

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 645 423**

51 Int. Cl.:

C12N 15/10 (2006.01)

C12Q 1/68 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **20.03.2015 PCT/EP2015/055961**

87 Fecha y número de publicación internacional: **24.09.2015 WO15140307**

96 Fecha de presentación y número de la solicitud europea: **20.03.2015 E 15711509 (8)**

97 Fecha y número de publicación de la concesión europea: **02.08.2017 EP 3119886**

54 Título: **Método de análisis de ARN de conservación del número de copias**

30 Prioridad:

21.03.2014 EP 14161135

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

05.12.2017

73 Titular/es:

**LEXOGEN GMBH (100.0%)
Campus-Vienna-Biocenter 5 Helmut Qualtinger-
Gasse 6
1030 Vienna, AT**

72 Inventor/es:

MOLL, PAMELA

74 Agente/Representante:

ARIAS SANZ, Juan

ES 2 645 423 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

Método de análisis de ARN de conservación del número de copias

Campo de la invención

5 La presente invención se refiere a un método de análisis de ARN, en particular a ensayos de la estimación del tipo y la cantidad de transcritos.

Antecedentes

10 La expresión génica es el proceso mediante el cual se usa la información de un gen en la síntesis de un producto génico funcional. Estos productos son ARN funcional de los cuales una clase importante principal consiste en los ARN mensajeros, ARNm, que codifican para proteínas, que se traducen en el proceso para dar toda clase de proteínas como enzimas, moléculas transportadoras, y otros. El conocimiento del contenido de ARNm y su fase de procesamiento en células y tejidos es importante para la comprensión de la génesis celular, el desarrollo de enfermedades, la respuesta a fármacos de organismos y otros procesos biológicos.

15 Los procesos celulares biológicos se ven afectados por numerosos parámetros internos y externos. En el presente documento, todo el ARN y en particular la reserva de ARNm (transcriptoma) desempeña un papel fundamental. Las células de mamífero típicas contienen entre 10 y 30 pg de ARN total que corresponde a $3,6 \cdot 10^5$ moléculas de ARNm en promedio. Las bases de datos actuales del genoma humano contienen 20769 anotaciones de genes codificantes, 48461 predicciones de genes de Genescan. Aunque los números para las anotaciones de genes y predicciones de genes son bastante estables, el número de transcritos (ahora 195565 transcritos) que se anotan aumenta continuamente debido a mejoras en el análisis de ARN [publicación de Ensembl 73, sept. de 2013]. El foco principal de muchas investigaciones es la cuantificación de ARN que codifica para proteínas, el ARNm o transcritos. Genes individuales pueden expresar numerosos transcritos diferentes, denominados variantes de corte y empalme, que se caracterizan a través de diferencias en su región de exón, y/o diferencias en los sitios de inicio y terminación de las regiones no traducidas que son importantes para procesos reguladores.

20 Se han desarrollado diferentes métodos para medir los niveles de expresión de o bien ARNm o bien genes con diferentes grados de precisión.

25 Las etiquetas de secuencia expresada, EST, son subsecuencias cortas de ADNc y resultan de la secuenciación de una sola vez de un ADNc clonado. Se usaron en el pasado para identificar transcritos de genes. Están disponibles millones de EST en bases de datos públicas y proporcionan información sobre las condiciones en las que se expresan los genes correspondientes. Las EST permiten el diseño de sondas para microalineamientos de ADN para medir la expresión génica.

30 Los métodos clásicos para mediciones de expresión génica tales como ensayos de hibridación de microalineamientos, o métodos más recientes tales como secuenciación de ARNm mediante secuenciación masiva en paralelo o secuenciación de nueva generación, NGS, están limitados a través de la inexactitud inherente de los métodos que, actualmente, sólo pueden compensarse en cierta medida a través de más mediciones, como secuenciación de mayor profundidad, lo que aumenta inevitablemente los costes en tal medida que no pueden llevarse a cabo análisis a grandes escalas de procesamiento de muestras. Sin embargo, la exactitud en las mediciones y también los costes son los requisitos más importantes en la investigación farmacéutica y grandes estudios a escala clínica. Los microalineamientos sólo pueden detectar genes a nivel de exón o de secuencia para lo que se han diseñado sondas de secuencias predeterminadas antes de los experimentos. El número limitado de tales sondas de hibridación y la hibridación errónea a menudo conducían a resultados ambiguos para experimentos de expresión génica de alta resolución. Los microalineamientos son limitados por diseño porque sólo pueden cubrir un cierto número de diferentes 3'UTR (región no traducida en 3') y no pueden identificar nuevas 3'UTR.

35 Al final de 1996 empezaron a surgir nuevas tecnologías de secuenciación de alto rendimiento [documento WO 98/44151] y pasaron a conocerse como secuenciación de nueva generación, NGS, en contraposición con el método de dideoxi habitual hasta ese momento por Sanger. El desarrollo de nuevas tecnologías de secuenciación hizo que fuese posible intentar la secuenciación de transcriptomas completos. NGS usa celdas de flujo miniaturizadas y paralelizadas para secuenciar millones de lecturas de extremos apareados, individuales o cortas, de entre 50 y 400 bases de longitud. Se secuencian moldes de ADN amplificadas de manera clonal espacialmente separadas mediante síntesis de tal modo que se produce descodificación mientras se añaden nucleótidos individuales a las hebras complementarias. Se usan barrido óptico (sistemas Illumina de Illumina, Inc., EE.UU.; sistemas SOLiD de Life Technologies, EE.UU.; Roche 454 de 454 Life Sciences, Roche Diagnostics Corp., EE.UU.) y la detección de minúsculos cambios de pH a través de transistores de efecto de campo de microchip en microalineamiento (Ion Torrent de Life Technologies, EE.UU.) en diferentes plataformas microfluídicas. Los millones de lecturas cortas deben alinearse con secuencias o bien conocidas o bien ensambladas *de novo*. Sin embargo, para la investigación de ARN, la situación es más compleja porque las secuencias de transcritos de genes individuales se solapan en gran medida. Las anotaciones de variantes de transcrito halladas previamente proporcionan marcos para guiar el posterior ensamblaje del transcrito basándose en el descubrimiento de exones individuales, uniones exón-exón y probabilidades de cobertura. Sólo el ensamblaje correcto del transcrito permite asignar lecturas a sus moléculas de

ARN parentales y, además, el cálculo de los números de copias respectivos.

Independientemente de la tecnología NGS, la determinación simultánea de información de secuencia y frecuencia es un problema importante en la investigación de mezclas de secuencias complejas. Dado que sólo su secuencia determina la naturaleza de la molécula, parece ser inevitable secuenciar repetitivamente moléculas idénticas de manera proporcional a su abundancia para contar sus números de copias respectivos. Un intervalo dinámico de seis órdenes requiere una secuenciación repetitiva a través de millones de moléculas idénticas altamente abundantes antes de llegar a valores estadísticamente sólidos para moléculas poco abundantes. Tales enfoques requieren mucho tiempo y recursos durante la secuenciación y el posterior análisis de datos. La profundidad de lectura requerida depende considerablemente de la complejidad de la muestra [Hopper, 2010; Wendl, 2009]. Después de todo, un reto importante es el enredo de alinear lecturas solapantes con múltiples anotaciones de transcritos solapantes dentro de genes individuales. Los esfuerzos y los costes en la profundidad de lectura y computación son enormes. Por tanto, se han desarrollado diferentes enfoques que eliminan la necesidad de alinear lecturas solapantes simplemente produciendo una lectura por molécula de ARNm. La agrupación y el recuento de tales lecturas simplifica las mediciones de expresión génica y de ARNm [documento WO02/059357].

La poliadenilación de pre-ARNm es una etapa importante de la expresión y regulación génica de eucariotas. Muchos genes producen ARNm con sitios de poliadenilación alternativos, APA y distintas 3'UTR que pueden regularse de diferente manera o que pueden codificar también para diferentes isoformas de proteína. Por tanto, para combinar la simplicidad de determinación de valores de expresión génica generando simplemente una lectura por ARNm con la identificación precisa de sitios de poliadenilación, se desarrollaron métodos para seleccionar como diana exclusivamente esos APA.

Uno de tales métodos identifica sitios de poliA de manera específica de hebra y de genoma completo [Wilkening, 2013]. En este caso, se preparan bibliotecas para secuenciación NGS a través de: fragmentación térmica de la muestra de ARN, inmovilización inversa en fase sólida, SPRI, purificación para detener la fragmentación adicional a través de cambio de tampón, transcripción inversa después de cebado con poliT(V)-cebador-adaptador biotinilado y anclado, purificación SPRI para retirar todos los fragmentos que no contienen poliA y para cambiar la disolución, el tratamiento con ARNasa H para degradar el ARN y usar los fragmentos de ARN más pequeños como secuencias de inicio aleatorias para la síntesis de segunda hebra con ADN polimerasa I que genera la doble hebra más larga posible porque todos los demás sitios de cebado extendidos interiores se desplazarán a través de desplazamiento de hebra, purificación SPRI, purificación por afinidad con estreptavidina y unión que permite los cambios de disolución después de cada una de las siguientes 3 etapas, reparación enzimática de extremos, prolongación de la cola con dA individuales, ligamiento de otro adaptador, seguido por una PCR de enriquecimiento y purificación SPRI.

Las bibliotecas NGS resultantes contienen simplemente una lectura por molécula de ARNm, aunque una lectura por ARNm marca el máximo teórico. En términos prácticos, dado que cada una de las muchas etapas de reacción de la generación de bibliotecas tiene una eficacia inferior al 100%, el resultado es una representación distorsionada, y en la realización a la que se aspira distorsionada proporcionalmente, de las abundancias de transcritos. Es importante que el número de lecturas por especie de transcrito sea proporcional a su número de copias y no a su longitud o cualquier otro sesgo específico de secuencia. El método que requiere mano de obra, productos químicos y productos de consumo es ventajoso para mediciones de expresión génica porque permite cuantificar abundancias de ARN a través de un simple recuento de lecturas porque sólo se produce una lectura a partir de cada transcrito. El método continúa con un protocolo de NGS particular que lee de manera silenciosa a través del tramo de poliT del cebador-adaptador antes de que se inicie la secuenciación real. Esta parte se denomina método de relleno de 3'T. Además, pueden detectarse los niveles de expresión de isoformas de sitio de poliA y cuantificarse con una resolución de secuencia de nucleótidos individuales, o después de fusionar sitios de poliA de estrecha proximidad a agrupaciones respectivas. Además de una mejor calidad en la generación de lecturas, la principal mejora en el protocolo fue la introducción de dicho relleno de 3'T que permitió la secuenciación desde el mismo extremo de los transcritos.

Se habían desarrollado antes otros métodos de enriquecimiento de sitios de poliA pero sin el relleno de 3'T mencionado anteriormente. Dado que faltan referencias internas de variantes de transcritos, es difícil evaluar las diferentes calidades de los métodos. Un método más sencillo es el análisis multiplexado de secuencias ligadas a poliA, MAPS [Fox-Walsh, 2011]. En el presente documento, se usa una secuencia de adaptador que contiene oligo-dT(NV) biotinilado para cebar la síntesis de ADNc. Tras la selección en fase sólida, se inicia la síntesis de segunda hebra usando un cebador al azar que se liga a otra secuencia de adaptador. Finalmente, se libera la biblioteca de perlas recubiertas con estreptavidina y se amplifica usando un cebador con código de barras junto con un cebador común. Este método tiene asimismo la capacidad para detectar de manera robusta la expresión génica. Aunque, la dirección de lectura se dirigía originariamente hacia el extremo 3' del ARNm, y sólo una selección de tamaño muy estrecho de la biblioteca permitiría la lectura en el sitio de poliA, el cambio de las secuencias de adaptador (cebador) y la combinación con el método de relleno de 3'T descrito anteriormente también permite la detección precisa de los sitios de poliA con todas las lecturas.

El método tiene varias deficiencias. Tiene como objetivo sintetizar ADNc de longitud completa, protege los extremos del ADNc con didesoxirribonucleósido trifosfato, ddNTP, antes de unir el ADNc a superficies de perlas con estreptavidina, purificar el ADNc mediante estos medios, cebar y extender segundas hebras con ADN polimerasa de

5 Taq. La ADN polimerasa de Taq degrada cualquier hebra que se encuentre en el sentido de 3' mediante una actividad exonucleasa 5'->3' y se ha elegido para garantizar que sólo se produce una segunda hebra por ADNc, la que se ha cebado más lejos del sitio de poliA, antes de purificar el producto bicatenario a través del método de unión por afinidad mencionado. Dado el ADNc largo, las bibliotecas NGS son largas como tendencia lo que conduciría a sesgos de longitud en la posterior generación de agrupaciones de NGS. Aunque la síntesis de segunda hebra se produce en las superficies de las perlas, está dificultada en particular en la región de la superficie de contacto hacia la secuencia de la primera secuencia de cebador biotinilada. Las múltiples etapas de purificación que se ven asistidas por reacciones confinadas a la superficie introducen una serie de sesgos de longitud y secuencia en la generación de lecturas auténticas de sitio de poliA.

10 Otro método basado en secuenciación de profundidad es la secuenciación cuantitativa de sitios de poliA, secuenciación PAS [Shepard, 2011]. Este método se inicia con una etapa de fragmentación para generar fragmentos de ARN del intervalo de tamaño deseado. De nuevo, la primera secuencia de adaptador forma parte del cebador de oligo-dT(NV) anclado. Este método se aprovecha de la actividad transferasa terminal de transcriptasas inversas. Tras alcanzar el extremo 5' del fragmento de ARNm, la transcriptasa inversa de VLMM-V añade unas cuantas desoxicitodinas sin molde a los extremos 3' del ADNc. Esos extremos hibridan con el segundo adaptador que contiene una secuencia de triple G. La transcriptasa inversa continúa cambiando el molde y sintetizando una copia del fragmento de ARNm que se extiende ahora por ambas secuencias de adaptador.

20 Un inconveniente importante de este método muy sencillo es su ineficacia de sólo el 1 - 10%, sesgo e inexactitud del cambio de molde. La baja eficacia dará como resultado pérdidas de transcritos poco abundantes. El cambio de molde no se acopla exclusivamente al cebador de cambio de molde y pueden generarse transcritos de fusión artificiales cambiando a diferentes moldes de ARN. Además, el cebador de cambio de molde ha de proporcionarse en un gran exceso, haciendo que sea esencial una etapa de purificación antes de la posterior amplificación de la biblioteca.

25 Derti *et al.* han descrito otro método de poliA-seq. El protocolo emplea síntesis de primera hebra con cebadores de poliT anclados que contienen la primera secuencia de adaptador, tratamiento con ARNasa H para digerir antes el ARN, cebado con un cebador al azar que contiene la segunda secuencia de adaptador, y extensión de Klenow para la síntesis de segunda hebra. Aunque el fragmento de ADN polimerasa I de Klenow carece de actividad exonucleasa 5'->3' contiene actividad de desplazamiento de hebra persistente. Por tanto, cada ADNc de primera hebra puede generar varias segundas hebras cebadas al azar. La correlación de recuento de lecturas y la abundancia de ARNm biyectiva inequívoca no se garantiza.

30 El documento US 6.406.891 B1 se refiere a un método para generar un ADNc de longitud completa con un método que comprende ciclar hacia atrás y hacia delante entre una enzima RT procesiva y una RT termoestable durante la síntesis de primera hebra.

35 El documento EP 1371726 A1 se refiere a un método de síntesis de primera y segunda hebra. Para la síntesis de primera hebra, se usan cebadores inmovilizados en perlas y para la síntesis de segunda hebra, hexámeros al azar. La síntesis de segunda hebra es con una mezcla de Klenow, que contiene actividad de desplazamiento de hebra.

