, por ejemplo.

#### 16. *Multiplexación*

Los procedimientos y soportes sólidos descritos en el presente documento pueden permitir la detección de alto rendimiento o el descubrimiento de secuencias en una pluralidad de secuencias diana. Multiplexación se refiere a la detección simultánea de más de una secuencia, un polimorfismo o una variación de secuencia. La multiplexación permite el procesamiento simultáneo de muchos moldes de secuenciación agrupándolos en las etapas más tempranas del procedimiento de preparación y los resuelve en secuencias individuales en la etapa más tardía posible del procedimiento de secuenciación, permitiendo de ese modo un alto rendimiento de los moldes con una reducción de etapas repetitivas. Se conocen métodos para realizar reacciones multiplexadas, particularmente junto con espectrometría de masas (véanse, por ejemplo, las patentes estadounidenses n.ºs 6.043.031, 5.547.835 y la solicitud PCT internacional n.º WO 97/37041).

La multiplexación puede realizarse, por ejemplo, para la misma secuencia de ácido nucleico diana usando reacciones de escisión específicas complementarias diferentes tal como se proporciona en el presente documento, o para secuencias de ácido nucleico diana diferentes, y los patrones de escisión pueden analizarse a su vez contra una pluralidad de secuencias de ácido nucleico de referencia. Varias mutaciones o variaciones de secuencia también pueden detectarse simultáneamente en una secuencia diana empleando los métodos proporcionados en el presente documento, donde cada variación de secuencia corresponde a un producto de escisión diferente con respecto al patrón de escisión de la secuencia de ácido nucleico de referencia. La multiplexación proporciona la ventaja de que pueden identificarse una pluralidad de variaciones de secuencia en tan sólo un único espectro de masas, en comparación con tener que realizar un análisis por espectrometría de masas independiente para cada variación de secuencia individual. Los métodos proporcionados en el presente documento se prestan por sí mismos a procedimientos de alto rendimiento y altamente automatizados para analizar variaciones de secuencia con alta velocidad y precisión, con la ventaja añadida de la identificación de secuencias que normalmente no pueden leerse

usando métodos basados en electroforesis en gel. En algunas implementaciones, el análisis múltiple de secuencias también puede combinarse con otros métodos no limitativos habitualmente conocidos en la técnica, tales como secuenciación de ADN por degradación con exonucleasas, por ejemplo.

Por tanto, en el presente documento se proporciona un método para analizar una secuencia de nucleótidos diana, que comprende (a) proporcionar un ácido nucleico a partir de una muestra; (b) preparar una mezcla del ácido nucleico de muestra con un soporte sólido descrito en el presente documento que tiene un ácido nucleico en fase sólida en condiciones en las que una única molécula del ácido nucleico de muestra se hibrida con una molécula de soporte sólido; (c) poner en contacto la mezcla con agentes de extensión en condiciones en las que se extiende el ácido nucleico en fase sólida hibridado con el ácido nucleico de muestra; (d) amplificar el ácido nucleico extendido; (e) analizar la secuencia del ácido nucleico amplificado; y (f) identificar dos más secuencias en el ácido nucleico de muestra. En algunas implementaciones, la muestra puede procesarse antes de la etapa (b), purificando el ácido nucleico en la muestra y/o fragmentando el ácido nucleico de muestra, por ejemplo. La parte (d) es opcional en determinadas implementaciones: el ácido nucleico en fase sólida extendido de (c) puede analizarse mediante secuenciación (e), sin amplificación (d), usando los métodos de secuenciación por síntesis descritos anteriormente.

#### 17. Monitorización de brotes epidémicos de enfermedad

Los procedimientos y soportes sólidos descritos en el presente documento pueden usarse para monitorizar brotes epidémicos de enfermedad. En tiempos de transporte y desplazamiento globales, los brotes epidémicos de organismos endémicos patógenos requieren la estrecha monitorización para prevenir su propagación mundial y permitir su control. El tipado basado en ADN mediante tecnologías de alto rendimiento (por ejemplo, usando chips de ADN, tecnologías de alineamientos de ADN, y similares) permiten un rendimiento rápido de la muestra en un tiempo comparativamente corto, tal como se requiere en una situación de brote epidémico. Actualmente, los métodos tradicionales de monitorización de brotes epidémicos de enfermedad pueden tardar hasta de 7 a 10 días para identificar los microorganismos patógenos. El uso de tecnologías de alto rendimiento puede ofrecer un ahorro de tiempo significativo en las etapas iniciales críticas de la monitorización de brotes epidémicos de enfermedad, reduciendo los tiempos de identificación desde de 7 a 10 días hasta menos de 2 días. La monitorización se realiza detectando una o más regiones de marcadores microbianos (por ejemplo, SNP, regiones singulares de ARNr, y similares) en una o más muestras. Puede monitorizarse un género, una especie, una cepa o un subtipo de un microorganismo, usando marcadores moleculares para identificar la presencia o ausencia de secuencias de ácido nucleico específicas de microorganismos patógenos conocidos.

