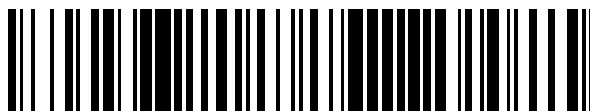


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 841 077**

51 Int. Cl.:

C12Q 1/6855 (2008.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **03.08.2016 PCT/EP2016/068546**

87 Fecha y número de publicación internacional: **09.02.2017 WO17021449**

96 Fecha de presentación y número de la solicitud europea: **03.08.2016 E 16745768 (8)**

97 Fecha y número de publicación de la concesión europea: **28.10.2020 EP 3332024**

54 Título: **Enriquecimiento de dianas por extensión del cebador de sonda individual**

30 Prioridad:

06.08.2015 US 201562201727 P

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

07.07.2021

73 Titular/es:

**F. HOFFMANN-LA ROCHE AG (100.0%)
Grenzacherstrasse 124
4070 Basel, CH**

72 Inventor/es:

GODWIN, BRIAN CHRISTOPHER

74 Agente/Representante:

LINAGE GONZÁLEZ, Rafael

ES 2 841 077 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

Enriquecimiento de dianas por extensión del cebador de sonda individual

5 CAMPO DE LA INVENCION

La divulgación se refiere en general al enriquecimiento de dianas de ácido nucleico en una muestra y, más en particular, al enriquecimiento de dianas para la secuenciación de ácido nucleico, incluyendo secuenciación de alto rendimiento.

10

ANTECEDENTES DE LA INVENCION

La invención pertenece a una clase de tecnologías que permiten a los usuarios centrarse en regiones de interés dentro del ácido nucleico que se va a secuenciar. Esto reduce costes asociados con reacciones de secuenciación y posterior análisis de datos. Actualmente existen tres tipos generales de tecnologías que capturan selectivamente regiones de interés dentro de un ácido nucleico presente en una muestra. La primera tecnología es la captura por hibridación en la que las regiones de interés se capturan a través de la hibridación de una sonda que se puede unir selectivamente a una superficie de captura. Esta captura permite la retirada de ácidos nucleicos distintos a la diana seguido de una liberación y recogida de las moléculas diana capturadas. Este tipo de tecnología tiene ventajas que incluyen la capacidad de capturar regiones de tamaño de exoma y regiones que contienen variaciones estructurales desconocidas. Las desventajas incluyen protocolos largos y complejos que tienden a tardar más de 8 horas en completarse. La complejidad está provocada principalmente por el requisito de preparar una genoteca fragmentada aleatoriamente antes de la hibridación. La etapa de hibridación sola puede tardar hasta tres días en completarse. Los ejemplos de este tipo de tecnología incluyen SeqCap EZ (NimbleGen, Madison, Wise.) y SureSelect Target Enrichment System (Agilent, Santa Clara, Cal.)

15

20

25

Otro procedimiento de enriquecimiento de dianas es la amplificación basada en cebadores de dianas dobles. En este procedimiento, las regiones de interés se enriquecen usando dos sondas en los límites de la diana. Los procedimientos tienden a tardar menos de 8 horas en completarse y son más simples que los procedimientos de captura por hibridación. Sin embargo, las tecnologías basadas en cebadores dobles no pueden enriquecer secuencias con variaciones estructurales desconocidas. El enfoque de cebadores dobles más establecido es la PCR múltiple. Es un procedimiento en una etapa individual muy simple pero solo puede amplificar decenas de dianas por tubo de reacción. Actualmente están disponibles otras tecnologías más nuevas, incluyendo los productos TruSeq Amplicon (Illumina, San Diego, Cal.) y Ion Torrent Ampliseq (Life Technologies, Grand Island, NY) que pueden amplificar de cientos a miles de dianas en un tubo de reacción individual y requieren solo unas pocas etapas de manipulación.

30

35

La tercera tecnología es la amplificación basada en cebador de diana individual. En este procedimiento, las dianas se enriquecen a través de la amplificación de una región que se define por un cebador de diana individual y un cebador universal ligado al extremo. Similar al enfoque basado en hibridación; estas tecnologías requieren que se genere una genoteca fragmentada aleatoriamente antes de la hibridación selectiva de un oligonucleótido diana. Sin embargo, en lugar de usar este oligonucleótido para capturar la diana y separar por lavado las moléculas distintas a la diana, se emplea una etapa de amplificación que amplifica selectivamente las regiones entre el extremo generado aleatoriamente y el oligonucleótido específico de diana. La ventaja de esta tecnología es que, a diferencia de las tecnologías de cebadores dobles, permite la detección de secuencias con variaciones estructurales desconocidas. También es más rápida y más sencilla que las tecnologías basadas en hibridación. Sin embargo, este tipo de tecnología todavía es más lenta y más complicada que los enfoques basados en cebadores dobles. Los ejemplos de este tipo de tecnología son Anchored Multiplex PCR de Archer (Archer Dx, Boulder, Colorado) y Ovation® Target Enrichment System (NuGen, San Carlos, Cal.).

40

45

El documento US-2007/0020640 enseña cómo se marcan moléculas de ácido nucleico (por ejemplo, fragmentos de ADN genómico) con marcas de secuencia distintas antes de la amplificación por PCR. Sigue existiendo una necesidad insatisfecha de obtener un procedimiento rápido y simple de enriquecimiento de dianas que también se adapte para variaciones estructurales desconocidas en una secuencia diana.

50

55 BREVE SUMARIO DE LA INVENCION

La invención se define por las reivindicaciones adjuntas.