Costa *et al.* [2010] se refiere a estudios del transcriptoma usando ARN-seq.

Mainul Hoque *et al.* [2012] se refiere al análisis de escisión y poliadenilación alternativas mediante extracción de la región en 3' y secuenciación de profundidad.

40 Bodescot *et al.* (1994) dan a conocer síntesis de ADNc de segunda hebra usando ADN polimerasa de T7. Para el recuento de la expresión génica, existe la necesidad de métodos fiables, eficientes, sencillos y rentables para producir amplicones de bibliotecas NGS que presenten una correlación biyectiva entre la de abundancia ARNm y el recuento de lecturas.

Sumario de la invención

45 La presente invención proporciona un método para generar un producto de ácido nucleico a partir de una molécula de ARN molde, que comprende después de haber obtenido un ARN molde

a) aparear un primer cebador oligonucleotídico en una región de ácido nucleico preseleccionada del ARN molde,

b) elongar el primer cebador oligonucleotídico de manera específica al molde obteniendo de ese modo una primera hebra elongada,

50 c) retirar el ARN molde,

d) aparear más de un cebador oligonucleotídico adicionales con la primera hebra elongada,

e) elongar los más de un cebador oligonucleotídico adicionales de manera específica al molde sin desplazamiento de los cebadores apareados con la primera hebra elongada usando una polimerasa que carece de actividad de desplazamiento de hebra, preferiblemente una ADN polimerasa de T7, T4 o Q5, y/o usando cebadores que tienen

resistencia al desplazamiento de hebra por una polimerasa, preferiblemente cebadores que tienen nucleótidos con modificaciones de LNA o 2'-fluor, generando de ese modo productos de elongación adicionales, en el que se realizan las etapas a) a e) en un volumen de fluido creciente consecutivamente,

- 5 f) aislar y/o amplificar un producto de elongación de dicho producto de elongación adicional que comprende una parte de ácido nucleico que se elonga complementaria al primer cebador oligonucleotídico.

Se definen realizaciones preferidas del método de la invención en las reivindicaciones dependientes 2 a 17.

- 10 La invención también se refiere al uso de un kit para realizar un método según una cualquiera de las reivindicaciones 1 a 17, en el que el kit comprende una transcriptasa inversa, dNTP, cofactores o sales de iones metálicos requeridos por una polimerasa, preferiblemente Mg^{2+} , un cebador, una ADN polimerasa sin actividad de desplazamiento de hebra, tal como ADN polimerasa de T7, Q5 o T4, y cebadores oligonucleotídicos al azar.

La siguiente descripción detallada reza sobre todos los aspectos y realizaciones de la presente invención. Las descripciones de métodos también rezan sobre el kit, que puede comprender partes adecuadas para realizar dicho método; los componentes del kit también pueden rezar sobre el método, que puede implementar o usar tales componentes según su función.

15 Descripción detallada de la invención

- La invención se refiere a un método sencillo y rentable para proporcionar (simplemente) un producto de amplificación por molécula de ARN, que se logra mediante la especificidad de extremo 3' de las reacciones de cebado si se usa, por ejemplo, un cebador que contiene oligo-dT o la especificidad de gen/transcrito si se selecciona como diana una secuencia específica de transcrito o gen durante la síntesis de primera hebra y el posterior aislamiento, selección o amplificación de tales productos teniendo como objetivo el producto de elongación que comprende una parte de ácido nucleico que se elonga complementaria (o en complementariedad) al primer cebador oligonucleotídico (por ejemplo, seleccionando una etiqueta de secuencia o secuencia de ligador o seleccionando otra secuencia del cebador tal como una secuencia que es complementaria al primer cebador). Impidiendo el desplazamiento de los cebadores apareados con la primera hebra elongada (o destruyendo las hebras desplazadas), sólo se obtiene un producto, es decir que se extiende desde el cebador de la etapa d) que se une más cerca de la región preseleccionada del molde, que cumple el criterio de selección, aislamiento, amplificación o generalmente procesamiento de la etapa f). Por tanto, la concentración o el número de copias de cada especie de ARN molde (que tiene la secuencia preseleccionada) se correlaciona directamente con el producto de elongación obtenido finalmente mediante el método de la invención. El método se describió anteriormente en el sumario y se describe adicionalmente en las reivindicaciones. El método se realiza en un único volumen creciente gradualmente, en particular simplemente añadiendo reactivos adicionales a la mezcla de reacción, sin necesidad de purificación mediante aislamiento de componentes de la mezcla en las etapas a) a e). Los reactivos adicionales pueden neutralizar en cierta medida o acumular los componentes que ya están en el fluido. Este enfoque no sólo simplifica el procesamiento sino que aumenta la fiabilidad del método ya que todos los productos de reacción intermedios se mantienen siempre en una fase de volumen coherente. Además de reducir en las preparaciones manuales cualquier purificación intermedia, esto facilita enormemente la implementación del método sobre un chip o dispositivo microfluídico adecuado para automatización.

- Se mide la expresión génica alineando y agrupando los productos de amplificación o lecturas en anotaciones de genes y contando esas lecturas sin necesidad de alinear las lecturas con armazones de transcritos, y sin aplicar algoritmos de normalización específicos de transcritos posteriores para intentar eliminar determinados sesgos específicos de longitud y secuencia, que son necesarios para los métodos con más de una lectura por molécula de ARN.

- 45 El método de la invención es corto ya que requiere dos eventos de cebado, dos reacciones de polimerasa, una hidrólisis intermedia de ARN, que va seguida por un aislamiento, amplificación o purificación final, que tiene como objetivo seleccionar los productos correspondientes a la región de ácido nucleico preseleccionada del ARN molde.

Se sabe que sólo una lectura por transcrito proporciona mayor exactitud en el recuento de la expresión génica. Un nuevo aspecto del método de la invención es la reducción en la profundidad de lectura requerida debido a la normalización de longitud que se produce, ya que sólo se analiza una parte, que tiene una longitud limitada en comparación con ADNc de longitud completa.

- 50 La restricción de la señal de expresión génica a la región de ácido nucleico preseleccionada del ARNm permite menores costes relativos de secuenciación, que se estima actualmente que son de aproximadamente 1/5 en comparación con la secuenciación de longitud completa convencional, y un valor de expresión génica más correcto debido a que se produce la normalización de longitud a nivel de la preparación de muestra. Por tanto, no es necesaria una precognición de la longitud de las variantes de transcrito para calcular valores correctos de FPKM (fragmentos por kilobase de transcrito por millón de lecturas mapeadas). El contenido de información proporcionado por la región de ARNm es suficiente para el fin de clasificar muestras en análisis a gran escala, aunque el contenido de información es menor que un análisis de transcritos a escala completa pero mayor de lo que proporcionaría una simple expresión génica.

El producto de ácido nucleico generado también puede considerarse un producto de amplificación puesto que se usan reacciones de amplificación de ácido nucleico, pero por supuesto el propio ARN molde no se copia en su totalidad y, así, no se multiplica mediante el procedimiento. El método tiene como objetivo “amplificar” o simplemente generar un polinucleótido que comprende una secuencia copiada de una región del ARN molde. Esta secuencia copiada es una parte del ARN molde y se encuentra en la dirección 5' de la región de ácido nucleico preseleccionada que se usa en la etapa de unión a cebador a). La secuencia copiada habitualmente tiene una longitud de aproximadamente 25 a 2000 nucleótidos, aproximadamente de cerca de 100, 200, 300, 500 ó 1000 nt. Los valores exactos pueden diferir y se ven influidos por parámetros y los reactivos usados por el profesional. En esencia, un profesional puede adaptar a medida la longitud de región obtenida modificando, por ejemplo, la cantidad y constitución del cebador usado en la reacción, especialmente los cebadores de la etapa d), que pueden ser cebadores al azar. El profesional puede adaptar a medida la longitud de región promedio para que sea óptima para la posterior NGS o cualquier otra secuenciación.

Cuando, según una posibilidad de la invención, cada transcrito (molécula de ARN molde) o secuencia seleccionada como diana (región de ácido nucleico preseleccionada, también pueden ser dos o más por transcrito) genera simplemente una lectura (basado en el producto de elongación) en comparación con múltiples lecturas a lo largo de todo el transcrito y estas lecturas pueden iniciarse o terminar todas con el mismo nucleótido, podría surgir ambigüedad si estas lecturas se originan a partir de diferentes copias de un transcrito o si se originan a partir de eventos de duplicación por PCR. Por tanto, opcionalmente pueden usarse códigos de barras durante la primera reacción de extensión para etiquetar (código de barras) cada evento de cebado con el fin de distinguir múltiples copias de transcrito de eventos de duplicación por PCR para determinar el verdadero alcance de una nueva toma de muestras [documento US 2011/0160078]. De manera ideal, tales códigos de barras se introducen como códigos de barras al azar en la secuencia de ligador. Preferiblemente no participan en la reacción de cebado. Cada lectura (o producto de elongación) tendrá entonces un código de barras único individualmente distinto de otras lecturas (productos de elongación).

La duplicación por PCR proporcional no está provocando la mayor exactitud reivindicada sino que indica que la profundidad de lectura aplicada supera la complejidad de la biblioteca de NGS, de modo que la serie de secuenciación se inicia para secuenciar (leer) copias de los mismos insertos de nuevo.

La duplicación por PCR *per se* no es un problema, pero ver lecturas que se inician y terminan con la misma secuencia pueden hacer creer al usuario que está secuenciando con demasiada profundidad, o que la complejidad de la biblioteca es demasiado baja. Debido a que todas las lecturas de un transcrito se inician con la misma secuencia adyacente a la cola de poliA (u otras secuencias seleccionadas como diana), y a menudo terminan en secuencias preferidas, las lecturas aparecen más a menudo como que son duplicados de PCR aunque no lo sean. Firmas tales como códigos de barras al azar que se introducen durante la síntesis de primera hebra permiten, por tanto, distinguir lecturas singulares genuinas de duplicados.

La etapa preliminar, obtener o proporcionar un ARN molde, es la provisión de una muestra que contiene cualquier ARN, tal como ARN total de una célula. También pueden seleccionarse fracciones de ARN especiales, tales como la fracción de ARNm o uno de los siguientes tipos de ARN. Aunque, el ARN es preferiblemente un transcrito o ARNm, se prefiere especialmente de manera que comprenda un sitio o una cola de poliA, por supuesto pueden usarse y analizarse otras moléculas de ARN, tales como pre-miARN, miARN, pre-ARNt, ARNt, pre-ARNr o ARNr, uno cualquiera de los mismos, solo o en combinación con otros tipos de ARN, puede estar comprendido en el ARN. Preferiblemente, el ARN molde comprende un sitio o una cola de poliA. Si no está presente *per se* en la especie de ARN, puede añadirse una cola artificialmente mediante una reacción de prolongación de cola. Por supuesto, también pueden añadirse otras colas distintas a poliA, por ejemplo mediante una reacción de ligamiento usando una ligasa (componente opcional del kit) tal como se describe, por ejemplo, en el documento WO 2007/062445. El primer cebador de la etapa a) debe aparearse entonces con una secuencia de esta cola (artificial). En las siguientes etapas, preferiblemente se genera un ADNc durante el método de la invención usando una ADN polimerasa, preferiblemente una ADN polimerasa dependiente de ARN. Alternativamente, pueden seleccionarse como diana regiones específicas de transcritos de interés, por ejemplo, transcritos implicados en la generación de enfermedades tales como cáncer, inmunodeficiencias, durante el cebado inicial del ARN usando cebadores específicos de transcrito.

El ARN molde puede ser de cualquier longitud, pero preferiblemente está en el intervalo de 20 a 100000 nt (nucleótidos), se prefiere especialmente de 30 a 50000 nt, se prefiere más de 50 a 25000 nt, de 75 a 10000 nt o de 100 a 8000 nt.

Preferiblemente, la prolongación de cola (opcional) del extremo 3' se realiza usando transferasa terminal (componente opcional del kit). Aunque también se dan a conocer otros métodos de prolongación de cola, como ligamiento de una secuencia de cola, que puede ser, por ejemplo, una secuencia preseleccionada definida. La transferasa terminal puede añadir un determinado número de nucleótidos preferiblemente seleccionados uniformemente de un tipo de nucleótido. También puede usarse cualquier otro medio para la prolongación de cola, añadiendo una secuencia de cola, por ejemplo, mediante ligamiento de la secuencia de cola que puede ser uniformemente de un tipo de nucleótidos o de nucleótidos diversos. Una cola de este tipo es preferiblemente una secuencia en el intervalo de entre 5 y 500 nucleótidos, se prefiere más de menos de 400, menos de 300, menos de 200, menos de 100, menos de 50 o menos de 30 nucleótidos. Cualquier cola de este tipo (o parte de la misma)

puede usarse como la región preseleccionada de ácido nucleico 3'-terminal en la etapa a) con la que puede aparearse un cebador.

El método de la invención es particularmente adecuado para analizar mezclas complejas de diversas moléculas de ARN con diferentes secuencias de ácido nucleico. Preferiblemente, se obtienen y/o usan 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 o más, especialmente al menos 20, al menos 30, al menos 40, al menos 50, al menos 75, al menos 100, al menos 200 o más ARN molde diferentes de diferentes secuencias en el método de la invención.

La etapa a), aparear un primer cebador oligonucleotídico en una región de ácido nucleico preseleccionada del ARN molde, contiene proporcionar un primer cebador que se aparea o hibrida en condiciones de hibridación (por debajo de la temperatura de fusión de la doble hebra) con/a la región preseleccionada. Por tanto, comprende una secuencia complementaria de longitud suficiente para la reacción de apareamiento o la hibridación. La región complementaria puede ser una cualquiera usada habitualmente en la técnica, tal como de 6 a 40 nt de longitud, preferiblemente al menos de 6, 7, 8, 9, 10 ó más nt. La región preseleccionada es una de una secuencia conocida o esperada, tal como una cola de poliA común para ARNm de eucariotas. Puede usarse cualquier otra secuencia conocida, tal como secuencia o secuencias específicas de transcrito o gen que selecciona una o más dianas específicas de interés. Tales una o más secuencias diana pueden usarse para crear paneles específicos de enfermedad, tales como, por ejemplo, para cáncer o inmunodeficiencias. La región de ácido nucleico preseleccionada puede estar presente en uno o múltiples ARN molde de interés. Estos moldes de interés podrían compartir una propiedad común, tal como que están relacionados con una enfermedad o un estado específico. Preferiblemente, este panel de moldes de interés comprende 2, 3, 4, 5, 6, 7, 8, 9, 10 moldes, o más tal como al menos 15, al menos 20 y cualquier intervalo entremedias, que comprenden una región de ácido nucleico preseleccionada. Podrían usarse múltiples primeros cebadores para aparearse con este panel en una reacción (multiplexación).

Preferiblemente, pero no necesariamente, es una región 3'-terminal preseleccionada del ARN molde, tal como una cola de poliA u otra cola añadida durante la prolongación de cola tal como se describió anteriormente. La región preseleccionada puede tener una secuencia que es característica para un tipo de ARN de interés (por ejemplo, tal como se dio a conocer anteriormente). También, tal como se afirmó anteriormente, la región preseleccionada puede unirse artificialmente al ARN molde, por ejemplo mediante ligamiento o prolongación de cola.