Por tanto, en el presente documento se proporciona un método para monitorizar un brote epidémico de enfermedad, que comprende (a) proporcionar un ácido nucleico a partir de una muestra; (b) preparar una mezcla del ácido nucleico de muestra con un soporte sólido descrito en el presente documento que tiene un ácido nucleico en fase sólida en condiciones en las que una única molécula del ácido nucleico de muestra se hibrida con una molécula de soporte sólido; (c) poner en contacto la mezcla con agentes de extensión en condiciones en las que se extiende el ácido nucleico en fase sólida hibridado con el ácido nucleico de muestra; (d) amplificar el ácido nucleico en fase sólida extendido; (e) analizar la secuencia del ácido nucleico amplificado en (d); y (f) comparar una secuencia determinada en (e) con una secuencia de referencia, mediante lo cual se monitoriza el brote epidémico de enfermedad. La muestra puede procesarse antes de la etapa (b), purificando los ácidos nucleicos y/o fragmentando los ácidos nucleicos, por ejemplo. El brote epidémico de enfermedad puede monitorizarse determinando si (i) hay nuevas secuencias tal como se determina en la parte (e) no presentes en una muestra de referencia (por ejemplo, lo que indica que están presentes nuevos patógenos en una población como parte de un brote epidémico de enfermedad) y (ii) hay menos secuencias tal como se determina en la parte (e) presentes en una muestra de referencia (por ejemplo, lo que indica que determinados patógenos ya no constituyen una amenaza como parte del brote epidémico de enfermedad). Una secuencia de referencia puede proceder de una muestra tomada del/de los mismo(s) individuo(s) en un punto de tiempo diferente (por ejemplo, punto de tiempo anterior). En algunas implementaciones, la muestra puede procesarse antes de la etapa (b), purificando el ácido nucleico en la muestra y/o fragmentando el ácido nucleico de muestra, por ejemplo. La parte (d) es opcional en determinadas implementaciones: el ácido nucleico en fase sólida extendido de (c) puede analizarse mediante secuenciación (e), sin amplificación (d), usando los métodos de secuenciación por síntesis descritos anteriormente. En algunas implementaciones, el análisis de secuencias comparativo multiplexado junto con el análisis por espectrometría de masas con tiempo de vuelo de desorción/ionización láser asistida por matriz (MALDI-TOF MS) pueden usarse con las implementaciones descritas en el presente documento para monitorizar brotes epidémicos de enfermedad.

#### 18. Control de calidad de vacunas y control de calidad de clones de producción

Los procedimientos y soportes sólidos descritos en el presente documento pueden usarse para controlar la identidad de clones de producción recombinantes, que pueden ser vacunas o, por ejemplo, insulina o cualquier otro clon de producción o producto biológico o médico. Puede analizarse toda la secuencia o una o más porciones de un clon o una vacuna en los lotes y las muestras de producción. Las secuencias determinadas mediante los procedimientos descritos en el presente documento pueden compararse con una secuencia de referencia para el clon o la vacuna para monitorizar el control de calidad. Una monitorización del control de calidad de este tipo puede permitir la detección de mutaciones espontáneos o un reordenamiento genético en diversas etapas de producción en

biorreactores a gran escala, permitiendo de ese modo una gestión de recursos más eficiente al permitir la detección y la parada tempranas de los procedimientos que se han monitorizado y muestran una desviación del producto esperado.

Por tanto, en el presente documento se proporciona un método para determinar la calidad de una muestra de clon o vacuna de producción, que comprende (a) proporcionar un ácido nucleico de muestra de clon o vacuna de producción; (b) preparar una mezcla del ácido nucleico de muestra con un soporte sólido descrito en el presente documento que tiene un ácido nucleico en fase sólida en condiciones en las que una única molécula del ácido nucleico de muestra se hibrida con una molécula de soporte sólido; (c) poner en contacto la mezcla con agentes de extensión en condiciones en las que se extiende el ácido nucleico en fase sólida hibridado con el ácido nucleico de muestra; (d) amplificar el ácido nucleico extendido; (e) analizar la secuencia de ácido nucleico amplificado de la parte (d); y (f) comparar una secuencia determinada mediante la parte (e) con una secuencia de referencia de clon o vacuna, mediante lo cual se determina la calidad del clon o la vacuna de producción basándose en la comparación en la parte (f). La comparación en la parte (f) puede ser el grado de identidad entre toda la secuencia o subsecuencia del clon o la vacuna de producción y una secuencia correspondiente en el clon o la vacuna de referencia. Una secuencia de referencia puede obtenerse a partir de un lote de producción diferente o a partir de una vacuna o un clon progenitor, por ejemplo. En algunas implementaciones, la muestra puede procesarse antes de la etapa (b), purificando el ácido nucleico en la muestra y/o fragmentando el ácido nucleico de muestra, por ejemplo. La parte (d) es opcional en determinadas implementaciones: el ácido nucleico en fase sólida extendido de (c) puede analizarse mediante secuenciación (e), sin amplificación (d), usando los métodos de secuenciación por síntesis descritos anteriormente.

## Ejemplos

### Ejemplo 1: Metodología de análisis de secuencias

A continuación se describe una metodología para realizar los análisis de secuencias de ácido nucleico descritos en el presente documento. Los oligonucleótidos sintetizados que contienen secuencias de sonda, de cebador y de identificación se unen a un soporte sólido (perlas, portaobjetos, chips, y similares, y en algunas realizaciones, perlas Dynal®) habitualmente disponible en la técnica, a través de una química de unión apropiada. En algunas realizaciones que usan perlas Dynal®, puede usarse la química de unión carboxilo-amino para unir los oligonucleótidos sintetizados a las perlas. La síntesis de oligonucleótidos se conoce bien en la técnica, y pueden elegirse una variedad de métodos para sintetizar oligonucleótidos y bibliotecas de oligonucleótidos, incluyendo métodos que incorporan nucleótidos modificados o derivatizados diseñados para aumentar la estabilidad biológica de las moléculas o para aumentar la estabilidad física del dúplex formado entre los ácidos nucleicos antisentido y sentido (por ejemplo, derivados de fosforotioato y nucleótidos sustituidos con acridina).

Las secuencias de nucleótidos para los oligonucleótidos sintetizados pueden incluir cualquier secuencia de ácido nucleico útil para procedimientos de investigación biológica o clínica (por ejemplo, SNP, secuencias de sonda conocidas específicas de microorganismos patógenos, y similares) incluyendo, pero sin limitarse a, las aplicaciones y los usos descritos en el presente documento. En algunas realizaciones, los oligonucleótidos sintetizados pueden unirse a un soporte sólido en condiciones diluidas de manera que uno o sólo algunos oligonucleótidos se unen a cada unidad individual de soporte sólido (1, 2, 3, 4, 5 o hasta 10 oligonucleótidos sintetizados unidos), cuando se usan perlas o partículas como soporte sólido, por ejemplo. En algunas realizaciones con más de un oligonucleótido unido, los oligonucleótidos unidos no son idénticos en cuanto a secuencia.