En el presente documento se divulga un procedimiento de amplificación de una secuencia diana que comprende las etapas de: poner en contacto el ácido nucleico diana con un cebador y una polimerasa, en el que el cebador comprende un sitio de unión a diana y una marca de identificación molecular única (UID); llevar a cabo una reacción de extensión por la polimerasa y una terminación para crear un producto de extensión del cebador monocatenario; ligar adaptadores a cada extremo del producto de extensión del cebador monocatenario para crear un producto de ligación, en el que los adaptadores comprenden al menos un sitio de cebado universal; amplificar el producto de ligación en una reacción de amplificación utilizando al menos un cebador que se une al al menos un sitio de cebado universal para crear la secuencia diana amplificada. En algunos modos de realización, el cebador y al menos uno de los adaptadores

60

65

comprenden sitios de ligación universal compatibles mutuamente. En algunos modos de realización, el sitio de unión a diana es una secuencia específica de diana prediseñada. En algunos modos de realización, el sitio de unión a diana es una secuencia aleatoria. En algunos modos de realización, la terminación se efectúa por un procedimiento seleccionado de la lista que consiste en cambio de temperatura, adición de un inhibidor enzimático específico, adición de un quelante, incorporación de bases que contienen uridina seguido de tratamiento con uracil-N-ADN glucosilasa. En algunos modos de realización, al menos un adaptador comprende un código de barras. El código de barras puede ser una ID de muestra múltiple (MID). La amplificación puede ser amplificación lineal o amplificación exponencial. En algunos modos de realización, el procedimiento comprende además una etapa de purificación después de al menos una de extensión del cebador y ligación.

También se divulga en el presente documento un kit para amplificar una secuencia diana que comprende: un cebador que comprende un sitio de unión a diana, una marca de identificación molecular única (UID) y un sitio de ligación universal; al menos un adaptador que comprende al menos un sitio de cebado universal, ID de muestra múltiple (MID) y un sitio de ligación universal. En algunos modos de realización, el kit comprende dos adaptadores que tienen sitios de cebado universal diferentes pero solo un adaptador que comprende el sitio de ligación universal y la MID. En algunos modos de realización, el kit comprende además uno o más de los siguientes: ácido nucleico polimerasa, ligasa, ADN polimerasa termoestable y cebadores universales.

BREVE DESCRIPCIÓN DE LOS DIBUJOS

La FIG. 1 es una representación esquemática de las etapas del procedimiento.

DESCRIPCIÓN DETALLADA DE LA INVENCION

Definiciones

Como se usa en el presente documento, "sonda" quiere decir cualquier molécula que se puede unir selectivamente a una biomolécula diana específicamente destinada, por ejemplo, una secuencia de ácido nucleico de interés que se va a unir, capturar o hibridar por las sondas.

Como se usa en el presente documento, "adaptador" quiere decir una secuencia de nucleótidos que se puede añadir a otra secuencia para importar propiedades adicionales a esa secuencia. Un adaptador puede ser mono o bicatenario, o puede tener tanto una porción monocatenaria como una porción bicatenaria.

Como se usa en el presente documento, "código de barras" quiere decir una secuencia de nucleótidos que confiere identidad a una molécula. Un código de barras puede conferir una identidad única a una molécula individual (y sus copias). Un código de barras de este tipo es una ID única (UID). Un código de barras puede conferir una identidad a toda una población de moléculas (y sus copias) provenientes de la misma fuente (por ejemplo, un paciente). Un código de barras de este tipo es una ID múltiple (MID).

Como se usa en el presente documento, "sitio de ligación" es una porción de una molécula de ácido nucleico (distinta de un extremo romo de una molécula monocatenaria) que puede facilitar la ligación. Los "sitios de ligación compatibles" presentes en dos moléculas permiten la ligación preferencial de las dos moléculas entre sí.

Como se usa en el presente documento, "ligación monocatenaria" es un procedimiento de ligación que comienza con al menos un sustrato monocatenario y que implica típicamente uno o más adaptadores bicatenarios o parcialmente bicatenarios.

Como se usa en el presente documento, "cebador universal" y "sitio de cebado universal" se refieren a un cebador y un sitio de cebado que no están presentes de forma natural en la secuencia diana. Típicamente, el sitio de cebado universal está presente en adaptadores o cebadores específicos de diana. El cebador universal se puede unir a y dirigir la extensión del cebador desde el sitio de cebado universal.

Los procedimientos de la presente invención se pueden usar como parte de un protocolo de secuenciación, que incluye un protocolo de secuenciación de molécula individual de alto rendimiento. El procedimiento de la invención genera un banco de ácidos nucleicos diana que se va a secuenciar. Los ácidos nucleicos diana en el banco pueden incorporar códigos de barras para la identificación molecular e identificación de muestras.

Extensión del cebador específico de diana

El procedimiento divulgado en el presente documento comprende una etapa de extensión del cebador lineal para el cebador específico de diana. La etapa de extensión lineal tiene varias ventajas sobre la amplificación exponencial practicada en la técnica. Cada ácido nucleico diana se caracteriza por una velocidad de síntesis única que depende de la velocidad de hibridación del cebador específico de diana y la velocidad con la que una polimerasa puede leer a través de una secuencia diana particular. Las diferencias en la velocidad de extensión y la velocidad de síntesis crean un sesgo que puede dar como resultado una ligera diferencia en una ronda individual de síntesis. Sin embargo, la

ligera diferencia se amplifica exponencialmente durante la PCR. La brecha resultante se denomina sesgo de PCR. El sesgo puede enmascarar cualquier diferencia en las cantidades iniciales de cada secuencia en la muestra e impedir cualquier análisis cuantitativo.

5 El presente procedimiento divulgado en el presente documento limita la extensión de cebadores específicos de diana (incluyendo cebadores específicos de gen y cebadores redundantes que por casualidad son específicos para un sitio de unión dentro del genoma) a una etapa individual. Cualquier amplificación exponencial se realiza con cebadores universales no sujetos a un sesgo dependiente del molde, o sujetos a un sesgo menor que el cebador específico de diana.

10 Con referencia a la figura 1, el procedimiento comprende la extensión del cebador. El procedimiento incluye una etapa de preparación de la reacción (etapa 1, hibridación del cebador) seguida de una etapa de adición de la polimerasa (etapa 2, extensión del cebador). Opcionalmente, las etapas de hibridación y extensión del cebador se realizan simultáneamente, es decir, como una etapa individual en las mismas condiciones de reacción. En otros modos de
15 realización, las etapas se realizan por separado como un procedimiento en dos etapas con distintas condiciones de reacción.