El cebador puede contener una región con una secuencia que se aparea con el ARN molde, por ejemplo una secuencia complementaria que hibrida mediante apareamiento de bases, y opcionalmente una región que no se une, por ejemplo al tener una secuencia no complementaria y/o una secuencia que se bloquea mediante un oligonucleótido en un híbrido con esta región que impide la hibridación adicional. Esta región que no se une incluye preferiblemente un código de barras (preferiblemente al azar) para distinguir múltiples copias de transcrito de los eventos de duplicación por PCR. Preferiblemente, tal código de barras está ubicado en la región bloqueada. La región de apareamiento puede ser, por ejemplo, una región de oligo-dT₈ a oligo-dT₃₅, preferiblemente una región de oligo-dT₁₅ a oligo-dT₃₀, por ejemplo oligo-dT₁₀ u oligo-dT₂₅, que aumenta la selectividad y disminuye los eventos de cebado interno, que pueden producirse en sitios ricos en A internos dentro de los ARNm, si no son regiones preseleccionadas deseadas del ARN molde.

Preferiblemente, el primer cebador es un cebador de ADN, opcionalmente modificado tal como se describe a continuación para cebadores al azar.

La etapa b), elongar el primer cebador oligonucleotídico de manera específica al molde obteniendo de ese modo una primera hebra elongada, puede realizarse con cualquier reacción de elongación de oligonucleótido específico de molde, preferiblemente usando una nucleótido polimerasa, preferiblemente una ADN polimerasa, especialmente una transcriptasa inversa con el ARN como molde para la transcripción inversa. La primera hebra elongada está entonces habitualmente en una doble hebra que comprende el ARN molde como hebra complementaria. La primera hebra elongada es un molde para una reacción de elongación de cebador adicional en las siguientes etapas. La transcripción inversa puede realizarse usando cualquier transcriptasa inversa, tal como se describe adicionalmente a continuación (con o sin actividad de desplazamiento de hebra) por ejemplo con RT de VLM-M. Preferiblemente, al menos cierto desplazamiento de hebra está presente para permitir que la polimerasa desenrolle estructuras secundarias de ARN. En caso de transcriptasas inversas que usan ARN molde, en realizaciones preferidas la transcripción inversa se lleva a cabo en condiciones que no permiten la formación de la estructura secundaria o terciaria del ARN molde (híbridos ARN:ARN) o en condiciones que permiten que se sometan a desplazamiento de hebra estas estructuras secundarias por la transcriptasa inversa. La polimerasa usada durante la reacción de elongación puede ser una polimerasa viral, y puede seleccionarse del grupo que consiste en RT de VMA (y mutantes de la misma tal como RT Thermoscript), RT de VLM-M (y mutantes de la misma incluyendo pero sin limitarse a Superscript I, II o III, Maxima RT, RevertAid, RevertAid Premium, Omniscript, GoScript), RT de VIH, RT de VSR, RT de VAIE, RT de VRA2, ADN polimerasa de Tth, ADN polimerasa de *C. hydrogeniformans*, virus de leucosis de sarcoma aviar (VLSA) y mutantes de ARNasa H- de las mismas. Pueden usarse mezclas de cualquiera de estas polimerasas. En particular, pueden usarse mezclas de polimerasas virales, tales como mezclas de VLM-M y VLSA, y/o sus análogos de ARNasa H reducida o ARNasa H negativa. En cualquiera de estos métodos y composiciones, pueden usarse dos o más polimerasas, incluyendo cualquier polimerasa tal como se describió anteriormente.

La etapa c), retirar el ARN molde al menos de la doble hebra, significa que la primera hebra elongada, al menos en una región 3'-terminal, se libera del ARN molde. La doble hebra puede fundirse y el ARN molde retirarse mediante purificación, pero tal purificación se prefiere menos ya que añade etapas laboriosas adicionales, o se digiere. La digestión puede avanzar por completo o parcialmente. Pueden quedar partes de ARN cortas en la primera hebra elongada ya que el método de la invención no requiere el acceso de longitud completa a la primera hebra elongada. La digestión puede realizarse usando una ARNasa o calentamiento, especialmente en presencia de agentes de desestabilización de ARN adicionales, tales como condiciones alcalinas o cationes divalentes, tales como Mn^{2+} . Preferiblemente, retirar el ARN molde comprende la digestión enzimática de ARN, preferiblemente mediante una ARNasa, degradación alcalina, preferiblemente mediante tratamiento con NaOH, o calentamiento en presencia de cationes divalentes, preferiblemente Mn^{2+} o Mg^{2+} .

La reacción en la etapa e) garantiza que sólo se produce un producto de elongación usando condiciones sin desplazamiento de hebra. Sin desplazamiento de hebra sólo el cebador más dirigido en sentido 3' a la región preseleccionada se extiende satisfactoriamente hasta la ubicación correspondiente al primer cebador. Con el fin de impedir que el ARN molde interfiera en esta reacción, preferiblemente después de la transcripción inversa se retira el ARN, preferiblemente se hidroliza, por completo o al menos en tal grado que sólo quedan fragmentos cortos que presentan menores temperaturas de fusión que los siguientes segundos cebadores durante la síntesis de segunda hebra.

El ARN experimenta degradación espontánea a altas temperaturas si están presentes cationes divalentes. Los cationes divalentes pueden enmascararse más tarde si no se retiran mediante agentes quelantes, tales como EDTA o EGTA. Si no las muestras se purifican adicionalmente con métodos basados en precipitación o columna, la concentración final del agente quelante debe equilibrarse de tal modo que proporcione protección frente a la degradación de cualquier producto más adelante y sin inhibir la posterior reacción enzimática que también puede requerir cationes divalentes para sus actividades (por ejemplo, Mg^{2+} para polimerasas). Se produce una hidrólisis rápida de ARN, por ejemplo, en presencia de cationes divalentes a temperaturas de al menos 70°C, por ejemplo, a 75°C y/o hasta 98°C.

La hidrólisis del ARN se realiza preferiblemente usando $MnCl_2$ y altas temperaturas, lo que deja intacto el ADNc mientras se destruye el ARN. Este es un enfoque mucho más rentable que usar ARNasas. También pueden usarse condiciones alcalinas, tales como mediante adición de NaOH, para hidrolizar el ARN, pero debe tenerse mucho más cuidado para proteger el ADNc frente a la degradación también, por ejemplo, usando menores temperaturas o un pH menos alcalino, que se ajusta de modo que se degrada el ARN pero no el ADNc.

La etapa d), aparear más de un cebador oligonucleotídico con la primera hebra elongada, requiere la unión de al menos un cebador a la primera hebra elongada. Esta etapa se realiza esencialmente según los mismos principios que los descritos para el apareamiento del primer cebador en la etapa a). La secuencia de los cebadores adicionales puede ser una que se conoce para el ARN molde de interés o puede ser desconocida. Se prefiere usar cebadores al azar, que no requieren el conocimiento de la secuencia complementaria. Como con la etapa a) descrita anteriormente, es posible la multiplexación. También es una opción en la etapa d) múltiples cebadores oligonucleotídicos adicionales, que opcionalmente se aparean específicamente con una o más regiones diana específicas de elección, permitiendo de ese modo la selección específica de un ARN molde o secuencia génica en el mismo por cebador oligonucleotídico adicional.

En el cebado al azar, se usa una población de oligonucleótidos de secuencia al azar, habitualmente una secuencia oligomérica al azar de pentámero, hexámero, heptámero, octámero, nonámero, decámero, undecámero, dodecámero o más larga para cebar la reacción de elongación en cualquier lugar dentro de la hebra de ácido nucleico molde, en este caso la primera hebra elongada. El cebador puede comprender por supuesto nucleótidos adicionales además de esta secuencia oligomérica al azar. Opcionalmente, pueden usarse códigos de barras al azar adicionales para distinguir múltiples copias de transcrito de los eventos de duplicación por PCR. Preferiblemente, los cebadores adicionales son ADN, opcionalmente modificado tal como se describe a continuación.

“Cebadores al azar” ha de entenderse como una mezcla de diferentes cebadores con diferentes partes de secuencia de cebador, con una alta variancia debido a una síntesis al azar de al menos una parte de la secuencia de cebador. Los cebadores al azar cubren potencialmente toda el área combinatoria para dicha secuencia. La parte de cebador de secuencia al azar del cebador al azar puede cubrir 1, 2, 3, 4, 5, 6, 7, 8 o más nucleótidos al azar o nucleótidos universales. Los nucleótidos al azar se seleccionan aleatoriamente de A, G, C o T (U) en una posición de nucleótido dad. En cuanto a la hibridación de secuencias de cebador, se usan T y U en secuencias de manera intercambiable en el presente documento. Las posibilidades combinatorias para una parte de secuencia al azar es m^n , en el que m es el número de tipos de nucleótidos usados (preferiblemente los cuatro de A, G, C, T(U)) y n es el número de los nucleótidos al azar. Por tanto, un hexámero al azar, en el que se representa cada posible secuencia, consiste en $4^6 = 4096$ secuencias diferentes. Un cebador al azar también puede comprender uno o más nucleótidos, que no se unen específicamente a un nucleótido complementario como lo hacen A, T, C o G. Tales nucleótidos también se denominan “bases de bamboleo” o “bases universales”. Pueden usarse nucleótidos con bases universales, tales como desoxi-inosina, 3-nitropirrol-2'-desoxinucleósido y 5-nitroindol-2'-desoxinucleósido. Las bases universales producirán apareamiento de bases con cualquier nucleótido de A, C, G, T(U) o al menos dos o tres nucleótidos de los mismos. No es necesario incluir todas las posibilidades para un cebador al azar de este tipo. En algunas

realizaciones, el cebador al azar comprende al menos un nucleótido al azar (permutación en una posición tal como se describió anteriormente) y/o al menos un nucleótido de bamboleo. En el contexto de cebadores al azar o cebadores de una secuencia seleccionada, se usan al menos 10, preferiblemente al menos 20, se prefiere especialmente al menos 100, cebadores diferentes.

5 Especialmente para, pero sin limitarse a una representación óptima en una elongación cebada al azar, pueden estar presentes cebadores en una concentración de desde 10 nM hasta 100 μ M, y se prefiere más a aproximadamente 1 μ M pero también puede ser de al menos 200 nM. En realizaciones preferidas, la razón (p/p) de cebador con respecto a ácidos nucleicos molde es de entre 5:1 y 1:1000, preferiblemente entre 2:1 y 1:500, preferiblemente entre 1:1 y 1:300, preferiblemente entre 1:2 y 1:250, preferiblemente entre 1:5 y 1:150, preferiblemente entre 1:10 y 1:100, preferiblemente entre 1:12 y 1:50. La razón molar de cebador con respecto a ácidos nucleicos molde puede ser de entre 100:1 y 1000000:1, preferiblemente entre 1000:1 y 1000000:1, entre 10000:1 y 500000:1, o entre 20000:1 y 300000:1. En un ejemplo, usando 100 ng de material de partida de ARNm y suponiendo una longitud del ARNm de 500 a 5000 nt con un valor medio de 2000 nt y añadiendo 1 nmol de cebadores, entonces están presentes los cebadores en un exceso molar de 6800:1.

15 Los cebadores adicionales pueden contener una región con una secuencia que se aparea con la primera hebra elongada, y opcionalmente una región que no se une, por ejemplo al tener una secuencia no complementaria y/o una secuencia que se bloquea mediante un oligonucleótido en un híbrido con esta región que impide la hibridación adicional. Esta región que no se une también puede incluir códigos de barras, preferiblemente códigos de barras al azar para distinguir múltiples copias de transcrito de los eventos de duplicación por PCR.

20 La etapa e), elongar los más de un cebador oligonucleotídico adicionales de manera específica al molde sin desplazamiento de los cebadores apareados con la primera hebra elongada usando una polimerasa que carece de actividad de desplazamiento de hebra, preferiblemente una ADN polimerasa de T7, T4 o Q5, y/o usando cebadores que tienen resistencia al desplazamiento de hebra por una polimerasa, generando de ese modo productos de elongación adicionales, sigue conceptos similares que los de la etapa b). Puede usarse cualquier método de elongación adecuado para el molde dado, por ejemplo ADN. En realizaciones preferidas se usa una polimerasa (dependiente de ADN si el molde es ADN o dependiente de ARN si el molde es ARN), preferiblemente una ADN polimerasa si el producto de elongación adicional va a ser ADN.

30 Como las ADN polimerasas pueden desplazar un oligonucleótido de ADN de una hebra molde de ADN al menos igual de bien que disuelven la estructura secundaria o terciaria, la hibridación del oligonucleótido puede potenciarse con el fin de detener el desplazamiento de hebra de la polimerasa. Una ADN polimerasa con actividad de desplazamiento de hebra particularmente intensa es la polimerasa de Klenow (fragmento de Klenow). Puede impedirse el desplazamiento usando modificaciones del propio oligonucleótido. Modificaciones de los oligonucleótidos que reducen o inhiben la actividad de desplazamiento de hebra de la polimerasa son, por ejemplo, 2'-fluoronucleósidos, PNA, ZNA, abrazadera de tipo G (documento US 6.335.439, un análogo de citosina que puede producir unión de tipo abrazadera a guanina) o LNA (documentos US 2003/0092905; US 7.084.125). Estas modificaciones aumentan en general la temperatura de fusión del oligonucleótido, aumentando la energía de hibridación local del oligonucleótido a la hebra de ADN o ARN molde en comparación con el mismo oligonucleótido sin la modificación o rigidiza la estructura principal de azúcar-fosfato del oligonucleótido. Algunas también rigidizan la estructura principal de azúcar-fosfato inhibiendo además el desplazamiento de hebra por la polimerasa. Se dan a conocer medios para detener el desplazamiento de hebra (SDS) en el documento WO 2013/038010 A1.

45 Alternativamente o además, la hibridación del cebador al molde (por ejemplo, primer producto de elongación) puede alterarse usando diferentes aditivos que se unen o intercalan a los ácidos nucleicos. Por ejemplo, pueden usarse bromuro de etidio, SybrGreen (documentos US 5.436.134; US 5.658.751; US 6.569.627) o acridina, preferiblemente intercaladores que son específicos para híbridos ARN:ADN o ADN:ADN. Otros compuestos que pueden unirse a dsNA son actinomicina D y análogos, aminoglicósidos de la familia de la neomicina (neomicina, ribostamicina, paromomicina y frameticina). También pueden incluirse de manera covalente aditivos que alteran las propiedades de hibridación del cebador en la estructura del cebador.

50 Pueden cambiarse la cinética y energía de hibridación para inhibir el desplazamiento de hebra por la polimerasa mediante la adición de proteínas de unión a ácido nucleico tales como proteína de unión monocatenaria tal como SSB de Tth o RecA de Tth.

Resultará evidente para los expertos en la técnica que esos aditivos son simplemente ejemplos y puede usarse cualquier otro compuesto, modificación de bases o enzima que conduzca a una estabilidad aumentada del híbrido cebador-molde para aumentar la Tf y, por tanto, inhibir el desplazamiento de hebra o inhibir la capacidad de desplazamiento de hebra de la enzima implicada.

55 El aumento de la Tf debe ser lo suficientemente fuerte como para impedir un desplazamiento de uno cualquiera de los nucleótidos del extremo 5' de la región del cebador apareada con el molde mediante una polimerasa de elongación. En particular, el aumento de Tf de la invención impide el desplazamiento del 3^{er}, 2^o y/o 1^{er} nucleótido en el sentido de 3' con respecto al extremo 5' de la región del cebador que se aparea con el molde (pueden existir nucleótidos en 5' no apareados adicionales, por ejemplo, regiones de ligador o códigos de barras, que no es

necesario modificar).

En determinadas realizaciones de la invención, es necesario detener el desplazamiento de hebra justo en el primer nucleótido en 5' del cebador en el sentido de 3'.