El ácido nucleico de muestra se prepara y pone en contacto con los oligonucleótidos sintetizados que contienen secuencias de sonda, de cebador y de identificación unidos a un soporte sólido (a continuación en el presente documento denominados oligonucleótidos en fase sólida, u oligos en fase sólida). El ácido nucleico de muestra puede prepararse mediante cualquier medio habitualmente conocido en la técnica, incluyendo pero sin limitarse a procedimientos de lisis celular usando métodos de lisis químicos, físicos o electrolíticos. Por ejemplo, los métodos químicos generalmente emplean agentes de lisis para romper las células y extraer los ácidos nucleicos a partir de las células, seguido del tratamiento con sales caotrópicas. Los métodos físicos tales como congelación/descongelación seguido de trituración, el uso de prensas celulares, y similares, también son útiles si se desean proteínas intactas. Habitualmente también se usan procedimientos de lisis con alto contenido de sales. Estos procedimientos pueden encontrarse en Current Protocols in Molecular Biology, John Wiley & Sons, N.Y., 6.3.1-6.3.6 (1989).

El ácido nucleico de muestra puede manipularse o prepararse adicionalmente después de la lisis celular y el aislamiento de ácido nucleico. Pueden usarse métodos de preparación de ácido nucleico incluyendo, pero sin limitarse a, cizalladura, fraccionamiento por tamaño, purificación, metilación o desmetilación, tratamiento con nucleasas de restricción, adición de nucleótidos o ligadores (definidos en el presente documento como oligómeros cortos de nucleótidos de secuencia específica o inespecífica), incorporación de marcadores detectables, y similares, para preparar ácidos nucleicos de muestra para la puesta en contacto con los oligonucleótidos en fase sólida. Por ejemplo, en algunas realizaciones, un ADN genómico de muestra puede someterse a cizalladura, diluirse y mezclarse con un exceso molar de perlas. En algunas realizaciones, el ADN genómico mezclado, en condiciones

diluidas, con un exceso molar de oligos en fase sólida (usando las razones molares descritas anteriormente) permite la unión de una molécula de ácido nucleico de muestra a una perla. El ácido nucleico de muestra y los oligos en fase sólida pueden hibridarse en cualesquiera condiciones de hibridación útiles conocidas en la técnica, cuyos ejemplos no limitativos se describieron anteriormente. En algunas realizaciones, después de la hibridación, puede aislarse el complejo oligo en fase sólida/ácido nucleico de muestra. El aislamiento de estos complejos puede permitir la eliminación de posibles contaminantes de la etapa de amplificación.

Las perlas y la muestra ADN pueden mezclarse con los componentes de la reacción en cadena de la polimerasa, y la mezcla puede emulsionarse con aceite mineral (por ejemplo, Margulies *et al.*, "Genome sequencing in open microfabricated high density picoliter reactors", *Nature*: 376-380 (2005); Kojima *et al.*, "PCR amplification from single DNA molecules on magnetic beads in emulsion: application for high-throughput screening of transcription factor targets", *Nucleic Acids Research* 33(17) (2005)), en condiciones que permitan la extensión y la amplificación (lineal o exponencial, según requiera el experto) del oligonucleótido en fase sólida usando el ácido nucleico de muestra hibridado como molde. La extensión de oligos en fase sólida, usando el ácido nucleico de muestra como molde, da como resultado un ácido nucleico en fase sólida extendido que es sustancialmente complementario (es decir, antisentido) al ácido nucleico de muestra.

Después de extender los ácidos nucleicos en fase sólida, los ácidos nucleicos extendidos pueden secuenciarse mediante cualquier protocolo de secuenciación conocido en la técnica incluyendo, pero sin limitarse a, las metodologías de secuenciación descritas anteriormente (por ejemplo, secuenciación por ligación, pirosecuenciación, secuenciación por síntesis) o tal como se describe en Bently, "Whole genome resequencing", *Curr Opin Genet Dev* 16(6):545-52 (2006); Shendure *et al.*, "Accurate multiplex polony sequencing of an evolved bacterial genome", *Science* 309(5741):1728-32 (2005); Ju *et al.*, "Four-color DNA sequencing by synthesis using cleavable fluorescent nucleotide reversible terminators", *Proc Natl Acad Sci USA* 103(52):19635-40 (2006)), por ejemplo. Los datos de secuencia sin procesar pueden almacenarse para su recopilación y análisis de secuencia posteriores.

Las lecturas de secuencia pueden recopilarse usando algoritmos de recopilación en secuencias de longitud completa (por ejemplo, Warren *et al.*, "Assembling millions of short DNA sequences using SSAKE", *Bioinformatics* 23(4):500-1 (2006); Jeck *et al.*, "Extending assembly of short DNA sequences to handle error", *Bioinformatics* 23(21):2942-4 (2007)). Después de analizar los datos de secuencia, pueden realizarse determinaciones específicas según el método, el procedimiento o la aplicación realizado por el experto (por ejemplo, detección de la presencia o ausencia de organismos patógenos, control de calidad para procedimientos en biorreactores, determinaciones de la frecuencia alélica, detección de variaciones alélicas en la expresión génica, y similares).

El ejemplo descrito anteriormente puede usarse para detectar, identificar y secuenciar ácidos nucleicos virales encontrados en mezclas virales (por ejemplo, encontrar genotipos o serotipos de la hepatitis B en una mezcla viral de hepatitis), o poblaciones virales mixtas que podrían encontrarse en muestras aisladas a partir de fuentes ambientales o a partir de organismos inmunodeficientes. La metodología para detectar, identificar y secuenciar ácidos nucleicos virales es sustancialmente similar a la descrita anteriormente para la secuenciación de ácidos nucleicos.