La etapa de hibridación del cebador está mediada por la región específica de diana del cebador. En algunos modos de realización, la región específica de diana se puede hibridar a la región de un gen localizado en un exón, intrón o una porción no traducida de un gen o en una porción no transcrita del gen, por ejemplo, un promotor o un potenciador. En algunos modos de realización, el gen es un gen que codifica una proteína, pero en otros modos de realización, el gen no es un gen que codifica una proteína, tal como un gen que codifica ARN o un seudogén. Aún en otros modos de realización, la región específica de diana se localiza en una región intergénica. Para dianas de ARN, el cebador puede comprender una secuencia de oligo-dT.

25 En lugar de una región específica de diana prediseñada, un cebador puede contener una secuencia redundante, es decir, una cadena de nucleótidos incorporados aleatoriamente. Un cebador de este tipo también puede encontrar un sitio de unión dentro del genoma y actuar como un cebador específico de diana para ese sitio de unión.

30 Además de la región específica de diana, el cebador puede comprender secuencias adicionales. En algunos modos de realización, estas secuencias se localizan en el extremo 5' de la región específica de diana. En otros modos de realización, puede ser posible incluir estas secuencias en otro lugar dentro del cebador siempre que la región específica de diana se pueda hibridar a la diana e impulsar la reacción de extensión del cebador como se describe a continuación. Las secuencias adicionales dentro del cebador pueden incluir una o más secuencias de códigos de
35 barras, tales como una secuencia de identificación molecular única (UID) o una secuencia de identificación de muestra múltiple (MID). Las secuencias de códigos de barras pueden estar presentes como una secuencia individual o como dos o más secuencias.

40 En algunos modos de realización, las secuencias adicionales incluyen secuencias que facilitan la ligación al extremo 5' del cebador. El cebador puede contener una secuencia de ligación universal que permite la ligación de un adaptador como se describe en la siguiente sección.

45 En algunos modos de realización, las secuencias adicionales incluyen uno o más sitios de unión para uno o más cebadores de amplificación universales.

La etapa de extensión del cebador se realiza por una ácido nucleico polimerasa. Dependiendo del tipo de ácido nucleico que se está analizando, la polimerasa puede ser una ADN polimerasa dependiente de ADN ("ADN polimerasa") o una ADN polimerasa dependiente de ARN ("transcriptasa inversa").

50 En algunos modos de realización se desea controlar la longitud de la hebra de ácido nucleico sintetizada en la reacción de extensión del cebador. (Figura 1, parada de extensión). Como se explica a continuación, la longitud de esta hebra determina la longitud del ácido nucleico sometido a las etapas posteriores del procedimiento y cualquier aplicación posterior. La reacción de extensión se puede terminar por cualquier procedimiento conocido en la técnica. La reacción se puede parar físicamente, por ejemplo, por un cambio de temperatura o la adición de un inhibidor de la polimerasa.

55 En algunos modos de realización, la reacción se para colocando la reacción en hielo. En otros modos de realización, la reacción se para elevando la temperatura para inactivar una polimerasa no termoestable. Aún en otros modos de realización, la reacción se para por la adición de un quelante, tal como EDTA que puede secuestrar un cofactor crítico para la enzima, u otro compuesto, sustancia química o biológica que puede inactivar de forma reversible o irreversible la enzima.

60 Otro procedimiento para controlar la longitud de los productos de extensión del cebador es privar la reacción de extensión limitando un componente crítico (por ejemplo, dNTP) para limitar directamente la longitud de extensión o Mg²⁺ para ralentizar la velocidad de extensión y mejorar la capacidad de controlar el punto de parada de extensión. Un experto en la técnica puede determinar de forma experimental o teórica la cantidad apropiada del componente crítico que permite la extensión del cebador limitada para proporcionar predominantemente el producto de longitud deseada.
65

Otro procedimiento para controlar la longitud de los productos de extensión del cebador es la adición de nucleótidos terminadores, incluyendo nucleótidos terminadores reversibles. Un experto en la técnica puede determinar de forma experimental o teórica una proporción apropiada de nucleótidos terminadores y no terminadores que permite la extensión del cebador limitada para producir predominantemente el producto de longitud deseada. Los ejemplos de nucleótidos terminadores incluyen dideoxinucleótidos, nucleótidos con fosfato en 2' como se describe en el documento US8163487, terminadores reversibles bloqueados en 3'-O y terminadores reversibles no bloqueados en 3' como se describe, por ejemplo, en el documento US20140242579 y Guo, J., *et al.*, *Four-color DNA sequencing with 3'-O-modified nucleotide reversible terminators and chemically cleavable fluorescent dideoxynucleotides*, P.N.A.S. 2008 105 (27) 9145-9150. Aún otro procedimiento para controlar la longitud de los productos de extensión del cebador es la adición de cantidades limitadas de uracilo (dUTP) a la reacción de extensión del cebador. El ADN que contiene uracilo se puede tratar a continuación con uracilo-N-ADN glucosilasa para producir sitios abásicos. El ADN con sitios básicos se puede degradar por tratamiento térmico con adición opcional de álcali para mejorar la eficacia de la degradación como se describe en el documento US8669061. Un experto en la técnica puede determinar de forma experimental o teórica una proporción apropiada de dUTP a dTTP en la reacción de extensión que permite la inclusión limitada de dUTP para proporcionar predominantemente el producto de longitud deseada tras el tratamiento con endonucleasa.

En algunos modos de realización, la longitud del producto de extensión se limita intrínsecamente por la longitud del ácido nucleico de entrada. Por ejemplo, el ADN libre circulante presente en el plasma sanguíneo materno tiene una longitud inferior a 200 pb teniendo la mayoría una longitud de 166 pb. Yu, S.C.Y., *et al.*, *Size-based molecular diagnostics using plasma DNA for noninvasive prenatal testing*, PNAS USA 2014; 111(23):8583-8. La mediana de la longitud del ADN libre circulante encontrado en el plasma de individuos sanos y pacientes con cáncer es de aproximadamente 185-200 pb. Giacona, M.B., *et al.*, *Cell-free DNA in human blood plasma: length measurements in patients with pancreatic cancer and healthy controls*, Pancreas 1998; 17(1):89-97. Las muestras mal conservadas o tratadas químicamente pueden contener ácidos nucleicos degradados química o físicamente. Por ejemplo, los tejidos incluidos en parafina fijados con formol (FFPET) típicamente proporcionan ácidos nucleicos con una longitud promedio de 150 pb.