5 Por tanto, se prefiere que la unión de los cebadores oligonucleotídicos tenga una Tf potenciada específicamente en sus extremos 5' de la región apareada con el molde para impedir que la polimerasa de elongación los desplace. Tales modificaciones incluyen pero no se limitan a LNA, PNA, ZNA, acridina o fluoróforos.

10 Los oligonucleótidos con una Tf aumentada en su extremo 5' tal como oligonucleótidos modificados con LNA permiten una detención justo en el inicio del siguiente cebador. Está dentro del alcance de la invención combinar la detención del desplazamiento de hebra usando los oligos modificados con LNA junto con una polimerasa sin actividad de desplazamiento de hebra así como la disminución de la temperatura de reacción y el uso de diferentes aditivos para aumentar la unión de cebadores al molde.

15 Preferiblemente se modifican nucleótidos C y/o G. Estos nucleótidos, incluso sin modificar, tienen una mayor Tf que A o T debido a un aumento de la formación de puentes de hidrógeno cuando se aparean de manera complementaria. En realizaciones preferidas, el cebador oligonucleotídico comprende al menos uno, al menos 2, al menos 3, al menos 4, al menos 5, al menos 6 nucleótidos modificados que se seleccionan de G o C. Estos nucleótidos modificados están preferiblemente en el extremo 5' de la secuencia de cebador que se aparea con el molde tal como se mencionó anteriormente.

20 Se logra la detención del desplazamiento de hebra más eficaz mediante bases G o C ya que aumentan la Tf local del cebador o agente de detención. Así, cebadores semialeatorios (hexámeros, heptámeros, octámeros, nonámeros, etc.) que contienen al menos dos, más preferiblemente tres o más G o C o una combinación de G y C. Lo más preferido es si estas G o C se modifican para aumentar la temperatura de fusión local, como en el caso cuando se usan bases modificadas con LNA. Lo más preferido es que se usen al menos 1, al menos 2 o al menos 3 bases modificadas con LNA en el extremo 5' de la región del cebador que se aparea. Por tanto, se prefiere que se usen al menos dos, al menos 3 nucleótidos modificados opcionalmente elegidos de G o C.

25 Existen varios métodos y medios para garantizar que se detiene la reacción de elongación cuando la reacción de elongación alcanza la posición de un cebador adicional apareado con el molde. Esta detención también se denomina impedir el desplazamiento de hebra en el presente documento. La etapa de la invención de impedir que la polimerasa produzca desplazamiento de hebra del/de los siguiente(s) cebador(es) de una parte de polinucleótido ya copiada garantiza que no se copia de nuevo ninguna parte de una molécula de polinucleótido que ya se copió, y en particular que no se copia de nuevo la parte orientada más en 5' en la primera hebra elongada, que corresponde a la parte más en 3' del ARN molde original. Por tanto, ninguna parte copiada del polinucleótido llega a estar sobrerrepresentada en la síntesis de segunda hebra, y en particular dicha parte copiada orientada más en 5' se sintetiza sólo una vez a partir de cada molde de primera hebra sintetizado. Esta inhibición del desplazamiento de hebra puede lograrse a través de diferentes medios, tales como disminución de la temperatura de reacción, uso de una polimerasa sin actividad de desplazamiento de hebra, aumento de la temperatura de fusión o la energía de hibridación del híbrido cebador:molde o aumento de la rigidez del ARN o cebador o estabilización de la hélice. En la práctica, habitualmente se selecciona una combinación de estos medios para lograr condiciones de reacción óptimas sin desplazamiento de hebra. Se posibilita que un experto en la técnica seleccione de manera correcta parámetros adecuados tal como se describe en el presente documento o se conoce en la técnica para adaptarse a un molde y condiciones de reacción particulares.

30 Una opción es modificar la temperatura de reacción. En general, una temperatura de reacción superior a 37°C, en particular superior a 70°C está favorecida durante la elongación para disolver mejor las estructuras secundarias en el molde, lo que conduce a una síntesis por desplazamiento más eficaz. En una realización, se logra la detención del desplazamiento de hebra del cebador disminuyendo la temperatura de reacción. Se usan temperaturas de reacción inferiores a 37°C y de tan sólo 25°C, y además de 4°C, para reducir el desplazamiento de hebra. Sin embargo, incluso a menores temperaturas de reacción la detención del desplazamiento de hebra no será completo cuando se usan polimerasas que tienen actividad de desplazamiento de hebra y /o se usa un simple oligonucleótido como agente de detención que no tiene modificaciones que alteren su temperatura de fusión. Se prefiere que se lleve a cabo la polimerización a entre 12°C y 37°C.

35 40 45 50 55 En una realización en lugar de, o además de, disminuir la temperatura de reacción para lograr una mejor detención de la elongación en dicha posición de un cebador adicional o un agente de detención (y reducir el desplazamiento de hebra) pueden usarse polimerasas que son deficientes en el desplazamiento de hebra. En caso de ADN-ADN polimerasas, preferiblemente se usa ADN polimerasa de T7, T4 o Q5 en la elongación del uno o más cebadores oligonucleotídicos de manera específica de molde. ADN polimerasa de T4 es especialmente eficaz y se prefiere en todas las realizaciones.

La polimerasa puede ser una polimerasa mesófila o termófila, especialmente ADN polimerasa.

Las polimerasas mutantes deficientes en el desplazamiento de hebra pueden ser capaces de desplazar el siguiente cebador para hasta 3 nt cuando no se modifican. Está dentro del alcance de la invención combinar la detención del

desplazamiento de hebra disminuyendo la temperatura de reacción con el uso de mutantes deficientes en la síntesis por desplazamiento o cualquier otra polimerasa con síntesis por desplazamiento afectada.

El aumento de la concentración de contraiones monovalentes también estabilizará cualquier híbrido molde-cebador (pero también la estructura secundaria). La concentración de iones positivos monovalentes se selecciona preferiblemente de al menos 20 mM, al menos 30 mM, al menos 40 mM, al menos 50 mM, al menos 60 mM, al menos 70 mM.

Alternativamente o en combinación con una cualquiera de las opciones anteriores, puede aumentarse el impedir el desplazamiento de hebra mediante la presencia de un agente de aglomeración, preferiblemente PEG. Los agentes de aglomeración son moléculas inertes que pueden usarse en altas concentraciones y pueden usarse para imitar los efectos de la aglomeración macromolecular en el interior de una célula. Ejemplos son PEG (polietilenglicol), PVP (polivinilpirrolidona), trehalosa, Ficoll y dextrano. Los agentes de aglomeración se dan a conocer, por ejemplo, en el documento US 5.554.730 o el documento US 8.017.339. Otros aditivos que actúan como agente de aglomeración son Tween-20, NP-40 podría añadirse adicionalmente o en lugar de PEG. Dentro del alcance de la invención se usan preferiblemente PEG-8000 final al 12%-25% (v/v). Puede usarse una variedad de pesos moleculares y compuestos de PEG, y el experimentador experto apreciará que pueden variarse la identidad y concentración del aditivo para optimizar los resultados. Los agentes de aglomeración están presentes preferiblemente en la etapa e) para disminuir el riesgo de desplazamiento de hebra. También pueden estar presentes alternativamente o en combinación en cualquier otra etapa, tal como en una etapa de purificación, especialmente en una de precipitación. El kit puede comprender un agente de aglomeración, preferiblemente en un tampón para la etapa de reacción e), por ejemplo un tampón que comprende un cofactor para una polimerasa, tal como Mg^{2+} . Después, el agente de aglomeración también puede estar presente en un tampón de precipitación en el que su concentración se aumentará hasta tal grado para que sea suficiente para la precipitación de polinucleótidos, especialmente de los productos de elongación.

La etapa f), aislar y/o amplificar un producto de elongación de dicho producto de elongación adicional que comprende una parte de ácido nucleico que se elonga complementaria (o en complementariedad) al primer cebador oligonucleotídico, es una selección de los productos de elongación correctos de la etapa e) que corresponden a la región de ácido nucleico preseleccionada del ARN molde original. Puesto que se disuade o impide el desplazamiento de hebra en la etapa e), sólo habrá esencialmente un producto de elongación resultante de la etapa e) para cada molde (directamente, la primera hebra elongada pero también implícitamente el ARN molde original) puesto que se impedirá que se elonguen otros cebadores apareados hasta la región correspondiente a la región de ácido nucleico preseleccionada debido a la acción de bloqueo del cebador más en 3' que puede elongarse hasta dicha región preseleccionada a la que se une el primer cebador, y puede elongarse además hasta a lo largo de toda la secuencia de dicho primer cebador. Otros eventos de bloqueo pueden deberse a fragmentos remanentes del ARN molde si la retirada de la etapa c) deja algunos productos de degradación cortos apareados con la primera hebra elongada, lo que debe evitarse que se produzca a través de una retirada completa del ARN molde puesto que si no también disminuye la probabilidad de elongar satisfactoriamente el último cebador más en 3' en la región preseleccionada deseada.

Esta "selección" del producto de elongación adicional correcto (por cada molde) puede lograrse por ejemplo mediante un aislamiento, una purificación o una amplificación o generalmente cualquier procesamiento específico para el producto de elongación adicional que comprende una parte de ácido nucleico que se elonga complementaria al primer cebador oligonucleotídico. Un aislamiento puede comprender, por ejemplo, una unión a una sonda inmovilizada. Una amplificación puede usarse para obtener también una hebra complementaria del producto de elongación adicional que se produce selectivamente reconociendo el producto de elongación adicional como hebra molde usando un cebador. La amplificación puede ser un ciclo de PCR que comprende una reacción de apareamiento de cebador y elongación de hebra adicional. Para el aislamiento o la amplificación, puede usarse una secuencia conocida como secuencia de reconocimiento, especialmente para la unión a oligonucleótido. Una secuencia conocida de este tipo es, por ejemplo, una secuencia que es idéntica a la región de ácido nucleico preseleccionada original del ARN molde de la etapa a), así corresponde a la región de secuencia selectiva del primer cebador, o corresponde a cualquier otra región que esté incluida en el primer cebador, tal como secuencias de ligador o código de barras que se describen adicionalmente a continuación.

Según una preferencia de cualquier realización de la invención, el primer cebador oligonucleotídico de la etapa a) y/o los cebadores oligonucleotídicos adicionales de la etapa d) contienen una secuencia de ligador o etiqueta de secuencia sin apareamiento. Tales etiquetas de secuencia o ligadores pueden usarse para la unión de cebador de amplificación en otra reacción de elongación, especialmente PCR. La etiqueta de secuencia o ligador también puede comprender una secuencia única para cada cebador o tipo de cebador (especialmente en caso de los cebadores al azar) o identificador de secuencia ubicuo, también denominado código de barras o secuencia de código de barras. Las identidades o identificadores de secuencia pueden identificar cebadores (y posteriormente productos de elongación) de un experimento o lote particular, o producto de elongación individual o grupos de productos de elongación. Las etiquetas de secuencia permiten un análisis adicional mediante secuenciación masiva en paralelo. "Sin apareamiento" puede lograrse seleccionando una secuencia que no se aparea con su molde de hibridación (el ARN molde, primera hebra o de elongación adicional) y/o hibridando la secuencia sin apareamiento con otro oligonucleótido, bloqueando por tanto esta parte del cebador e impidiendo la hibridación con el molde.

También es posible que el ARN molde antes del apareamiento del primer cebador (en la etapa a) sea un ARN fragmentado, es decir un ARN obtenido por ejemplo a partir de una muestra se trata para que experimente fragmentación para proporcionar el ARN molde. Puede llevarse a cabo tal fragmentación usando cualquier medio conocido en la técnica. Puede iniciarse la fragmentación de manera dependiente de secuencia, por ejemplo mediante digestión con endonucleasa, o mediante medios independientes de secuencia tales como mediante medios físicos como sonicación o cizalladura, o tal como mediante medios químicos como hidrólisis. Si se usa un método dependiente de secuencia, por ejemplo digestión con endonucleasa de restricción o amplificación específica de secuencia, los extremos de fragmento presentarán un sesgo de secuencia. Una realización preferida es la fragmentación mediante hidrólisis o degradación limitada tal como se describió anteriormente para la etapa c), tal como calentamiento en presencia de cationes divalentes o en condiciones alcalinas, pero durante un tiempo limitado y/o a menores temperaturas para preservar fragmentos de ARN más grandes. Tales fragmentos pueden tener, por ejemplo, una longitud promedio de aproximadamente 100 a 5000 nt, preferiblemente de 300 a 3000 nt, se prefiere especialmente de 500 a 2000 nt. Es necesario mantener los fragmentos con una longitud mínima de más de 50 nt, preferiblemente más de 100 nt, de manera especialmente preferible más de 150 nt para preservar la parte de secuencia selectiva para que esté intacta en el fragmento de extremo 3' seleccionado como diana, para proporcionar una secuencia suficientemente larga para el cebado al azar durante la etapa e), y para mantener un inserto de secuencia lo suficientemente largo entre la primera secuencia complementaria y la secuencia de segundo cebador que puede mapearse en una anotación de genoma y/o transcriptoma. Se prefiere que el inserto de secuencia sea de más de 10 nt, más de 15 nt, preferiblemente más de 20 nt de largo, que se fija a menudo como requisito de longitud mínima para algoritmos de alineación de lecturas de secuenciación NGS de bioinformática.

Preferiblemente, el método comprende realizar una PCR con el producto de elongación adicional usando un cebador (o par de cebadores) específico para una etiqueta de secuencia o secuencias de ligador de dicho producto de elongación. Tales ligadores o etiquetas de secuencia pueden introducirse mediante los cebadores en la etapa a) y/o cebadores en la etapa d).

Estos cebadores adicionales de la PCR adicional pueden comprender etiquetas de secuencia o secuencias de ligador adicionales, que pueden usarse de nuevo en la amplificación por PCR. También pueden comprender estas etiquetas o ligadores un identificador de secuencia tal como un código de barras tal como se describió anteriormente para los primeros ligadores o etiquetas de secuencia mencionados.

En una preferencia, el método de la invención comprende además en la etapa f) purificar el producto de elongación de la etapa e). Tal purificación puede ser una selección de polinucleótidos con una longitud que corresponde a la longitud esperada del producto de elongación de la etapa e), por ejemplo de 150 a 500 nt. La longitud del producto de elongación de la etapa e) puede controlarse por ejemplo modificando las concentraciones de cebador de los cebadores al azar, es decir más cantidad de cebadores al azar conducirá a más eventos de cebado en la primera hebra y, así, el producto de elongación deseado que se selecciona entonces en la etapa f) estará más cerca del primer sitio de cebado y, por tanto, será de menor longitud. La purificación puede realizarse mediante precipitación de los productos de elongación mientras se mantienen cortos los polinucleótidos con una longitud de, por ejemplo, menos de 100 nt en disolución y retirada de los polinucleótidos en disolución. De 50 a 100 nt es una longitud típica para los productos de cebador-cebador que incluyen dos etiquetas de secuencia o ligadores, uno a cada lado. En una preferencia, la purificación es para retirar tales polinucleótidos cortos con una longitud de 70 nt o menos o de 50 nt o menos o de 40 nt o menos o de 30 nt o menos. El intervalo entre 40 y 70 nt es una longitud típica para cebadores y cebadores que incluyen una etiqueta de secuencia o ligador. Preferiblemente, el método comprende una inmovilización reversible en fase sólida que se une y libera selectivamente polinucleótidos de intervalos de tamaño definidos a y de superficies sólidas tales como superficies con un resto de grupos hidroxilo [Hawkins, 1995]. Puede imponerse la precipitación de polinucleótido dependiente de tamaño a través de un agente de aglomeración, preferiblemente a través de PEG usando condiciones de tampón específicas que incluyen concentraciones de sal y valores de pH definidos. Preferiblemente, los polinucleótidos cortos que van a retirarse no se unen, por ejemplo precipitados, sobre perlas recubiertas y se retiran con el sobrenadante, antes de liberarse los oligonucleótidos más largos deseados en un nuevo tampón que no contiene, o contiene menos agente de aglomeración. Preferiblemente tales perlas contienen un núcleo magnético [documento US 5705628]. Otros métodos de purificación incluyen métodos cromatográficos dependientes de tamaño tales como cromatografía de exclusión molecular.