## REIVINDICACIONES

1. Colección de soportes sólidos que comprende especies de ácido nucleico monocatenario,
  - (a) en la que los soportes sólidos son perlas o partículas,
  - (b) en la que cada soporte sólido en la colección comprende más de una especie de ácido nucleico monocatenario covalentemente unida al soporte sólido,
  - (c) en la que cada especie de ácido nucleico monocatenario unida a un soporte sólido común comprende una secuencia identificadora y una secuencia de sonda,
  - (d) en la que todas las especies de ácido nucleico monocatenario unidas al soporte sólido común comparten una secuencia identificadora común, y
  - (e) en la que la secuencia identificadora de las más de una especies de ácido nucleico monocatenario unidas al soporte sólido común en la colección es diferente de las secuencias identificadoras de las especies de ácido nucleico monocatenario unidas a los demás soportes sólidos en la colección, mediante lo cual cada soporte sólido en la colección está asociado con una secuencia identificadora singular que difiere, cada una, de todas las demás secuencias identificadoras en uno o más nucleótidos, cuando las secuencias identificadoras se alinean entre sí y en la que sólo está presente una especie de secuencia de identificación singular por especie de soporte sólido,
  - (f) en la que la secuencia de sonda de cada una de las más de una especies de ácido nucleico monocatenario unidas al soporte sólido permite que los más de un ácidos nucleicos monocatenarios se hibriden con una subsecuencia en un ácido nucleico de muestra, y
  - (g) en la que las secuencias de sonda de las más de una especies de ácido nucleico monocatenario en cada soporte sólido comprenden al menos 10 especies de secuencia de sonda diferentes, en la que cada especie de secuencia de sonda difiere de todas las demás especies de secuencia de sonda en uno o más nucleótidos, cuando las especies de secuencia de sonda se alinean entre sí.
2. Colección de soportes sólidos según la reivindicación 1, en la que la especie de ácido nucleico monocatenario de cada soporte sólido en la colección comprende además una secuencia de cebador.
3. Colección de soportes sólidos según la reivindicación 2, en la que la secuencia de cebador, la secuencia identificadora y la secuencia de sonda están orientadas 5'-(secuencia de cebador)-(secuencia identificadora)-(secuencia de sonda)-3'.
4. Colección de soportes sólidos según la reivindicación 2, en la que las especies de ácido nucleico monocatenario unidas a un soporte sólido común en la colección comparten una secuencia de cebador común, o en la que las especies de ácido nucleico monocatenario de todos los soportes sólidos en la colección comparten una secuencia de cebador común.
5. Colección de soportes sólidos según la reivindicación 1, en la que los soportes sólidos son microperlas, nanopérlas, micropartículas o nanopartículas.
6. Colección de soportes sólidos según la reivindicación 1, en la que las perlas o partículas comprenden un material magnético.
7. Colección de soportes sólidos según la reivindicación 1, en la que las más de una secuencias de sonda son complementarias a una secuencia de ADN, una secuencia de ADN fragmentado o un ácido nucleico monocatenario.
8. Sustrato que comprende una colección de soportes sólidos según la reivindicación 5, en el que:
  - (a) los soportes sólidos son perlas; y
  - (b) las perlas están orientadas en una matriz.
9. Sustrato según la reivindicación 8, en el que la especie de ácido nucleico monocatenario de cada perla en la colección comprende además una secuencia de cebador.
10. Sustrato según la reivindicación 9, en el que la secuencia de cebador, la secuencia identificadora y la secuencia de sonda están orientadas 5'-(secuencia de cebador)-(secuencia identificadora)-(secuencia de sonda)-3'.



11. Sustrato según la reivindicación 9, en el que las especies de ácido nucleico monocatenario unidas a una perla común en la colección comparten una secuencia de cebador común, o
- 5 en el que las especies de ácido nucleico monocatenario de todas las perlas en la colección comparten una secuencia de cebador común.
12. Sustrato según la reivindicación 8, en el que el soporte sólido es una microperla o una nanopera, y/o en el que la colección de soportes sólidos está comprendida en una emulsión.
- 10
13. Sustrato según la reivindicación 8, en el que la perla comprende un gel o un material magnético.
14. Sustrato según la reivindicación 8, en el que las más de una secuencias de sonda son complementarias a una secuencia de ADN o secuencia de ADN fragmentado.
- 15
15. Método de fabricación de la colección de soportes sólidos según la reivindicación 1, que comprende
- (a) unir secuencialmente nucleótidos a un nucleótido covalentemente unido a cada soporte sólido en la colección, mediante lo cual se prepara cada una de las especies de ácido nucleico monocatenario en asociación con cada soporte sólido; o
- 20
- (b) unir cada especie de ácido nucleico monocatenario en fase de disolución a cada soporte sólido en la colección, mediante lo cual las más de una especies de ácido nucleico monocatenario en fase sólida están en asociación con cada uno de los soportes sólidos; en el que:
- 25
- (i) cada especie de ácido nucleico monocatenario en fase sólida comprende una secuencia identificadora y una secuencia de sonda, en el que las especies de ácido nucleico en fase sólida del soporte sólido comparten una secuencia de sonda común o no comparten una secuencia de sonda común,
- 30
- (ii) en el que todas las especies de ácido nucleico monocatenario unidas al soporte sólido común comparten una secuencia identificadora común, y
- 35 en el que la secuencia identificadora de las más de una especies de ácido nucleico monocatenario unidas al soporte sólido común en la colección es diferente de la secuencia identificadora de las especies de ácido nucleico monocatenario unidas a los demás soportes sólidos en la colección, mediante lo cual cada soporte sólido en la colección está asociado con una secuencia identificadora singular que difiere, cada una, de todas las demás secuencias identificadoras en uno o más nucleótidos, cuando las secuencias identificadoras se alinean entre sí y en la que sólo está presente una especie de secuencia de identificación singular por especie de soporte sólido, y
- 40
- (iii) en el que la secuencia de sonda de cada una de las más de una especies de ácido nucleico monocatenario unidas al soporte sólido permite que los más de un ácidos nucleicos monocatenarios se hibriden con una subsecuencia en un ácido nucleico de muestra, y
- 45
- (c) en el que las secuencias de sonda de las más de una especies de ácido nucleico monocatenario en cada soporte sólido comprenden al menos 10 especies de secuencia de sonda diferentes, en el que cada especie de secuencia de sonda difiere de todas las demás especies de secuencia de sonda en uno o más nucleótidos, cuando las especies de secuencia de sonda se alinean entre sí.
- 50