30 *Purificación*

En algunos modos de realización, el procedimiento divulgado en el presente documento incluye una o más etapas de purificación después de la extensión del cebador por ADN polimerasa o transcriptasa inversa. La purificación retirará moléculas de cebador sin usar y la molécula molde usada para crear el producto de extensión del cebador. En algunos modos de realización, el ácido nucleico molde y todos los fragmentos de ácido nucleico distintos del cebador extendido se retiran por digestión con exonucleasas. En ese modo de realización, el cebador usado en la extensión del cebador puede tener una modificación en el extremo 5' que hace que el cebador y cualquier producto de extensión sean resistentes a la digestión con exonucleasas. Los ejemplos de dicha modificación incluyen enlace fosforotioato. En otros modos de realización, el molde de ARN se puede retirar por tratamiento enzimático que ahorrará ADN, por ejemplo, digestión con ARNasa, incluyendo digestión con ARNasaH. Aún en otros modos de realización, los cebadores y el ADN molde de gran tamaño se separan de los productos de extensión por un procedimiento de exclusión por tamaño, por ejemplo, electroforesis en gel, cromatografía o isotacoforesis.

En algunos modos de realización, la purificación es por unión por afinidad. En variaciones de este modo de realización, la afinidad es para la secuencia diana específica (captura de secuencia). En otros modos de realización, el cebador comprende una marca de afinidad. Se puede usar cualquier marca de afinidad conocida en la técnica, por ejemplo, biotina o un anticuerpo o un antígeno para el que existe un anticuerpo específico. El compañero de afinidad para la marca de afinidad puede estar presente en solución, por ejemplo, en partículas o perlas suspendidas, o unido a un soporte sólido. En el transcurso de la purificación por afinidad, los componentes no unidos de la mezcla de reacción se separan por lavado. En algunos modos de realización, se toman etapas adicionales para retirar el cebador sin usar.

50 *Ligación de secuencias de cebado universal*

En algunos modos de realización, el procedimiento divulgado en el presente documento incluye una etapa de ligación. Por ejemplo, es posible añadir una cola de homopolímero al extremo 3' de un ácido nucleico. En este modo de realización, el homopolímero puede servir como un sitio de unión para el homopolímero de complemento inverso (similar a la cola poli-A con cebador poli-T para ARNm). La ligación añade una o más secuencias del adaptador al producto de extensión del cebador generado en la etapa precedente. La secuencia del adaptador proporciona uno o más sitios de cebado universal (para amplificación o secuenciación) y, opcionalmente, uno o más códigos de barras. El modo exacto de ligación del adaptador es irrelevante siempre que el adaptador se asocie con el producto de extensión del cebador y permita las etapas posteriores descritas a continuación.

En algunos modos de realización descritos anteriormente, el procedimiento implica un cebador específico de diana que incluye una secuencia de cebado universal ("sitio de cebado") y proporciona un producto de extensión del cebador con un sitio de cebado individual. En dichos modos de realización, solo se necesita proporcionar una secuencia de cebado adicional ("sitio de cebado") para permitir la amplificación exponencial. En otros modos de realización, el

cebador específico de diana no incluye un sitio de cebado universal. En dichos modos de realización, es necesario proporcionar dos sitios de cebado para permitir la amplificación exponencial. Los adaptadores con sitios de cebado universal se pueden añadir por cualquier procedimiento de ligación monocatenaria disponible en la técnica.

5 Se puede usar un ejemplo de un procedimiento de ligación monocatenaria en modos de realización donde el cebador de extensión comprende un sitio de ligación universal. En dichos modos de realización, el adaptador que tiene una región bicatenaria y un saliente monocatenario complementario al sitio de ligación universal en el cebador se puede hibridar y ligar como se muestra en la figura 1, etapa 4. La hibridación del saliente en 3' del adaptador al sitio de ligación universal en el extremo 5' del cebador crea una región bicatenaria con una mella en la hebra que contiene el producto de extensión del cebador. Las dos hebras se pueden ligar en la mella por una ADN ligasa u otra enzima, o un reactivo no enzimático que puede catalizar una reacción entre el 5'-fosfato del producto de extensión del cebador y el 3'-OH del adaptador. Al conectar el adaptador, la ligación proporciona un sitio de cebado universal en un extremo del producto de extensión del cebador.

15 Se puede usar otro ejemplo de un procedimiento de ligación monocatenaria para añadir el sitio de cebado universal al extremo opuesto del producto de extensión del cebador (o, en modos de realización donde el cebador de extensión no comprende un sitio de ligación universal, a ambos lados del producto de extensión). Para este modo de realización, uno o ambos extremos del producto de extensión del cebador que se va a ligar no tienen un sitio de ligación universal. Además, en algunos modos de realización, al menos un extremo del producto de extensión del cebador que se va a ligar tiene una secuencia desconocida (por ejemplo, debido a un acontecimiento de terminación aleatorio o una variación de secuencia desconocida). En dicho modo de realización, se emplea un procedimiento de ligación monocatenaria independiente de secuencia. Un procedimiento ejemplar se describe en una publicación de solicitud de EE. UU. n.º 20140193860. Esencialmente, el procedimiento usa una población de adaptadores donde el saliente del extremo 3' monocatenario en lugar de tener un sitio de ligación universal, tiene una secuencia aleatoria, por ejemplo, una secuencia hexámera aleatoria. En algunos modos de realización de ese procedimiento, el adaptador también tiene una estructura de horquilla. Otro ejemplo es un procedimiento permitido por el kit Accel-NGS™ 1S DNA Library Kit (Swift Biosciences, Ann Arbor, Mich.).

30 La etapa de ligación del procedimiento utiliza una ligasa u otra enzima con una actividad similar o un reactivo no enzimático. La ligasa puede ser una ADN o ARN ligasa, por ejemplo, de origen vírico o bacteriano tal como la ligasa T4 o de *E. coli*, o las ligasas termoestables *Afu*, *Taq*, *Tfl* o *Tth*. En algunos modos de realización, se puede usar una enzima alternativa, por ejemplo, topoisomerasa. Además, se puede usar un reactivo no enzimático para formar el enlace fósforo-diéster entre el 5'-fosfato del producto de extensión del cebador y el 3'-OH del adaptador como se describe y se hace referencia en el documento US20140193860.