En los métodos de la invención, las etapas a) a e) del método se realizan en un volumen de fluido creciente consecutivamente, por ejemplo, en un pocillo, un envase o un tubo. Todas las etapas de reacción que comienzan después de la provisión del ARN en una alícuota de disolución hasta el aislamiento o la amplificación del producto de elongación deseado que comprende las partes de ácido nucleico que se elongan complementarias al primer cebador oligonucleotídico se llevan a cabo en dicha disolución a la que se añaden gradualmente los demás componentes de reacción a través de la adición de disoluciones adicionales. La adición de los reactivos necesarios para realizar el método de la invención, por ejemplo, sustancias de partida, enzimas y cofactores en un fluido, crea una mezcla de reacción. Esencialmente, el método de las etapas a) a e) puede realizarse sin etapas de purificación adicionales. Por la presente, las propias acciones emprendidas para las etapas a) a e) no se consideran purificación, especialmente no para la retirada del ARN de la etapa c) puesto que los productos degradados del ARN después de la degradación pueden permanecer en disolución. En particular, el método de la invención de las etapas a) a e) realizado en un volumen de fluido creciente es preferiblemente un método sin retirada de fluido de la mezcla de reacción. Tal método simplifica la manipulación del procedimiento significativamente. También ayuda a mantener la

muestra y los posteriores productos de reacción al no purificar los productos intermedios ni dividir el volumen de reacción.

La presente descripción proporciona además un kit que comprende una transcriptasa inversa, dNTP, cofactores o sales de iones metálicos requeridos por una polimerasa, preferiblemente Mg^{2+} , un cebador, preferiblemente un cebador de poli-T, una ADN polimerasa sin actividad de desplazamiento de hebra tal como ADN polimerasa de T7, Q5 o T4, cebadores oligonucleotídicos al azar. El kit puede ser adecuado para realizar el método de la invención según cualquier realización con una cualquiera o una combinación de características preferidas. El cebador, preferiblemente cebador de poli-T o uno o una mezcla de múltiples cebadores específicos de transcrito o gen, es adecuado para la etapa a) y los cebadores al azar son adecuados para usarse en la etapa d). Estos cebadores o preparación de cebadores difieren en la composición de cebador y se proporcionan preferiblemente en recipientes independientes, tales como viales.

El kit puede comprender además un agente de degradación de ARN, tal como una enzima o un catión divalente, que va a usarse para la degradación de ARN a temperatura elevada, y/o un agente de aglomeración tal como PEG. Por supuesto, cualquier componente descrito anteriormente puede usarse en realizaciones alternativas o más preferidas.

Una ventaja adicional que logra la generación del producto de ácido nucleico según la invención es una normalización de longitud (todos los productos de elongación adicionales obtienen una longitud similar o una estrecha distribución de longitud) que libera el espacio de secuenciación que se habría usado alternativamente por lecturas que se generan a partir de los transcritos completos y más largos. La ganancia entendida como espacio de secuenciación ahorrado para obtener la misma información sobre la distribución de sitios de extremo de transcripción y expresión génica es una relación entre la variación de longitud de transcrito y el intervalo dinámico o de concentración de los transcritos en una muestra. La relación se ilustra de la mejor manera cuando se examinan dos condiciones de contorno. i) Si todos los transcritos tuvieran la misma longitud conocida, por ejemplo, todos los transcritos son de 1 kb de largo y todas las lecturas/fragmentos generados se mapean de manera única, entonces la normalización de longitud, a por ejemplo 100 pb, no tendría beneficio. Así, la ganancia no depende de la disminución de la longitud promedio sino de la disminución de la distribución de longitud. ii) Si todos los transcritos tuvieran longitudes diferentes y desconocidas, por ejemplo los transcritos son de entre 500 pb y 10000 kb de largo y muchas lecturas/fragmentos generados no pueden asignarse de manera única porque se mapean en exones que se comparten por varias variantes de transcrito del mismo gen, entonces la normalización de longitud, a por ejemplo 100 pb, tendría el beneficio de contar de manera inequívoca las lecturas y determinar el valor correcto de expresión génica en relación con el número global de lecturas. Por tanto, la distribución de longitud ponderada para la concentración, la exactitud de la anotación de secuencia (conocimiento previo) y la capacidad de mapear de manera inequívoca las lecturas que pueden asignarse de manera inequívoca al transcrito correcto son las medidas relevantes para determinar la complejidad de un transcriptoma. Esta complejidad puede reducirse significativamente mediante el método de la invención.

La oportunidad comercial del método de la invención se observa al reemplazar la obtención de perfiles de expresión génica por microalineamientos y proporcionar una herramienta de análisis intermedio por debajo de un análisis de transcriptoma de ARNm completo. La toxicogenómica y farmacogenómica son ejemplos de posibles aplicaciones. Usando cebadores específicos de transcrito, región o gen o en la etapa a) y/o en la etapa d) son posibles paneles de secuenciación dirigida. Cualquier combinación con primeros y segundo cebadores dirigidos (específicos de secuencia, se preselecciona/predefine) y no dirigidos (por ejemplo, a secuencia ubicua como una secuencia compartida por todas las secuencias de ARN de interés, como poliA, o secuencia al azar) es posible, tal como a) primer cebador específico de diana (=dirigido) y segundo cebador no dirigido; b) primer cebador no dirigido y un segundo cebador dirigido; c) primer cebador dirigido y segundo cebador dirigido; d) primer cebador no dirigido y segundo cebador no dirigido. Por supuesto, el beneficio de realizar el método de la invención en un volumen de fluido creciente consecutivamente se aplica a todas estas variantes, especialmente sin retirada/lavado de fluido. Un uso para b) es, por ejemplo, detectar la variación potencial en el lado 3', tal como eventos de corte y empalme alternativos, PAA alternativos de último exón alternativo así como eventos de fusión. Un uso para a) es, por ejemplo, detectar la variación potencial en el lado 5', tal como eventos de corte y empalme alternativos, primer exón alternativo así como eventos de fusión.

La presente invención se ilustra adicionalmente en las siguientes figuras y ejemplos, sin limitarse a estas realizaciones de la invención.

Figuras

Figura 1: esquema del método de la invención. La reacción avanza a través de la adición posterior de los reactantes y se representa en a) se inicia la síntesis de ADNc mediante cebado en una región o etiqueta conocida que o bien está presente ya (en este caso, poli(A) de ARNm) o bien se une en una reacción anterior. P1 es complementaria a dicha región conocida y contiene además una secuencia específica no complementaria en su extremo 5' que sirve como etiqueta universal. b) ARN se somete a transcripción inversa para dar ADNc por una polimerasa dependiente de ARN, es decir una transcriptasa inversa. c) Después de la síntesis de ADNc, se hidroliza el ARN molde o se degrada o bien por ARNasas, cambios de pH (NaOH y calor) o cationes divalentes (Mn^{2+} , Mg^{2+} y calor). d) Entonces, se ceba el ADNc monocatenario mediante múltiples cebadores al azar, P(n), P(n+1), ..., P(n+n), e) y se sintetiza una

segunda hebra usando una polimerasa dependiente de ADN sin desplazamiento de hebra. f) La falta de desplazamiento de hebra garantiza que sólo el fragmento más en 3' contendrá ambas etiquetas, una del cebador de síntesis de ADNc y la otra del cebador de síntesis de 2ª hebra. Este fragmento más en 3' se aísla y/o amplifica.

5 Figura 2: Representación esquemática de moléculas de ácido nucleico a modo de ejemplo que se producen en el ensayo. Dos oligos (Seq ID: 2 y Seq ID: 3) se hibridan a un ADN molde monocatenario (Seq ID: 9). Una polimerasa sin desplazamiento de hebra generará un fragmento de 65 nt de largo (Seq ID: 4) resultante de la elongación de Seq ID:3 y un fragmento de 85 nt de largo (Seq ID: 5) de la elongación de Seq ID: 2. Una polimerasa con desplazamiento de hebra generará un fragmento de 150 nt de largo (Seq ID: 6) de la elongación de Seq ID: 2 y el desplazamiento de Seq ID: 3. Adicionalmente, se producirían un fragmento de 65 nt de largo (Seq ID: 4) resultante de la elongación de Seq ID:3 y en caso de desplazamiento de hebra ineficaz, productos de entre 85 nt (Seq ID: 5) y 150 nt (Seq ID: 6).

10 Figura 3: Comparación de polimerasas con y sin actividad de desplazamiento de hebra a diferentes temperaturas de reacción. Se sometieron a prueba tres polimerasas diferentes, ADN polimerasa de T7 y T4, tanto sin desplazamiento de hebra como con desplazamiento de hebra a 3'-5' exo-fragmento Klenow a diferentes temperaturas de reacción. Las flechas rellenas blancas indican productos de detención de desplazamiento de hebra parcialmente desplazados. Las estructuras secundarias del molde monocatenario sin la etapa de desnaturalización se indican con una flecha negra.

15 Figura 4: Comparación adicional de polimerasas con y sin actividad de desplazamiento de hebra a diferentes temperaturas de reacción. 25°C para T4 es más recomendable que 37°C porque a 37°C la exonucleasa inherente toma el control de la reacción.

20 Figura 5: Comparación de diferentes métodos de degradación de ARN con MnCl₂, temperatura elevada sólo, tratamiento con NaOH o ARNasas. Se añadieron cantidades conocidas del ARN total aislado de hígado de ratón a un oligo de ADN de 111 nt monocatenario (ADNmc) (ID Seq ID: 7) véanse el carril 1 y el carril 10. El tratamiento con calor en un tampón para RT convencional, Tris-HCl 50 mM (pH 8,3 a 25°C), KCl 75 mM, MgCl₂ 3 mM y DTT 10 mM durante 30 minutos a 95°C y durante 5 minutos a 98°C, 10 minutos a 98°C, 20 minutos a 98°C, y 30 minutos a 98°C, da como resultado la degradación del ARN, pero no una retirada completa del ARN. La incubación de la mezcla de ARN/ADNmc con la mezcla de ARNasa H/ A / T1 durante 30 minutos o bien a 25°C o bien a 37°C retira por completo el ARN sin degradar el ADN monocatenario. La incubación a temperaturas elevadas durante 10 minutos (55°C) en presencia de NaOH 0,1 N degrada el ARN, aunque no por completo. Después de 10 minutos a 95°C en NaOH 0,1 N se retira por completo el ARN, sin embargo también comienza a degradarse el ADNmc (carriles 7-10). La adición de MnCl₂ 10 mM a la mezcla de ARN/ADNmc/tampón para RT y el tratamiento con calor durante 5, 10, 20 y 30 minutos a 98°C da como resultado la degradación completa del ARN sin degradar el ADNmc.

30 Figura 6: Efecto de la fragmentación de ARN inicial sobre la eficacia y el tamaño de biblioteca. Se fragmentaron a) y b) en presencia de MnCl₂ 6 mM y finalmente se amplificó durante 14 ciclos de PCR, mientras que se fragmentaron c) y d) en presencia de MnCl₂ 4 mM y requirieron 2 ciclos de PCR más. Todas las fragmentaciones avanzaron durante 3 min a 85°C.

35 Figura 7: Influencia de la concentración de cebador de RT y concentración de cebador de síntesis de 2ª hebra sobre el rendimiento y el tamaño de biblioteca. a) Cebador de polidT (RT) anclado 50 nM (SEQ ID No: 8) y oligo de síntesis de 2ª hebra 1 µM+ rc (SEQ ID No:9 y 10), b) cebador de polidT (RT) anclado 25 nM y oligo de síntesis de 2ª hebra 0,5 µM+ rc, biblioteca diluida 1:3 antes de la carga, c) cebador de polidT (RT) anclado 50 nM y oligo de síntesis de 2ª hebra 0,5 µM+ rc, biblioteca diluida 1:3 antes de la carga, d) cebador de polidT (RT) anclado 25 nM y oligo de síntesis de 2ª hebra 0,1 µM+ rc.

40 Figura 8: Influencia del método de degradación de ARN sobre la calidad de bibliotecas NGS. El ARN se ha degradado a través de a) MnCl₂ 10 mM durante 10 min a 95°C, b) 5000 U de ARNasa H durante 30 min a 37°C, c) tampón para la transcripción inversa durante 10 min a 95°C, d) NaOH 100 mM durante 10 min a 95°C, y e) MnCl₂ 200 mM. ML, marcador de bajo peso molecular; MH, marcador de alto peso molecular; P, cebador restante; LL, fragmentos de ligador-ligador.

45 Figura 9: Comparación entre purificación SPRI y en columna de sílice y después de la síntesis de 2ª hebra. a) Purificación SPRI con perlas magnéticas modificadas con hidroxilo y un tampón de sal-PEG, y b) purificación en columna de sílice con un sistema de tampón de pH.

50 Figura 10: Secuencias de nucleótidos.

Ejemplos

Ejemplo 1: Generación de biblioteca NGS 3' terminal para una plataforma de secuenciación Illumina.

Brevemente, se lleva a cabo la generación de biblioteca principal tal como se describe en la figura 1. a) Se inicia la síntesis de ADNc mediante cebado en una región o etiqueta conocida que o bien está presente ya (en este caso, poli(A) de ARNm) o bien se une en una reacción anterior. P1 es complementaria a dicha región conocida y contiene además una secuencia específica no complementaria en su extremo 5' que sirve como etiqueta universal. b) ARN se

somete a transcripción inversa para dar ADNc por una polimerasa dependiente de ARN, es decir una transcriptasa inversa. c) Después de la síntesis de ADNc, se hidroliza el ARN molde o se degrada o bien por ARNasas, cambios de pH (NaOH y calor) o cationes divalentes (Mn^{2+} , Mg^{2+} y calor). d) Entonces, se ceba el ADNc monocatenario mediante múltiples cebadores al azar, P(n), P(n+1), ..., P(n+n), e) y se sintetiza una segunda hebra usando una polimerasa dependiente de ADN sin desplazamiento de hebra. La falta de desplazamiento de hebra garantiza que sólo el fragmento más en 3' contendrá ambas etiquetas, una del cebador de síntesis de ADNc y la otra del cebador de síntesis de 2ª hebra.

Se describen las etapas de reacción individuales en más detalle a continuación.

La generación de la biblioteca se inicia con la síntesis de ADNc de primera hebra a través de transcripción inversa en la que se hibrida un cebador de oligodT que contiene una secuencia compatible con Illumina en su extremo 5' al ARN después de que tenga lugar la transcripción inversa. Para este fin, para una preparación de biblioteca individual se mezclaron 5 µl de ARN con 5 µl de mezcla 1 de síntesis de ADNc de primera hebra que contenía todos los componentes necesarios para una transcripción inversa incluyendo un cebador de oligo dT sin la enzima en un pocillo de una placa para PCR, alternativamente en un pocillo de una tira de 8 pocillos o en cualquier otro tubo compatible con termociclador. Si se usa un menor volumen de ARN, se añade agua libre de ARNasa para obtener un volumen total de 10 µl. Entonces se mezcla bien la disolución mediante pipeteado y se sella la placa para PCR. Se aplica el sello de manera apretada. Se centrifuga la placa de modo que se acumula todo el líquido en la parte inferior de los pocillos. Entonces se desnaturaliza la mezcla de ARN/RT durante 3 min a 85°C en un termociclador y luego se enfría hasta 37°C para permitir la hibridación del cebador de RT a). Se centrifuga la placa para asegurarse de que se acumula todo el líquido en la parte inferior de los pocillos antes de retirar cuidadosamente la lámina de sellado.