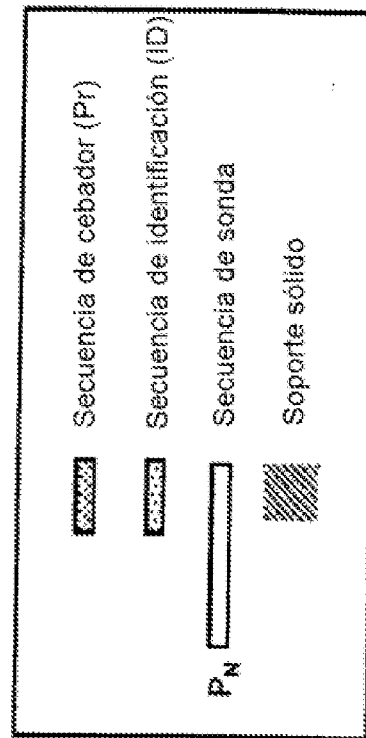
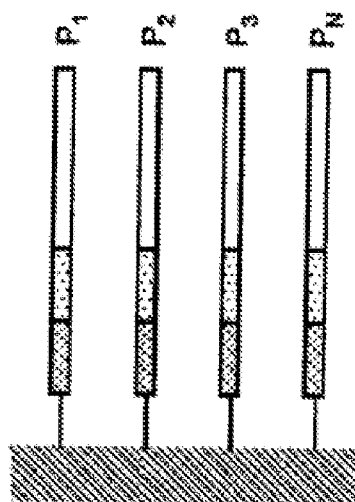
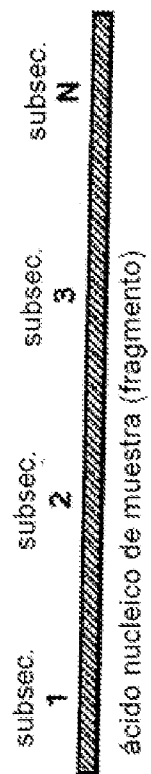