35 *Extensión del cebador y ligación en extremo romo opcionales*

40 En algunos modos de realización del procedimiento, la primera ligación del adaptador está seguida de una extensión del cebador opcional. El adaptador ligado tiene un extremo 3' libre que se puede extender para crear un ácido nucleico bicatenario. El extremo opuesto al adaptador será adecuado a continuación para la ligación en extremo romo de otro adaptador. Evitando la necesidad de un procedimiento de ligación monocatenaria, este extremo bicatenario de la molécula se puede ligar a un adaptador bicatenario por cualquier ligasa u otro medio enzimático o no enzimático. La secuencia del adaptador bicatenario proporciona uno o más sitios de cebado universal (para amplificación o secuenciación) y, opcionalmente, uno o más códigos de barras.

45 *Purificación*

50 En algunos modos de realización, el procedimiento divulgado en el presente documento incluye una o más etapas de purificación después de la etapa de ligación. La purificación retirará moléculas adaptadoras sin usar. Los adaptadores y productos ligados de gran tamaño se separan de los productos de extensión por un procedimiento de exclusión por tamaño, por ejemplo, electroforesis en gel, cromatografía o isotacoforesis.

55 En algunos modos de realización, la purificación es por unión por afinidad. En variaciones de este modo de realización, la afinidad es para la secuencia diana específica (captura de secuencia). En otros modos de realización, el adaptador comprende una marca de afinidad. Se puede usar cualquier marca de afinidad conocida en la técnica, por ejemplo, biotina o un anticuerpo o un antígeno para el que existe un anticuerpo específico. El compañero de afinidad para la marca de afinidad puede estar presente en solución, por ejemplo, en partículas o perlas suspendidas, o unido a un soporte sólido. En el transcurso de la purificación por afinidad, los componentes no unidos de la mezcla de reacción se separan por lavado. En algunos modos de realización, se toman etapas adicionales para retirar el adaptador sin usar.

60 *Amplificación*

65 En algunos modos de realización, el procedimiento divulgado en el presente documento comprende una etapa de amplificación. Esta etapa puede implicar una amplificación lineal o exponencial, por ejemplo, PCR. Los cebadores para la amplificación pueden incluir cualquier secuencia que esté presente dentro del ácido nucleico que se está

amplificando y pueden soportar la síntesis de una o ambas hebras. La amplificación puede ser isotérmica o implicar termociclado.

En algunos modos de realización, la amplificación es exponencial e implica PCR. Se desea reducir el sesgo de amplificación por PCR. Si se usan uno o más cebadores específicos de gen, para reducir el sesgo, el procedimiento implica un número limitado de ciclos de amplificación, por ejemplo, aproximadamente 10 ciclos o menos. En otras variaciones de estos modos de realización, se usan cebadores universales para sintetizar ambas hebras. Las secuencias de cebador universal pueden ser parte del cebador de extensión original de uno o ambos adaptadores ligados. Se pueden usar uno o dos cebadores universales. El cebador de extensión y uno o ambos adaptadores descritos anteriormente se pueden genomanipular para tener el mismo sitio de unión de cebador. En ese modo de realización, se puede usar un cebador universal individual para sintetizar ambas hebras. En otros modos de realización, el cebador de extensión (o adaptador) en un lado y el adaptador en el otro lado de la molécula que se va a amplificar contienen diferentes sitios de unión del cebador universal. Un cebador universal se puede emparejar con otro cebador universal (de la misma o diferente secuencia). En otros modos de realización, el cebador universal se puede emparejar con un cebador específico de gen. Debido a que la PCR con cebadores universales tiene un sesgo de secuencia reducido, no es necesario limitar el número de ciclos de amplificación al mismo grado que en la PCR con cebadores específicos de gen. El número de ciclos de amplificación donde se usan cebadores universales puede ser bajo, pero también puede ser tan alto como aproximadamente 20, 30 o más ciclos.

20 *Códigos de barras*

El procedimiento divulgado en el presente documento incluye el uso de códigos de barras moleculares. Los códigos de barras típicamente consisten en de 4 a 36 nucleótidos. En algunos modos de realización, los códigos de barras se diseñan para tener una temperatura de fusión dentro de 10 °C o menos entre sí. Los códigos de barras se pueden diseñar para formar un conjunto de hibridación mínimamente cruzada, es decir, una combinación de secuencias que, en las condiciones de reacción deseadas, forman la menor cantidad posible de híbridos estables entre sí. El diseño, colocación y uso de códigos de barras para identificación y recuento de secuencias son conocidos en la técnica. Véanse, por ejemplo, las patentes de EE. UU. n.ºs 7.393.665, 8.168.385, 8.481.292, 8.685.678, y 8.722.368.

Se pueden usar códigos de barras para identificar cada molécula de ácido nucleico en la muestra y su progenie (es decir, un conjunto de moléculas de ácido nucleico que se producen usando la molécula de ácido nucleico original). Dichos códigos de barras son "ID únicas" (UID).

Los códigos de barras también se pueden usar para identificar una muestra de la que se deriva la molécula de ácido nucleico que se está analizando. Dichos códigos de barras son "ID de muestra múltiple" ("MID"). Todas las moléculas derivadas de la misma muestra comparten las mismas MID.

Los códigos de barras comprenden una secuencia única de nucleótidos característica de cada código de barras. En algunos modos de realización, las secuencias de códigos de barras están prediseñadas. En otros modos de realización, las secuencias de códigos de barras son aleatorias. Todos o algunos nucleótidos dentro del código de barras pueden ser aleatorios. Una secuencia aleatoria y una base nucleotídica aleatoria dentro de una secuencia conocida se denominan "secuencia redundante" y "base redundante" respectivamente. En algunos modos de realización, una molécula comprende dos o más códigos de barras: uno para identificación molecular (UID) y otro para identificación de muestras (MID). A veces, la UID o la MID comprenden cada una varios códigos de barras que, cuando se toman conjuntamente, permiten la identificación de la molécula o la muestra.