Después de eso, se mezclan 10 µl de una de dilución transcriptasa inversa con cada reacción mediante pipeteado antes de sellarse la placa de nuevo. Es necesario centrifugar el líquido y en la etapa b) se incuba la placa a 37°C durante 15 minutos.

c) Se retira el ARN molde. Durante esta etapa se destruye el ARN molde que es esencial para la síntesis de segunda hebra eficaz. Antes de retirar la lámina de sellado después de la reacción de síntesis de primera hebra, se centrifuga rápidamente la placa para garantizar que se acumula todo el líquido en la parte inferior de los pocillos. Se añaden directamente 5 µl de una disolución de retirada de ARN a la reacción de síntesis de primera hebra y se mezcla bien y vuelve a sellarse la placa usando una nueva lámina. Debe incubarse la placa durante 10 minutos a 95°C antes de enfriarse hasta 25°C, y centrifugarse. Ahora, se retira cuidadosamente la lámina de sellado, se añaden 5 µl de disolución de retirada 2 (que neutraliza o retira básicamente el componente al que se le añadió la disolución de retirada 1) y se mezcla la disolución bien de nuevo.

d) Durante la siguiente síntesis de segunda hebra, se convierte la biblioteca en ADNbc. Se inicia la síntesis de segunda hebra mediante un cebador al azar que contiene una secuencia de ligador compatible con Illumina en su extremo 5'. Un complemento inverso impide que la secuencia de ligador participe en la hibridación. En esta fase, se recomienda llevar las perlas de purificación (perlas de SPRI) hasta temperatura ambiente para proporcionarles tiempo suficiente para equilibrarse. Se añaden 15 µl de mezcla de síntesis 1 de segunda hebra (que contiene todos los componentes necesarios para una reacción de polimerización dependiente de ADN), se mezclan bien mediante pipeteado y se sella la placa. Ahora, se incuba la placa durante 1 min a 98°C en un termociclador y se enfría lentamente hasta 25°C fijando la velocidad de rampa en 0,5°C por segundo que corresponde a muchos termocicladores al 10% de la velocidad de rampa máxima. Se incuba la reacción durante 30 min a 25°C, se centrifuga rápidamente antes de retirar la lámina de sellado de la placa.

e) Se añaden 5 µl de una dilución de polimerasa dependiente de ADN. Se incuba la reacción a 25°C durante 15 minutos. Hasta la etapa e) se ha llevado a cabo toda la reacción en un volumen creciente sucesivamente. La reacción continúa entrando en la etapa de selección f).

f) La biblioteca bicatenaria que contiene todavía dobles hebras no deseadas que no contienen la secuencia P1 se purifica usando perlas magnéticas. Las perlas de purificación (PB) deben haberse equilibrado durante 30 min a temperatura ambiente antes de su uso. PB pueden haberse sedimentado y resuspendido apropiadamente antes de añadirse a la reacción. Después de eso, se lleva a cabo una purificación SPRI según las instrucciones del fabricante (AMPure Beads; Agentcourt). Se eluyen las bibliotecas en 20 µl de agua o Tris 10 mM, pH 8,0 y se transfieren 17 µl del sobrenadante transparente con la biblioteca a una nueva placa para PCR. Debe tenerse cuidado de no transferir ninguna perla a la nueva placa. Pueden almacenarse las bibliotecas a -20°C para su posterior amplificación.

Se aíslan los fragmentos más en 3' a través de amplificación por PCR. También se amplifica la biblioteca para añadir las secuencias de adaptador completas requeridas para la generación de agrupaciones en máquinas NGS y para generar suficiente material para control de calidad y las posteriores mezclas de carriles. Se lleva a cabo una reacción PCR convencional usando una ADN polimerasa termoestable y después de eso, se purifican de nuevo los productos mediante una purificación final (purificación SPPRI según las instrucciones del fabricante) en la que se aísla la biblioteca terminada de cualquier componente de PCR restante y en la que el material de ADN de entrada que no contenía ni la secuencia P1 ni Pn se desplaza en la representación global a través del proceso de dilución

relativa de la PCR. Las secuencias remanente que no contienen ambas secuencias P1 y Pn no podrán generar agrupaciones en el proceso de NGS porque la generación de agrupaciones usa una amplificación por PCR que parte de moléculas individuales que sólo contienen ambas secuencias. Se eluyen las bibliotecas finales en 20 µl de EB se añaden, y se resuspenden las perlas apropiadamente en EB antes de incubar durante 2 minutos a temperatura ambiente. Se coloca la placa sobre una placa magnética para acumular las perlas, se acumulan durante 5 minutos o hasta que el sobrenadante sea transparente por completo. Se transfiere el sobrenadante a una nueva placa para PCR. En este punto, se terminan las bibliotecas y están listas para control de calidad, combinación, generación de agrupaciones y secuenciación adicional.

Ejemplo 2: Comparación de polimerasas sin detención con actividad de desplazamiento de hebra (Klenow) y polimerasas con detención sin actividad de desplazamiento de hebra (T4 y T7) a diferentes temperaturas (figuras 2, 3 y 4).

Descripción del ensayo:

Se muestra una representación esquemática de la configuración del ensayo y se describe en la figura 2. Se evaluaron diferentes polimerasas con desplazamiento de hebra, con y sin capacidad para destruir la hebra desplazada y polimerasas sin desplazamiento de hebra usando el ensayo descrito en la figura 2. Brevemente, se hibridaron Seq ID: 1, Seq ID: 2 y Seq ID: 3 en el tampón correspondiente de las diferentes polimerasas. Después de la hibridación, se añadieron polimerasas (3 U de ADN polimerasa de T4, 10 U de ADN polimerasa de T7 o 5 U de fragmento Klenow (3'-5' exo-), y se realizó la reacción tal como se indica en la figura 3 y la figura 4. El tiempo de reacción fue de 10 minutos a la temperatura indicada. Después de eso, se purificaron las reacciones mediante columnas de sílice para retirar componentes de tampón y enzimas sin ninguna selección de tamaño. Se cargaron las muestras sobre un gel de PAA al 10% (mezclado con colorante de carga y desnaturalizado durante 2 minutos a 95°C) y se ejecutó a 100 V durante 10 minutos y entonces a 180 V durante 120 minutos a 58°C. Se tiñeron los gels con GYBR Gold.

En la figura 3, se demuestra la importancia de una etapa de desnaturalización y apareamiento lento (desde 95°C hasta la temperatura de reacción con una rampa más lenta (lleva 15 min) como para polimerasas sin desplazamiento de hebra, las estructuras secundarias en el molde representan un obstáculo importante. Las flechas rellenas blancas indican productos de detención de desplazamiento de hebra parcialmente desplazados. Las estructuras secundarias del molde monocatenario sin la etapa de desnaturalización se indican con una flecha negra. Klenow siempre muestra desplazamiento de hebra (especialmente a mayores temperaturas y, en este caso, también una etapa de desnaturalización para impedir las estructuras secundarias en el ADNc no es esencial ya que la enzima puede disolver esas estructuras secundarias con su desplazamiento de hebra inherente (figura 2). La figura 4 demuestra el desplazamiento de hebra de cinco polimerasas diferentes. Se realizó el ensayo tal como se describió anteriormente con los tampones correspondientes, una etapa de desnaturalización inicial y a temperaturas de reacción tal como se indica en la figura 4. Polimerasas con desplazamiento de hebra: 8 unidades de ADN polimerasa de Bst, fragmento grande, carril: 2; 5 unidades de ADN polimerasa I, fragmento grande (Klenow), carriles 7-8; y 200 unidades de VLM-M, carril 11; y polimerasas sin desplazamiento de hebra: carril 3, 3 unidades de ADN polimerasa de T4, carriles 4-6, o una polimerasa que destruye la hebra desplazada tal como 5 unidades de ADN polimerasa de Bst, de longitud completa, carril 3, 10 unidades de ADN polimerasa I de *E. coli*, carriles 9-10. Se comprobaron diferentes temperaturas de reacción: 12°C: carril: 4; 25°C: carriles: 5, 7, 9; 37°C: carriles: 2-3, 6, 8, 10-11). La ADN polimerasa de T4 contiene una actividad exonucleasa que se vuelve prominente a 37°C. Son visibles productos de degradación de Seq ID: 1, Seq ID: 5, Seq ID: 2 y Seq ID: 3. Con ADN polimerasa de Bst, fragmento grande, ADN polimerasa I, fragmento grande (Klenow) y VLM-M hay un desplazamiento de hebra y es visible Seq ID 6 (producto de longitud completa). Adicionalmente, es visible un producto mayor que Seq ID: 5 con aquellas polimerasas resultantes de un desplazamiento de hebra parcial. Con otras polimerasas de desplazamiento de hebra, tal ADN polimerasa de Bst, de longitud completa, (carril 3) y ADN polimerasa I de *E. coli* (carril 9-10) también se destruye la hebra desplazada. La ADN polimerasa de T4 no contiene desplazamiento de hebra y Seq ID 5 es claramente visible. 25°C para T4 es más recomendable que 37°C porque a 37°C la exonucleasa inherente toma el control de la reacción (figura 4).

Ejemplo 3: Degradación de ARN con MnCl₂, temperatura elevada sólo, tratamiento con NaOH o ARNasas (figura 5) la degradación de ARN también depende de las condiciones de tampón.

Se añadieron cantidades conocidas del ARN total aislado de hígado de ratón a un oligo de ADN de 111 nt monocatenario (ADNmc) (ID Seq ID: 7) véanse el carril 1 y el carril 10. El ARN total es largo, así sólo son visibles bandas de ARN más pequeñas en el gel, los fragmentos de ARN largos permanecen en la ranura. Tras la fragmentación, se degradan los fragmentos de ARN más largos y son visibles como una mancha en el gel de poliacrilamida.

El tratamiento con calor en un tampón para RT convencional, Tris-HCl 50 mM (pH 8,3 a 25°C), KCl 75 mM, MgCl₂ 3 mM y DTT 10 mM durante 30 minutos a 95°C y durante 5 minutos a 98°C, 10 minutos a 98°C, 20 minutos a 98°C, y 30 minutos a 98°C, da como resultado la degradación del ARN, pero no una retirada completa del ARN. La incubación de la mezcla de ARN/ADNmc con la mezcla de ARNasa H/ A / T1 durante 30 minutos o bien a 25°C o bien a 37°C retira por completo el ARN sin degradar el ADN monocatenario. La incubación a temperaturas elevadas

durante 10 minutos (55°C) en presencia de NaOH 0,1 N degrada el ARN, aunque no por completo. Después de 10 minutos 95°C en NaOH 0,1 N se retira por completo el ARN sin embargo también comienza a degradarse el ADNmc (carriles 7-10). La adición de MnCl₂ 10 mM a la mezcla de ARN/ADNmc/tampón para RT y el tratamiento con calor durante 5, 10, 20 y 30 minutos a 98°C da como resultado la degradación completa del ARN sin degradar el ADNmc.

5 Se cargaron las muestras en un gel PAA de al 10% gel sin purificación (mezclado con colorante de carga y desnaturizado durante 2 minutos a 95°C) y se ejecutó a 100 V durante 10 minutos y entonces a 180 V durante 70 minutos a 58°C. Se tiñeron los geles con GYBR Gold.

Ejemplo 4: Influencia del método de degradación de ARN sobre la calidad de las bibliotecas NGS sintetizadas con una polimerasa sin desplazamiento de hebra (figura 6).

10 Se mezclaron 500 ng de ARN total con Seq ID: 8 (concentración final en 20 µl: 25 nM) y se calentó hasta 85°C durante 3 minutos en un volumen de 10 µl que contenía 4 µl de tampón de RT 5x. Después de enfriar hasta 37°C, se añadieron 1 µl de dNTP 1 mM, 200 unidades de VLM-M y se incubó durante 15 minutos a 37°C. La posterior hidrólisis de ARN difiere tal como se muestra en la figura 6. Se llevó a cabo una captura con biotina-estreptavidina en el tampón de reacción de RT durante 20 min (en un agitador a 1250 rpm a 25°C) usando 5 µg de perlas de estreptavidina de NEB. Se lavaron las perlas 2 veces con tampón de lavado, y se liberaron las muestras de las perlas mediante calentamiento hasta 80°C durante minutos en 10 µl de MB-H₂O. Se recogieron las perlas usando un imán y se transfirió el sobrenadante transparente a un tubo diferente en el que se llevó a cabo la síntesis de ADNc de segunda hebra usando 3 unidades de ADN polimerasa de T4, Seq ID: 9 y 10 (concentración final en 20 µl de 0,1 mM), PEG al 8%, MgCl₂ 10 mM y dNTP 0,5 mM en un volumen total de 20 µl. Antes de añadir la polimerasa, se incluyó una etapa de desnaturalización a 98°C, 1 minuto y un apareamiento lento (rampa descendente hasta 25°C en el plazo de 15 minutos). Entonces se purificaron en sílice las muestras según la guía de usuario de SENSE (purificación de sección después de síntesis de segunda hebra) y se eluyeron en 25 µl de Tris 10 mM pH 8. Se añadieron 10 µl del producto purificado y luego se amplificó durante 18 ciclos según la Seq-PCR de ARNm SENSE usando Seq ID: 11 y 12 como cebadores de PCR. Entonces se purificaron en sílice las muestras según la guía de usuario de SENSE (purificación de sección después de síntesis de segunda hebra) y se eluyeron en 15 µl de Tris 10 mM pH8. Se cargó 1 µl de producto de PCR purificado, en chip de ADN de alta sensibilidad (Agilent) según las instrucciones del fabricante.

La degradación con ARNasas y MnCl₂ del ARN da como resultado un rendimiento mucho mayor que la hidrólisis con NaOH que o bien daña el ADNc o bien da como resultado modificaciones de bases que hacen que no pueda amplificarse el ADNc.

Ejemplo 5: La fragmentación inicial del ARN en el tampón de RT determina el tamaño de biblioteca y la eficacia del protocolo (figura 6).

El volumen en el que se desnaturaliza el ARN también tiene influencia sobre la concentración de MgCl₂ que está presente durante la desnaturalización y esto también determina cómo será de largo el ADNc que se genera a partir del ARN ligeramente fragmentado

Se mezclaron 500 ng de ARN total con Seq ID: 8 (concentración final en 20 µl: 25 nM) y se calentó hasta 85°C durante 3 minutos en un volumen de 10 µl (a y b) o 15 µl (c y d) que contenía 4 µl de tampón de RT 5x. Después de enfriar hasta 37°C, se añadieron 1 µl de dNTP 1 mM, 200 unidades de VLM-M y se incubó durante 15 minutos a 37°C. Se hidrolizó el ARN en presencia de MnCl₂ 10 mM mediante calentamiento hasta 98°C durante 10 minutos. Después de eso, se añadió EDTA 10 mM. Se realizó la síntesis de segunda hebra mediante la adición de los componentes de síntesis de segunda hebra a la reacción resultante en una concentración final de MgCl₂ 10 mM, dNTP 0,5 mM, PEG al 8%, SEQ ID: 9 y 10 cada uno a una concentración final de 100 nM y 3 unidades de ADN polimerasa de T4 en un volumen total de 40 µl. De nuevo, se añadió la polimerasa después de calentamiento hasta 98°C durante 1 minuto y un apareamiento lento (rampa descendente hasta 25°C en el plazo 15 minutos. Se llevó a cabo purificación, amplificación por PCR y posterior purificación por PCR y la ejecución del bioanalizador tal como se describe en el ejemplo 4.