Figura 1A

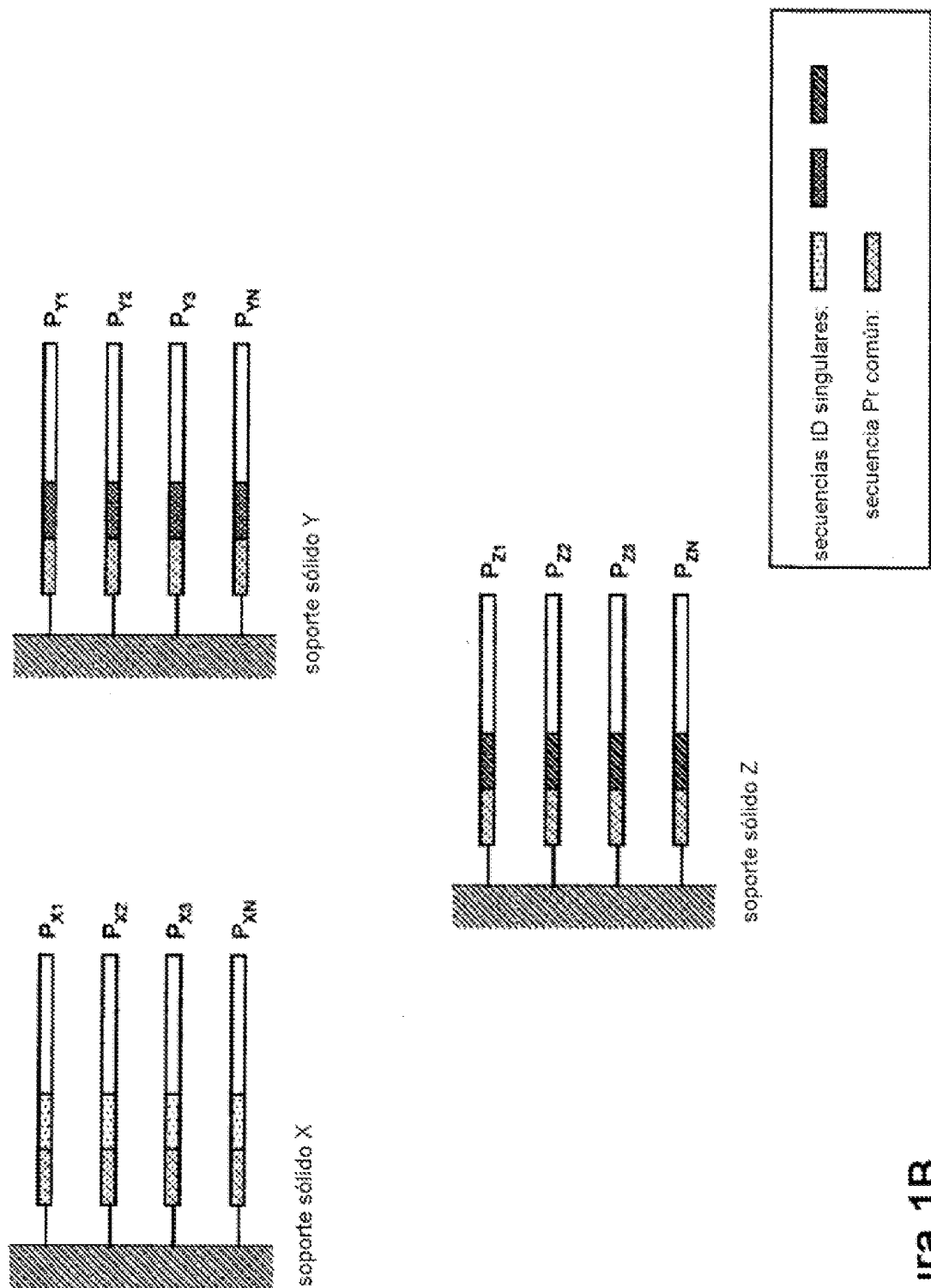


Figura 1B

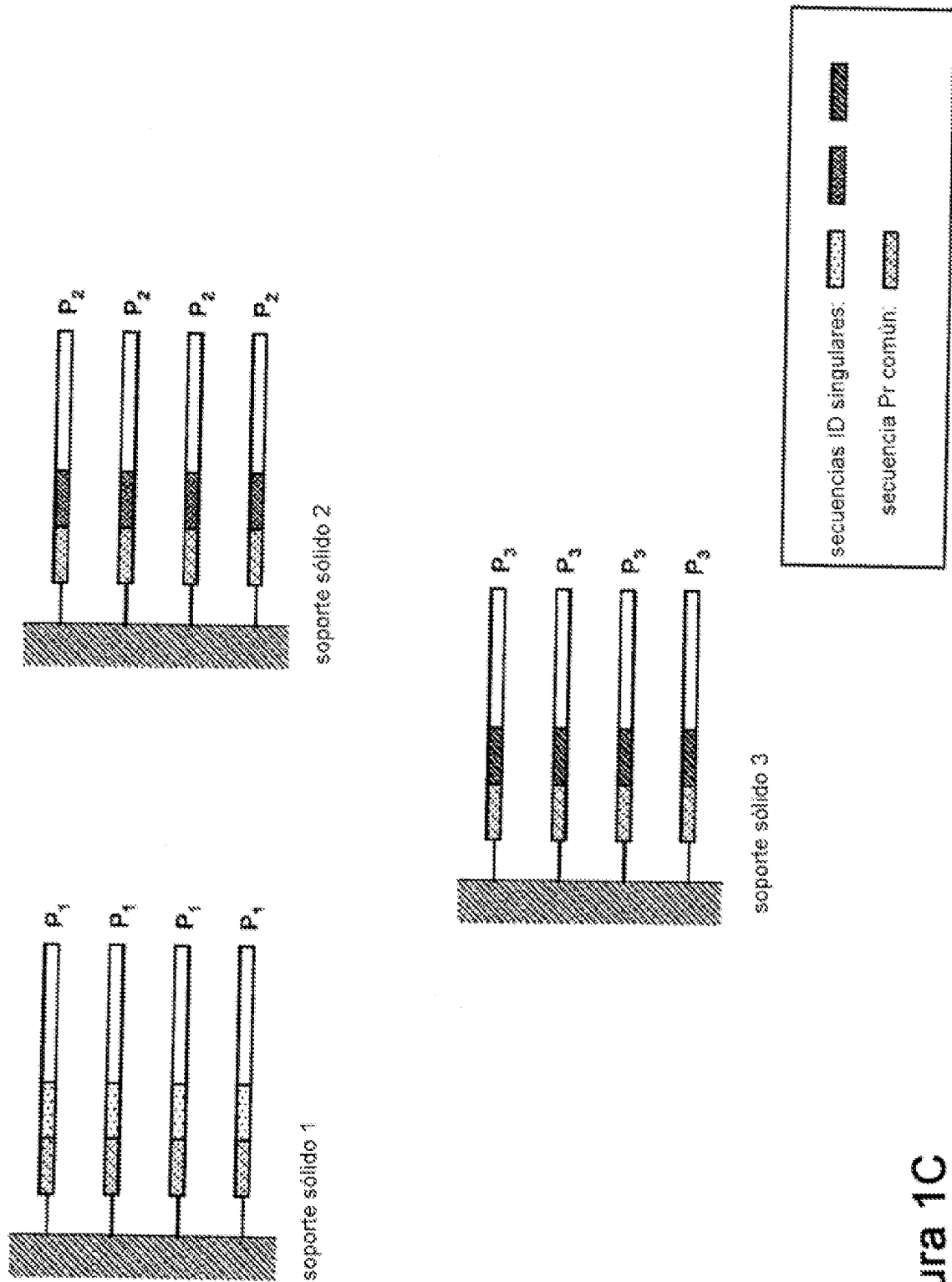


Figura 1C

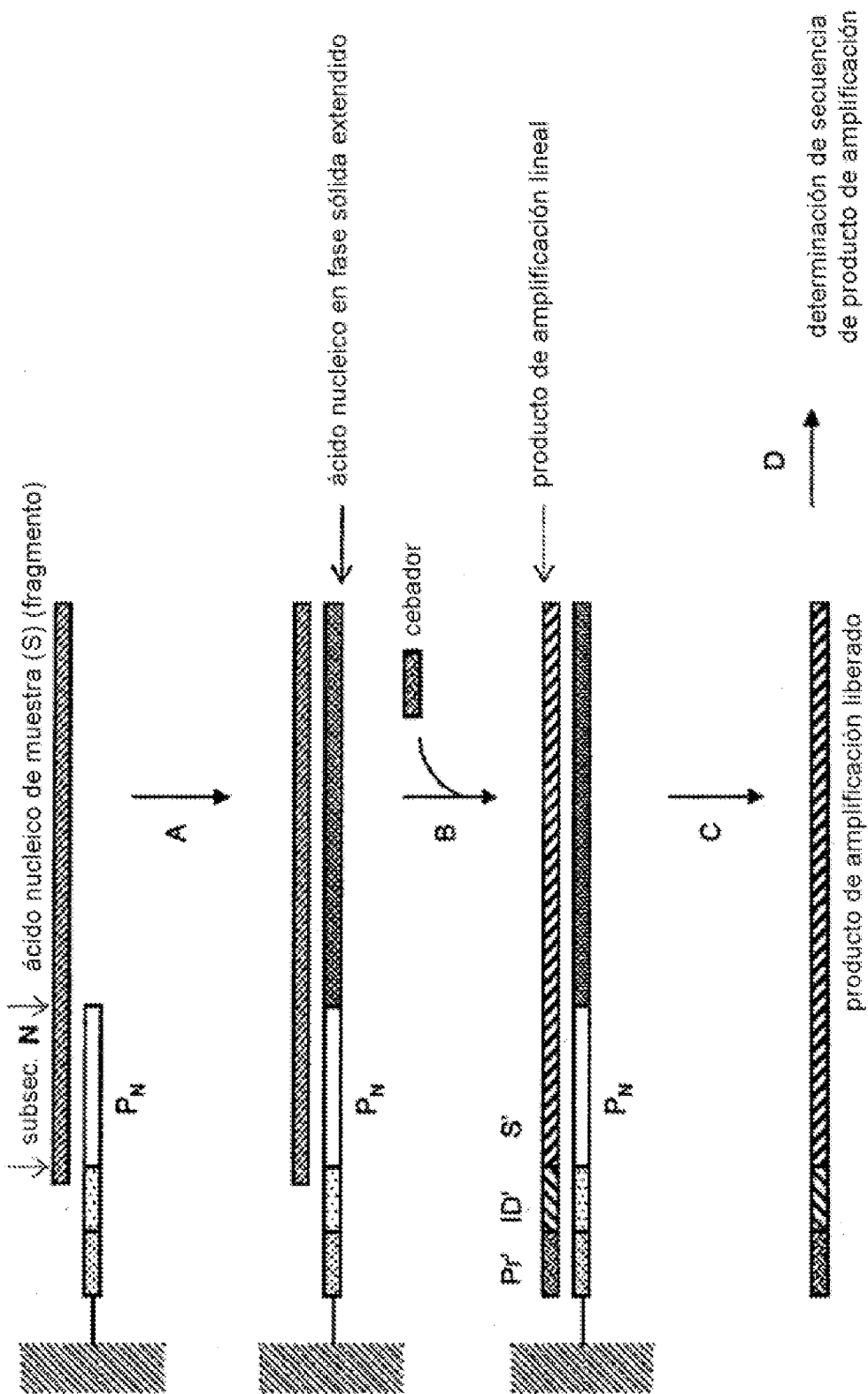


Figura 2

## REFERENCIAS CITADAS EN LA DESCRIPCION

La lista de referencias citadas por el solicitante es, únicamente, para conveniencia del lector. No forma parte del documento de patente europea. Si bien se ha tenido gran cuidado al recopilar las referencias, no pueden excluirse errores u omisiones y la OEP declina toda responsabilidad a este respecto.

## Documentos citados en la descripción

- \* US 20050112590, Van Den Boom [0025]
- \* US 5786146 A [0029]
- \* US 20030180779 A [0029]
- \* US 20030082600 A [0029]
- \* US 8133438 A, Koster [0039]
- \* WO 9106307 A, Van Ness [0039] [0063]
- \* US 4488863 A [0040]
- \* US 5536621 A [0040]
- \* US 5541306 A [0040]
- \* US 5637683 A [0040]
- \* US 5637684 A [0040]
- \* US 5760922 A [0040]
- \* US 5717083 A [0040]
- \* US 5716262 A [0040]
- \* US 5739308 A [0040]
- \* US 5773601 A [0040]
- \* US 5886165 A [0040]
- \* US 5926226 A [0040]
- \* US 5977296 A [0040]
- \* US 6140432 A [0040]
- \* WO 0050740 A [0040]
- \* WO 0114396 A [0040]
- \* US 5614622 A [0040]
- \* US 5739314 A [0040]
- \* US 5955599 A [0040]
- \* US 5962674 A [0040]
- \* US 6117992 A [0040]
- \* WO 0076372 A [0040]
- \* US 20040110191, Winkler [0069]
- \* US 20050214625, Stuehnagel [0069]
- \* US 7189314 B [0082]
- \* EP 0268520 A [0096] [0105]
- \* EP 0655501 A [0096] [0105]
- \* WO 9400562 A [0096] [0105]
- \* WO 9213666 A [0117]
- \* US 6043031 A [0139]
- \* US 5547835 A [0139]
- \* WO 9737041 A [0139]