En algunos modos de realización, el número de UID en la reacción puede ser superior al número de moléculas que se van a marcar. En algunos modos de realización, se usan uno o más códigos de barras para agrupar secuencias. Por ejemplo, en algunos modos de realización, se usan uno o más UID para agrupar secuencias, en las que las secuencias de cada agrupación contienen la misma UID, es decir, son amplicones derivados de una molécula diana individual. En algunos modos de realización, se usan UID para alinear secuencias. En otros modos de realización, se usa la región específica de diana para alinear secuencias. En algunos modos de realización del presente procedimiento divulgado en el presente documento, las UID se introducen en el acontecimiento de extensión del cebador inicial mientras que los códigos de barras de muestra (MID) se introducen en los adaptadores ligados.

55 *Secuenciación*

Después de que se haya realizado la ligación, es decir, después de la etapa 4 o la etapa opcional 5 (figura 1), se pueden secuenciar los productos de ácido nucleico. La secuenciación se puede realizar por cualquier procedimiento conocido en la técnica. Es especialmente ventajosa la secuenciación de molécula individual de alto rendimiento. Los ejemplos de dichas tecnologías incluyen la plataforma GS FLX de 454 Life Sciences (454 Life Sciences, Branford, Conn.), la plataforma HiSeq de Illumina (Illumina, San Diego, Cal.), la plataforma Ion Torrent (Life Technologies, Grand Island, NY), la plataforma Pacific BioSciences que utiliza SMRT (Pacific Biosciences, Menlo Park, Cal.) y cualquier otra tecnología de secuenciación de molécula individual actualmente existente o futura que implique o no la secuenciación por síntesis. En variaciones de estos modos de realización, la secuenciación utiliza un sitio de cebadores universales presente en una o ambas secuencias del adaptador o en una o ambas secuencias de cebador.

Aún en otras variaciones de estos modos de realización, se usa un cebador específico de gen para la secuenciación. Sin embargo, cabe señalar que los cebadores universales se asocian con un sesgo de secuenciación reducido en comparación con los cebadores específicos de gen.

5 En algunos modos de realización, la etapa de secuenciación implica alineación de secuencias. En algunos modos de realización, se usa la alineación para determinar una secuencia consenso a partir de una pluralidad de secuencias, por ejemplo, una pluralidad que tiene la misma ID molecular única (UID). En algunos modos de realización, se usa la alineación para identificar variaciones de secuencia, tales como variaciones mononucleotídicas (SNV). En algunos modos de realización, se determina una secuencia consenso a partir de una pluralidad de secuencias que tienen todas una UID idéntica. En otros modos de realización, se usa la UID para eliminar artefactos, es decir, variaciones existentes en la progenie de una molécula individual (caracterizada por una UID particular). Dichos artefactos resultantes de errores de PCR o errores de secuenciación se pueden eliminar usando UID.

15 En algunos modos de realización, el número de cada secuencia en la muestra se puede cuantificar cuantificando números relativos de secuencias con cada UID entre la población que tiene la misma ID de muestra múltiple (MID). Cada UID representa una molécula individual en la muestra original y contando diferentes UID asociadas con cada variante de secuencia puede determinar la fracción de cada variante de secuencia en la muestra original, donde todas las moléculas comparten la misma MID. Un experto en la técnica podrá determinar el número de lecturas de secuencia necesarias para determinar una secuencia consenso. En algunos modos de realización, se lee el número pertinente por UID ("profundidad de secuencia") necesario para un resultado cuantitativo exacto. En algunos modos de realización, la profundidad deseada es de 5-50 lecturas por UID.

Muestra

25 Una muestra usada en el procedimiento divulgado en el presente documento comprende cualquier muestra individual (por ejemplo, humano, paciente) o ambiental que contiene ácidos nucleicos. Los polinucleótidos se pueden extraer de la muestra o la muestra se puede someter directamente a los procedimientos de la invención. La muestra de partida también se puede extraer o aislar de ácidos nucleicos, ADN o ARN. La muestra puede constituir cualquier tejido o líquido obtenido de un organismo. Por ejemplo, la muestra puede ser una biopsia tumoral o una muestra de sangre o plasma. En algunos modos de realización, la muestra es una muestra incluida en parafina fijada con formol (FFPE). La muestra puede comprender ácidos nucleicos de una o más fuentes, por ejemplo, uno o más pacientes. En algunos modos de realización, los tejidos se pueden infectar con un patógeno y por tanto contener ácidos nucleicos del huésped y del patógeno.

35 Los procedimientos de extracción de ADN son bien conocidos en la técnica. Véase J. Sambrook *et al*, "Molecular Cloning: A Laboratory Manual," 1989, 2.^a ed., Cold Spring Harbor Laboratory Press: New York, N.Y.). Una variedad de kits están disponibles comercialmente para extraer ácidos nucleicos (ADN o ARN) de muestras biológicas (por ejemplo, BD Biosciences Clontech (Palo Alto, Cal.), Epicentre Technologies (Madison, Wisc.); Gentra Systems, Inc. (Minneapolis, Minn.); y Qiagen, Inc. (Valencia, Cal.), Ambion, Inc. (Austin, Tex.); BioRad Laboratories (Hercules, Cal.); y más.

45 En algunos modos de realización, la muestra de partida usada en el procedimiento divulgado en el presente documento es un banco, por ejemplo, una genoteca genómica o una genoteca de expresión que comprende una pluralidad de polinucleótidos. En otros modos de realización, se crea un banco por el procedimiento de la invención. Siendo el material de partida una muestra biológica, el procedimiento crea un banco de amplificación o una colección de amplicones que representan variedad o secuencias. Un banco se puede almacenar y usar múltiples veces para amplificación o secuenciación adicional de los ácidos nucleicos en el banco.