Ejemplo 6: Ajuste del rendimiento y tamaño de biblioteca mediante la concentración de cebador de RT y concentración de oligo de síntesis de 2ª hebra (figura 7).

Todas las bibliotecas 17 ciclos. Todas las condiciones de reacción según el ejemplo 5 concentraciones variables de a) Seq ID: 8 50 nM durante la etapa de transcripción inversa (RT) y Seq ID: 9 y 10 0,1 mM, en síntesis de segunda hebra (SSS) b) Seq ID: 8 25 nM durante RT y Seq ID: 9 y 10 0,5 mM durante SSS, biblioteca diluida 1:3 antes de la carga, c) Seq ID: 8 50 nM durante RT y Seq ID: 9 y 10 0,5 mM durante SSS, biblioteca diluida 1:3 antes de la carga, d) Seq ID: 8 25 nM durante RT y Seq ID: 9 y 10 0,1 mM durante SSS.

Ejemplo 7: Purificación en sílice frente a SPRI (figura 8).

55 Todas las etapas tal como se describen en el ejemplo 6 (c) pero con diferente purificación. Se realizaron purificaciones en sílice tal como se describe en el ejemplo 5 y purificación SPRI con perlas magnéticas modificadas con hidroxilo y un tampón de sal-PEG según las instrucciones del fabricante (AMPure XP Beads de Agentcourt)

instrucciones.

Bibliografía

Bodescot M. *et al* (1994) ADN and Cell Biology 13(9) :977-985

Costa V, *et al.* (2010) J Biomed Biotech 7(7): 1299-20.

5 Derti A, *et al.* (2012) Genome Res. 22(6): 1173-83.

Fox-Walsh K, *et al.* (2011) Genomics. 98(4): 266-71.

Hawkins TL, *et. al.* (1995) Nucleic Acids Res. 23: 4742-4743.

Mainul Hoque *et al.*, (2012) Nature Methods 10(2): 133-139.

Shepard PJ, *et al.* (2011) RNA 17: 761-772.

10 Wendl MC y Wilson RK (2009) BMC Genomics 10: article 485.

Wilkening S, *et al.* (2013) Nucleic Acids Res. 41(5): e65.

Documento WO 98/044151

Documento WO 02/059357

Documento US 5705628

15 Documento US 2011/0160078

Documento US 6406891 B1

Documento EP 1371726 A1

Documento WO 2013/038010 A2

Lista de secuencias

20 <110> Lexogen GmbH

<120> método de análisis de ARN de conservación del número de copias

<130> r67217

<150> Documento EP 14161135.0

<151> 21.03.2014

25 <160> 12

<170> PatentIn versión 3.5

<210> 1

<211> 145

<212> ADN

30 <213> Secuencia artificial

<220>

<223> constructo modelo

<400> 1

actgtaaaac gacggccagt atagttcgct cggtgaccct ggtctttctg gtgcttgtct 60

cactgaccgg cctgtatgct atccagagtg agtgcctctt gtcaagaaca tatagatagg 120

ctccgttgct aatctgaagg tcgaa 145

| | | |
|----|--|-----------|
| | <210> 2 | |
| | <211> 51 | |
| | <212> ADN | |
| | <213> Secuencia artificial | |
| 5 | <220> | |
| | <223> constructo modelo | |
| | <400> 2 | |
| | gtccagctac ctaggctcatg tactacggag cctatctata tgttcttgac a | 51 |
| | <210> 3 | |
| 10 | <211> 35 | |
| | <212> ADN | |
| | <213> Secuencia artificial | |
| | <220> | |
| | <223> constructo modelo | |
| 15 | <400> 3 | |
| | cagtgagaca agcaccagaa agaccagggt caccg | 35 |
| | <210> 4 | |
| | <211> 65 | |
| | <212> ADN | |
| 20 | <213> Secuencia artificial | |
| | <220> | |
| | <223> constructo modelo | |
| | <400> 4 | |
| | cagtgagaca agcaccagaa agaccagggt caccgagcga actatactgg ccgtcgtttt | 60 |
| | acagt | 65 |
| 25 | <210> 5 | |
| | <211> 85 | |
| | <212> ADN | |
| | <213> Secuencia artificial | |
| | <220> | |
| 30 | <223> constructo modelo | |
| | <400> 5 | |
| | gtccagctac ctaggctcatg tactacggag cctatctata tgttcttgac aagaggcaact | 60 |
| | cactctggat agcatacagg ccggt | 85 |
| | <210> 6 | |

ES 2 645 423 T3

<211> 150
 <212> ADN
 <213> Secuencia artificial
 <220>
 5 <223> constructo modelo
 <400> 6
 gtccagctac ctaggctcatg tactacggag cctatctata tgttcttgac aagaggcact 60
 cactctggat agcatacagg ccggtcagtg agacaagcac cagaaagacc agggtcaccg 120
 agcgaactat actggccgtc gttttacagt 150
 <210> 7
 <211> 111
 10 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> constructo modelo
 <400> 7
 gttgtcacca gcatccctag acccgtacag tgcccactcc ccttcccagt ttccgactgt 60
 ccccggcctc ctgcgggcca ttccaaaaaa aaaaaaaaaa aaaaaaaaaa a 111
 15 <210> 8
 <211> 44
 <212> ADN
 <213> Secuencia artificial
 20 <220>
 <223> constructo modelo
 <400> 8
 tccctacacg acgctcttcc gatctttttt tttttttttt tttv 44
 <210> 9
 25 <211> 24
 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> constructo modelo
 30 <220>
 <221> misc_feature
 <222> (24) .. (24)

<223> dt invertido
 <400> 9
agatcgggaag agcacacgtc tgat 24
 <210> 10
 5 <211> 36
 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> constructo modelo
 10 <220>
 <221> misc_feature
 <222> (25) .. (36)
 <223> n es a, e, g o t
 <400> 10
 15 **ttcagacgtg tgctcttccg atctnnnnnn nnnnnn** 36
 <210> 11
 <211> 58
 <212> ADN
 <213> Secuencia artificial
 20 <220>
 <223> constructo modelo
 <400> 11
aatgatacgg cgaccaccga gatctacact ctttcctac acgacgctct tccgatct 58
 <210> 12
 25 <211> 58
 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> constructo modelo
 30 <400> 12
caagcagaag acggcatacag agatgtgact ggagttcaga cgtgtgctct tccgatct 58

REIVINDICACIONES

1. Un método para generar un producto de ácido nucleico a partir de una molécula de ARN molde, que comprende después de haber obtenido un ARN molde
 - 5 a) aparear un primer cebador oligonucleotídico en una región de ácido nucleico preseleccionada del ARN molde,
 - b) elongar el primer cebador oligonucleotídico de manera específica al molde obteniendo de ese modo una primera hebra elongada,
 - c) retirar el ARN molde,
 - d) aparear más de un cebador oligonucleotídico adicionales con la primera hebra elongada,
 - 10 e) elongar los más de un cebador oligonucleotídico adicionales de manera específica al molde sin desplazamiento de los cebadores apareados con la primera hebra elongada usando una polimerasa que carece de actividad de desplazamiento de hebra, preferiblemente una ADN polimerasa T7, T4 o Q5, y/o usando cebadores que tienen resistencia al desplazamiento de hebra por una polimerasa, preferiblemente cebadores que tienen nucleótidos con modificaciones de LNA o 2'-fluor, generando de ese modo productos de elongación adicionales,
 - 15 en el que las etapas a) a e) se realizan en un volumen de fluido creciente consecutivamente,
 - f) aislar y/o amplificar un producto de elongación de dicho producto de elongación adicional que comprende una parte de ácido nucleico que se elonga complementaria al primer cebador oligonucleotídico.
2. El método según la reivindicación 1, en el que la región de ácido nucleico preseleccionada es una región de ácido nucleico 3'-terminal, que comprende preferiblemente una cola de poli-A.
3. El método según la reivindicación 1 ó 2, en el que el método comprende además la etapa de unir una cola de polinucleótido en 3' al extremo 3' del ARN molde en el que dicha región de ácido nucleico 3'-terminal preseleccionada comprende dicha cola de polinucleótido en 3'.
4. El método según una cualquiera de las reivindicaciones 1 a 3, en el que el primer cebador oligonucleotídico y/o los cebadores oligonucleotídicos adicionales es/son ADN.
5. El método según una cualquiera de las reivindicaciones 1 a 4, en el que el primer cebador oligonucleotídico y/o los cebadores oligonucleotídicos adicionales contienen una etiqueta de secuencia o una secuencia de ligador sin apareamiento, que se usa preferiblemente para la unión de cebador de amplificación.
6. El método según la reivindicación 5, en el que la etiqueta de secuencia o la secuencia de ligador sin apareamiento contiene un código de barras, preferiblemente un código de barras aleatorio.
7. El método según una cualquiera de las reivindicaciones 1 a 6, en el que b) elongar el primer cebador oligonucleotídico de manera específica al molde es mediante transcripción inversa y la primera hebra elongada es una hebra de ADN.
8. El método según una cualquiera de las reivindicaciones 1 a 7, en el que c) retirar el ARN molde comprende digestión enzimática de ARN, preferiblemente por una ARNasa, degradación alcalina, preferiblemente mediante tratamiento con NaOH, o calentamiento en presencia de cationes divalentes, preferiblemente Mn^{2+} o Mg^{2+} .
9. El método según una cualquiera de las reivindicaciones 1 a 8, en el que los más de un cebador oligonucleotídico adicionales comprenden cebadores al azar y/o al menos 10 cebadores diferentes, preferiblemente al menos 20, de manera especialmente preferida al menos 100.
10. El método según una cualquiera de las reivindicaciones 1 a 9, en el que la región de ácido nucleico preseleccionada está presente en uno o múltiples ARN moldes de interés.
11. El método según una cualquiera de las reivindicaciones 1 a 10, en el que los más de un cebador oligonucleotídico adicionales son cada uno específicos para un ARN molde o secuencia génica en el mismo.
12. El método según una cualquiera de las reivindicaciones 1 a 11, en el que los más de un cebador oligonucleotídico adicional se aparean con regiones específicas de uno o múltiples ARN de interés.
13. El método según una cualquiera de las reivindicaciones 1 a 12, en el que e) elongar los más de un cebador oligonucleotídico adicionales de manera específica al molde se realiza en presencia de un agente de aglomeración, preferiblemente PEG.

14. El método según una cualquiera de las reivindicaciones 1 a 13, en el que se fragmenta el ARN molde antes del apareamiento del primer cebador.
15. El método según una cualquiera de las reivindicaciones 1 a 14, que comprende además una etapa de purificar el producto de elongación de la etapa e).
- 5 16. El método según una cualquiera de las reivindicaciones 1 a 15, que comprende realizar una PCR con el producto de elongación adicional usando un cebador específico para etiquetas de secuencia o secuencias de ligador de dicho producto de elongación.
17. El método según la reivindicación 16, en el que al menos un cebador de la PCR comprende una etiqueta de secuencia o secuencia de ligador adicional.
- 10 18. Uso de un kit para realizar un método según una cualquiera de las reivindicaciones 1 a 17, en el que el kit comprende una transcriptasa inversa, dNTP, cofactores o sales de iones metálicos requeridos por una polimerasa, preferiblemente Mg^{2+} , un cebador, una ADN polimerasa sin actividad de desplazamiento de hebra, tal como ADN polimerasa T7, Q5 o T4, y cebadores oligonucleotídicos al azar.
- 15 19. El uso según la reivindicación 18, en el que el kit comprende además un agente de degradación de ARN y/o un agente de aglomeración tal como PEG.