## Literatura no patente citada en la descripción

- \* CARUTHERS. *Science*, 1985, vol. 230, 281-286 [0038]
- \* CLAEBOE et al. *Nucleic Acids Res.*, 2003, vol. 31 (19), 5885-5891 [0038]
- \* S. S. WONG. *Chemistry of Protein Conjugation and Cross-Linking*. CRC Press, 1991 [0039]
- \* G. T. HERMANSON. *Bioconjugate Techniques*. Academic Press, 1995 [0039]
- \* MEYERS ; MILLER. *CABIOS*, 1989, vol. 4, 11-17 [0042]
- \* NAKANO et al. Single-molecule PCR using water-in-oil emulsion. *Journal of Biotechnology*, 2003, vol. 102, 117-124 [0060] [0069] [0080]
- \* *Nucleic Acid Hybridization, A Practical Approach*. IRL Press, 1985 [0063]
- \* VENTER et al. *Science*, 2001, vol. 291, 1304-1351 [0074]
- \* MARGULIES, M. et al. *Nature*, 2005, vol. 437, 378-380 [0076]
- \* HARRIS TD et al. *Science*, 2008, vol. 320, 106-109 [0076]
- \* SONI GV ; MELLER A. *Clin Chem*, 2007, vol. 53, 1996-2001 [0076]
- \* *Dear Brief Funct Genomic Proteomic*, 2003, vol. 1, 397-416 [0076]
- \* BRASLAVSKY et al. *PNAS*, 2003, vol. 100 (7), 3950-3964 [0082]
- \* WANG et al. *Science*, 1998, vol. 280, 1077-1082 [0094]
- \* RATNER et al. *Nature*, 1985, vol. 313, 227-284 [0098] [0105]
- \* WAIN HOBSON et al. *Cell*, 1995, vol. 40, 9-17 [0098] [0105]
- \* GUYADER et al. *Nature*, 1987, vol. 328, 682-689 [0098] [0105]
- \* CHAKRABARTI et al. *Nature*, 1987, vol. 328, 543-547 [0098] [0105]
- \* GUST et al. *Immunology*, 1983, vol. 20, 1-7 [0098] [0105]
- \* YIN et al. Role of hepatitis B virus genotype mixture, subgenotypes C2 and B2 on hepatocellular carcinoma: compared with chronic hepatitis B and asymptomatic carrier state in the same area. *Carcinogenesis*, 2008, vol. 29 (9), 1685-1691 [0103]
- \* ROHWER ; EDWARDS. The phage proteomic tree: a genome based taxonomy for phage. *Journal of Bacteriology*, 2002, vol. 184, 4529-4535 [0104]
- \* GRIFFIN et al. Pathogenic human viruses in coastal waters. *Clinical Microbiol. Rev.*, 2003, vol. 16, 129-143 [0104]
- \* HEYM et al. *Lancet*, 1994, vol. 344, 293 [0107]

- MORRIS et al. *J. Infect. Dis.*, 1995, vol. 171, 954 [0107]
- BANERJEE et al. *Science*, 1994, vol. 263, 227 [0107]
- J. L. WEBER, P. E. CAN. *Am. J. Hum. Genet.*, 1989, vol. 44, 388 [0114]
- J. WEISSENSACH et al. *Nature*, 1992, vol. 359, 794 [0114]
- C. T. CASKEY et al. *Science*, 1992, vol. 258, 784 [0115]
- P. J. MCKINNEN. *Hum. Genet.*, 1987, vol. 1 (75), 197 [0115]
- J. GERMAN et al. *Clin. Genet.*, 1989, vol. 35, 57 [0115]
- EDWARDS et al. *Nucl. Acids Res.*, 1991, vol. 19, 4791 [0117] [0118]
- BECKMANN et al. *Genomics*, 1992, vol. 12, 627-631 [0117]
- NAKAMURA et al. *Science*, 1987, vol. 235, 1616-1622 [0117]
- JEFFREYS et al. *Nature*, 1985, vol. 314, 67-73 [0117]
- LUTY et al. *Nucleic Acids Res.*, 1991, vol. 19, 4308 [0117]
- LITT et al. *Nucleic Acids Res.*, 1990, vol. 18, 4301 [0117]
- LITT et al. *Nucleic Acids Res.*, 1990, vol. 18, 5921 [0117]
- LUTY et al. *Am. J. Hum. Genet.*, 1990, vol. 46, 775-783 [0117]
- TAUTZ. *Nucl. Acids Res.*, 1989, vol. 17, 6463-6471 [0117]
- WEBER et al. *Am. J. Hum. Genet.*, 1989, vol. 44, 388-396 [0117]
- POLYMERPOULOS et al. *Nucl. Acids Res.*, 1991, vol. 19, 125 [0118]
- POLYMERPOULOS et al. *Nucl. Acids Res.*, 1991, vol. 19, 4308 [0118]
- NISHIMURA et al. *Nucl. Acids Res.*, 1992, vol. 20, 1187 [0118]
- POLYMERPOULOS et al. *Nucl. Acids Res.*, 1991, vol. 19, 4019 [0118]
- ZULIANI et al. *Nucl. Acids Res.*, 1990, vol. 18, 4958 [0118]
- POLYMERPOULOS et al. *Nucl. Acids Res.*, 1990, vol. 18, 7489 [0118]
- PLOOS et al. *Nucl. Acids Res.*, 1990, vol. 18, 4957 [0118]
- ANKER et al. *Hum. Mol. Genet.*, 1992, vol. 1, 137 [0118]
- SCHECHTER et al. *Nature Genetics*, 1994, vol. 6, 29-32 [0125]
- REYMER et al. *Nature Genetics*, 1995, vol. 10, 28-34 [0125]
- BIRD, A. *Genes Dev.*, 2002, vol. 16, 5-21 [0127]
- Current Protocols in Molecular Biology. John Wiley & Sons, 1998, 8.3.1-8.3.6 [0146]
- MARGULIES et al. Genome sequencing in open microfabricated high density picoliter reactors. *Nature*, 2005, 376-380 [0156]
- KOJIMA et al. PCR amplification from single DNA molecules on magnetic beads in emulsion: application for high-throughput screening of transcription factor targets. *Nucleic Acids Research*, 2005, vol. 33 (17) [0156]
- BENTLY. Whole genome resequencing. *Curr Opin Genet Dev*, 2006, vol. 16 (6), 545-52 [0151]
- SHENDURE et al. Accurate multiplex polony sequencing of an evolved bacterial genome. *Science*, 2005, vol. 309 (5741), 1728-32 [0151]
- JU et al. Four-color DNA sequencing by synthesis using cleavable fluorescent nucleotide reversible terminators. *Proc Natl Acad Sci USA*, 2006, vol. 103 (52), 19635-40 [0151]
- WARREN et al. Assembling millions of short DNA sequences using SSAKE. *Bioinformatics*, 2006, vol. 23 (4), 590-1 [0152]
- JECK et al. Extending assembly of short DNA sequences to handle error. *Bioinformatics*, 2007, vol. 23 (21), 2942-4 [0152]