EJEMPLOS

50 **Ejemplo 1 (hipotético) Enriquecimiento de dianas con cebador específico de gen y amplificación lineal**

Se aíslan ácidos nucleicos de una muestra de plasma sanguíneo humano usando el kit DNeasy Blood & Tissue Kit (Qiagen, Valencia, Cal.). Se añade un cebador específico de gen. El cebador se diseña con una porción específica de gen que se hibrida al exón 19 del gen EGFR humano. El cebador también tiene una secuencia de identificación única (UID) hexámera y una secuencia de ligación universal. El cebador se modifica en el extremo 5' para evitar la digestión con exonucleasas. Se deja que el cebador se hibride en tampón de amplificación isotérmica (New England Biolabs, Ipswich, Mass., "NEB") a 60 °C durante 20 minutos y se añade la Bst polimerasa 2.0 (NEB), una ADN polimerasa no termoestable y se incuba durante 20 segundos a 65 °C. La reacción se termina termoinactivando la polimerasa a 95 °C durante 3 minutos. Las hebras molde de ácido nucleico se digieren con una combinación de la exonucleasa RecJF (NEB) específica de ADNmc en 5' y la exonucleasa lambda específica de ADNbc en 5'. Los cebadores no extendidos se retiran usando purificación con perlas Ampure (Beckman Coulter, Brea, Cal.).

65 Las hebras individuales que resultan de la extensión del cebador se purifican y se añaden a una reacción de ligación. Se añaden dos tipos de adaptadores de ligación. El adaptador 5' se diseña para contener el sitio de ligación universal, un sitio de cebador universal para amplificación y un sitio de cebador universal para secuenciación. El adaptador en

5' también contiene una ID de muestra múltiple (MID). El adaptador en 3' se diseña para contener un sitio de cebador universal para amplificación y un sitio de cebador universal para secuenciación. La ligación monocatenaria se realiza usando reactivos del kit Accel-NGS™ 1S DNA Library Kit (Swift Biosciences, Ann Arbor, Mich.).

- 5 Los adaptadores no ligados se separan de los productos de ligación por medio de purificación Ampure como se describe anteriormente.

10 Para la amplificación lineal, los productos de ligación se ponen en contacto con una mezcla de reacción que comprende un cebador universal individual correspondiente al sitio de unión de cebador en el adaptador 3' terminal. Después de la amplificación, se transfiere una muestra de la mezcla de reacción a la reacción de secuenciación que comprende un cebador de secuenciación universal.

Ejemplo 2 (hipotético) Enriquecimiento de dianas con un cebador redundante y amplificación exponencial

15 Se aíslan ácidos nucleicos de una muestra de plasma sanguíneo humano usando el kit DNeasy Blood & Tissue Kit (Qiagen, Valencia, Cal.). Se añade un cebador que contiene una secuencia redundante. El cebador se diseña con una secuencia aleatoria de seis nucleótidos, una secuencia de identificación única (UID) hexámera y una secuencia de ligación universal. El cebador se modifica en el extremo 5' para presentar digestión con exonucleasas. Se deja que el cebador se hibride en el tampón de amplificación isotérmica (NEB) a 60 °C durante 20 minutos y se añade Bst Polimerasa 2.0 (NEB), una ADN polimerasa no termoestable y se incuba durante 20 segundos a 65 °C. La reacción se termina termoinactivando la polimerasa a 95 °C durante 3 minutos. Las hebras molde de ácido nucleico se digieren con una combinación de la exonucleasa RecJF (NEB) específica de ADNmc en 5' y la exonucleasa lambda específica de ADNbc en 5'. Los cebadores no extendidos se retiran usando purificación con perlas Ampure como se describe en el ejemplo 1.

25 Las hebras individuales que resultan de la extensión del cebador se purifican y se añaden a una reacción de ligación. Se añaden dos tipos de adaptadores de ligación. El adaptador 5' se diseña para contener el sitio de ligación universal, un sitio de cebador universal para amplificación y un sitio de cebador universal para secuenciación. El adaptador en 5' también contiene una ID de muestra múltiple (MID). El adaptador en 3' se diseña para contener un sitio de cebador universal para amplificación y un sitio de cebador universal para secuenciación. La ligación monocatenaria se realiza esencialmente como se describe en la publicación US20140193860.

Los adaptadores no ligados se retiran usando purificación con perlas Ampure como se describe en el ejemplo 1.

35 Para la amplificación exponencial, los productos de ligación se ponen en contacto con una mezcla de reacción de PCR que comprende un par de cebadores de amplificación universales. Después de la amplificación, se transfiere una muestra de la mezcla de reacción a la reacción de secuenciación que comprende un cebador de secuenciación universal.

REIVINDICACIONES

1. Un procedimiento de amplificación de una secuencia diana que comprende las etapas de:
- 5 a) poner en contacto el ácido nucleico diana con un cebador y una polimerasa, en el que el cebador comprende un sitio de unión a diana y una marca de identificación molecular única (UID);
- b) llevar a cabo una reacción de extensión por la polimerasa y una terminación para crear un producto de extensión del cebador monocatenario;
- 10 c) ligar adaptadores a cada extremo del producto de extensión del cebador monocatenario para crear un producto de ligación, en el que los adaptadores comprenden al menos un sitio de cebado universal;
- d) amplificar el producto de ligación en una reacción de amplificación utilizando al menos un cebador que se une al al menos un sitio de cebado universal para crear la secuencia diana amplificada
- 15 en el que la terminación se efectúa por un procedimiento seleccionado de la lista que consiste en cambio de temperatura, adición de un inhibidor enzimático específico, adición de un quelante, incorporación de bases que contienen uridina seguido de tratamiento con uracil-N-ADN glucosilasa.
- 20 2. El procedimiento de la reivindicación 1, en el que el cebador y al menos uno de los adaptadores comprenden sitios de ligación universal mutuamente compatibles.
3. El procedimiento de la reivindicación 1-2, en el que el sitio de unión a diana es una secuencia específica de diana prediseñada.
- 25 4. El procedimiento de la reivindicación 1-2, en el que el sitio de unión a diana es una secuencia aleatoria.
5. El procedimiento de la reivindicación 1-4, en el que al menos un adaptador comprende un código de barras.
- 30 6. El procedimiento de la reivindicación 5, en el que el código de barras es una ID de muestra múltiple (MID).
7. El procedimiento de la reivindicación 1-6, en el que la amplificación es amplificación lineal.
- 35 8. El procedimiento de la reivindicación 1-6, en el que la amplificación es amplificación exponencial.
9. El procedimiento de la reivindicación 1-8, que comprende además una etapa de purificación después de al menos una de las etapas b) y c).

Fig. 1

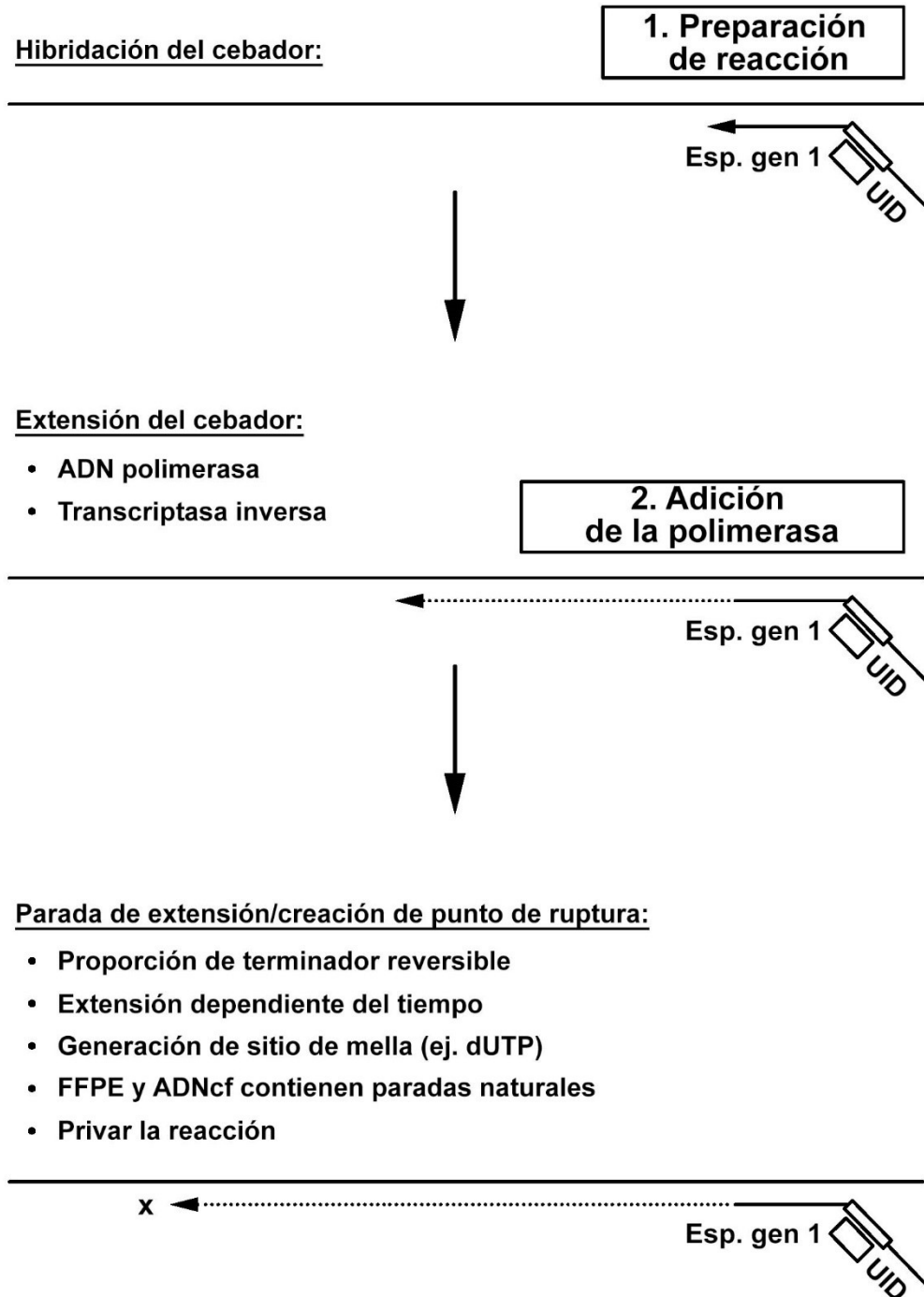
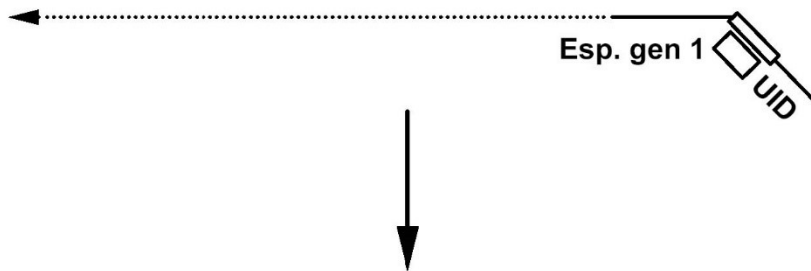


Fig. 1 continuación

Retirada de ADN no extendido:

- 5' exo si el cebador original tiene bloqueo en 5'
- ADNmc endo si el cebador está protegido ADNmc se mantendrá (no mostrado)

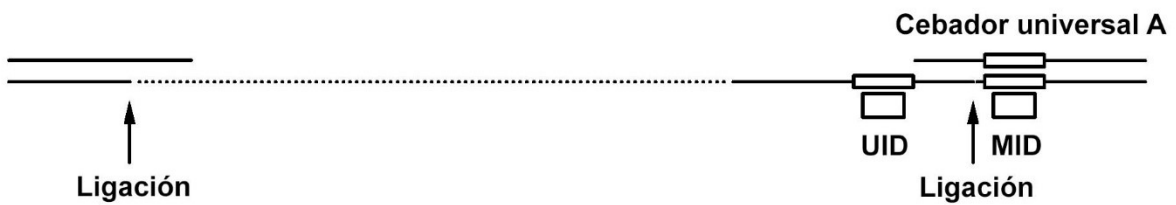
3. Adición de exonucleasas



Ligación de ADNmc en 3':

- Kit Swift Bio 1S
- Horquilla de Penn State

4. Adición de ligasa y adaptador



PCR ML:

5. Reactivos de PCR y termociclo