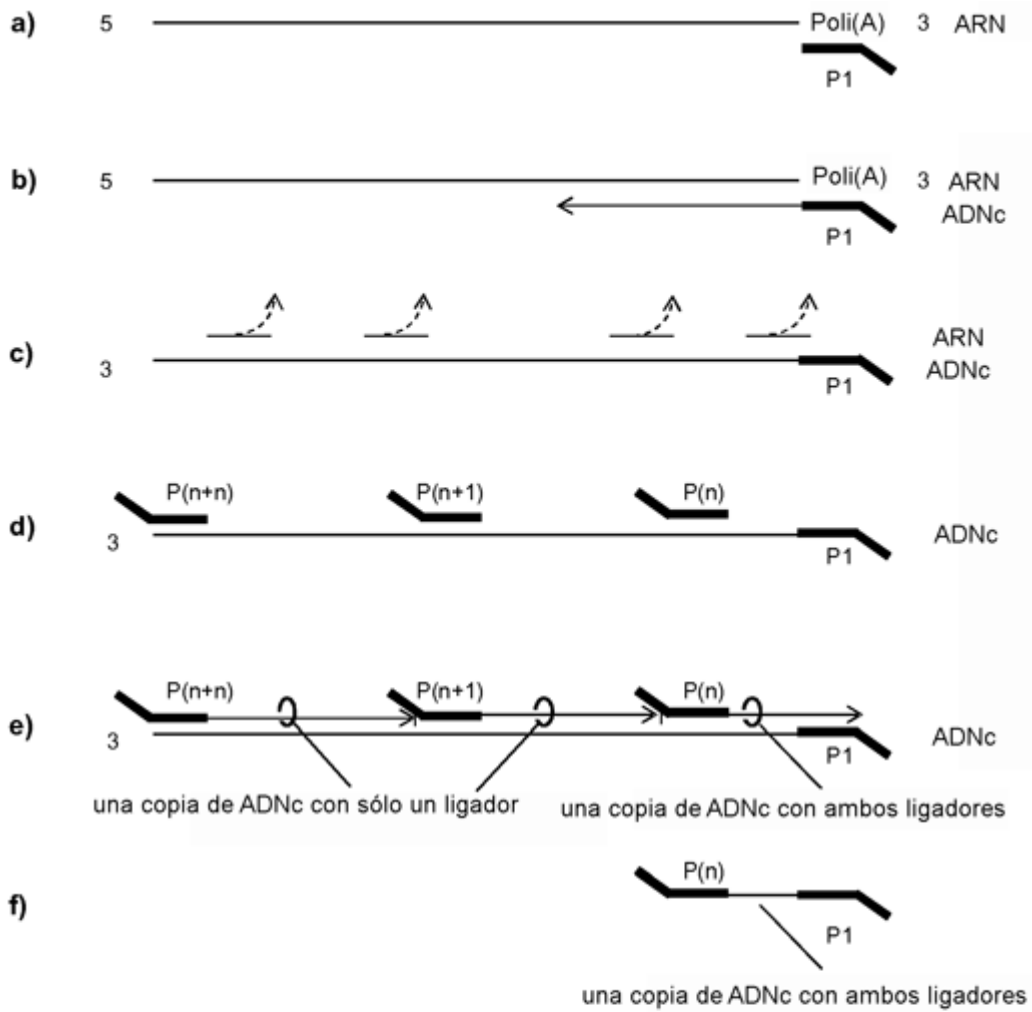


Figura 1

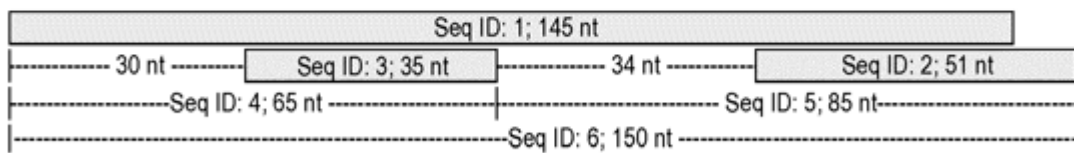


Figura 2

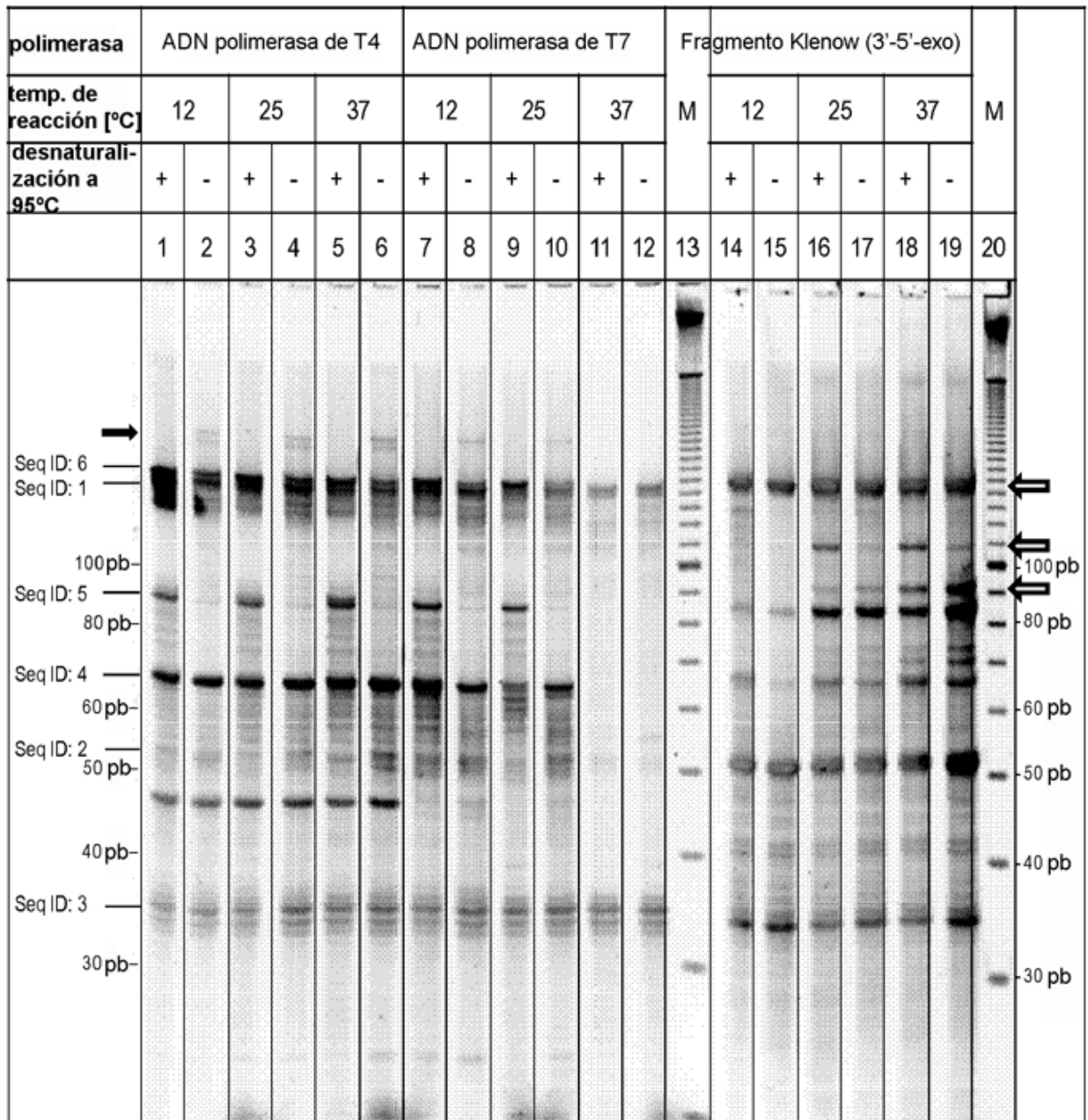


Figura 3

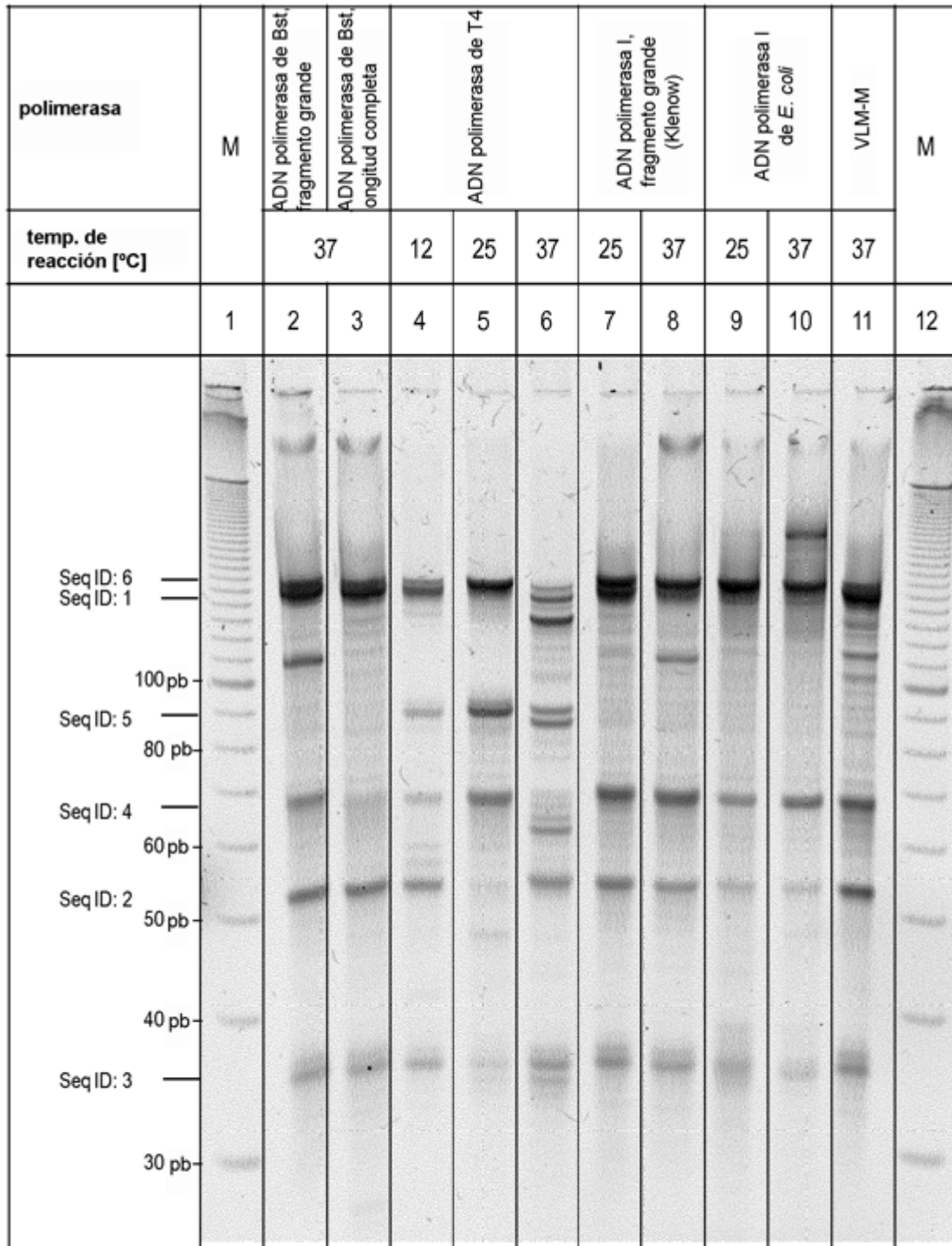


Figura 4

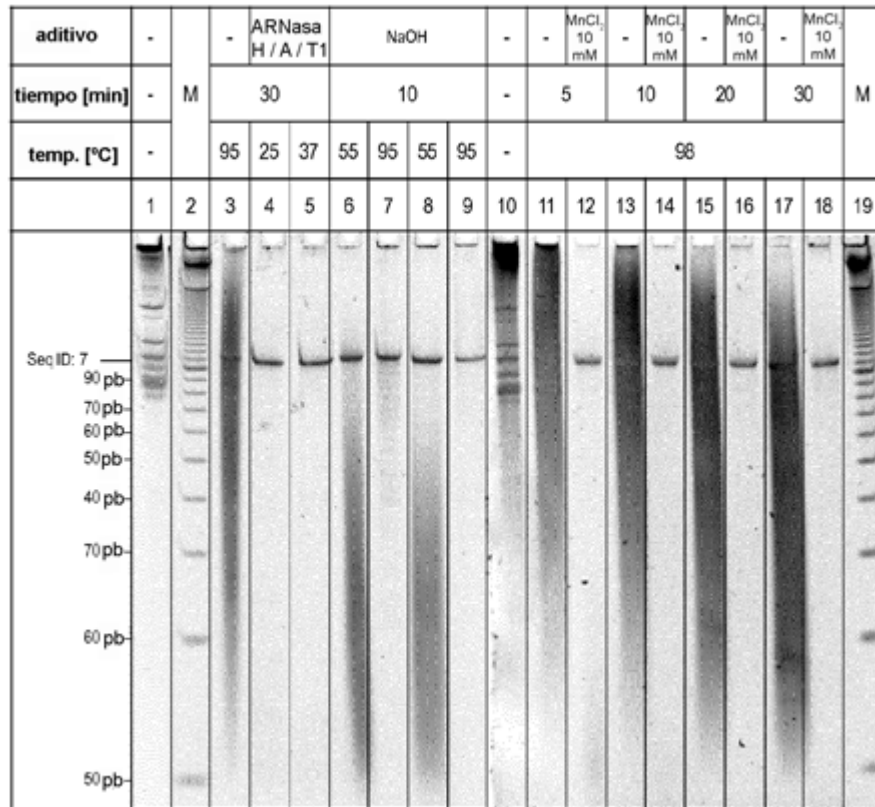


Figura 5

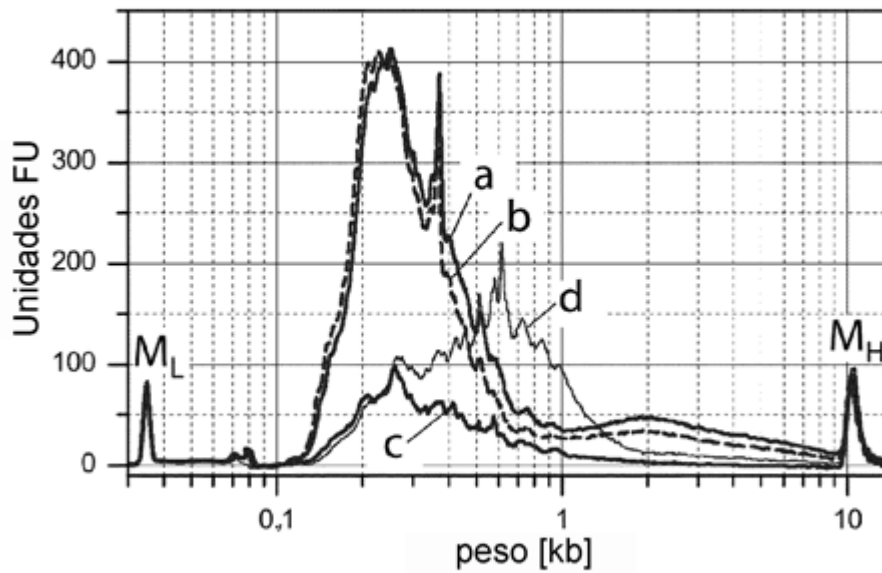


Figura 6

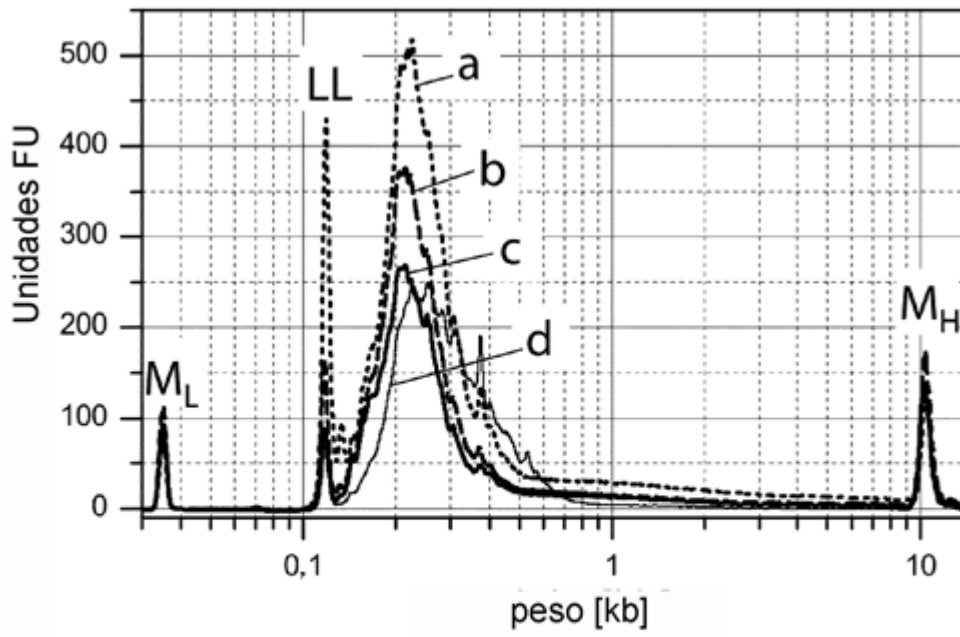


Figura 7

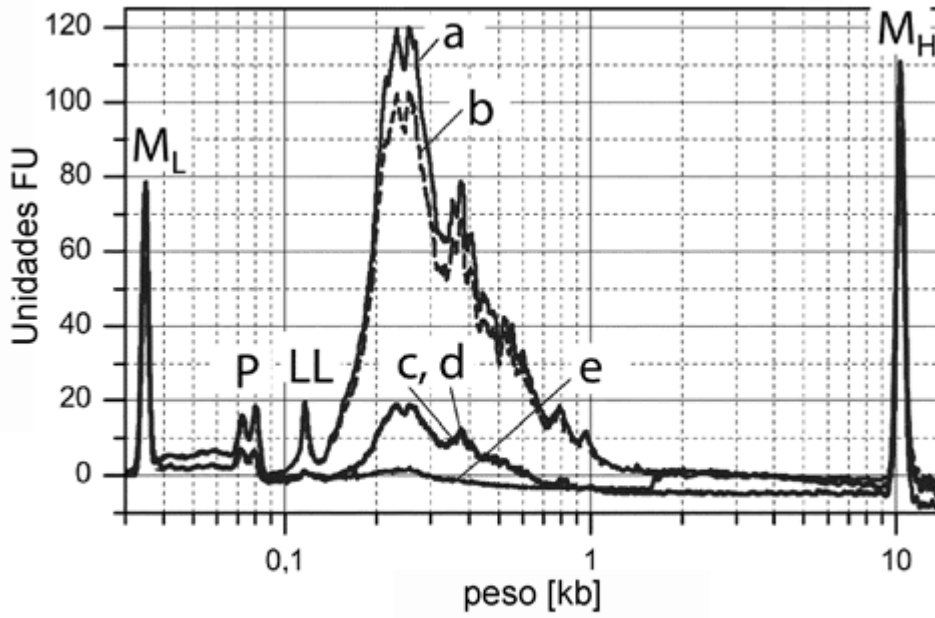


Figura 8

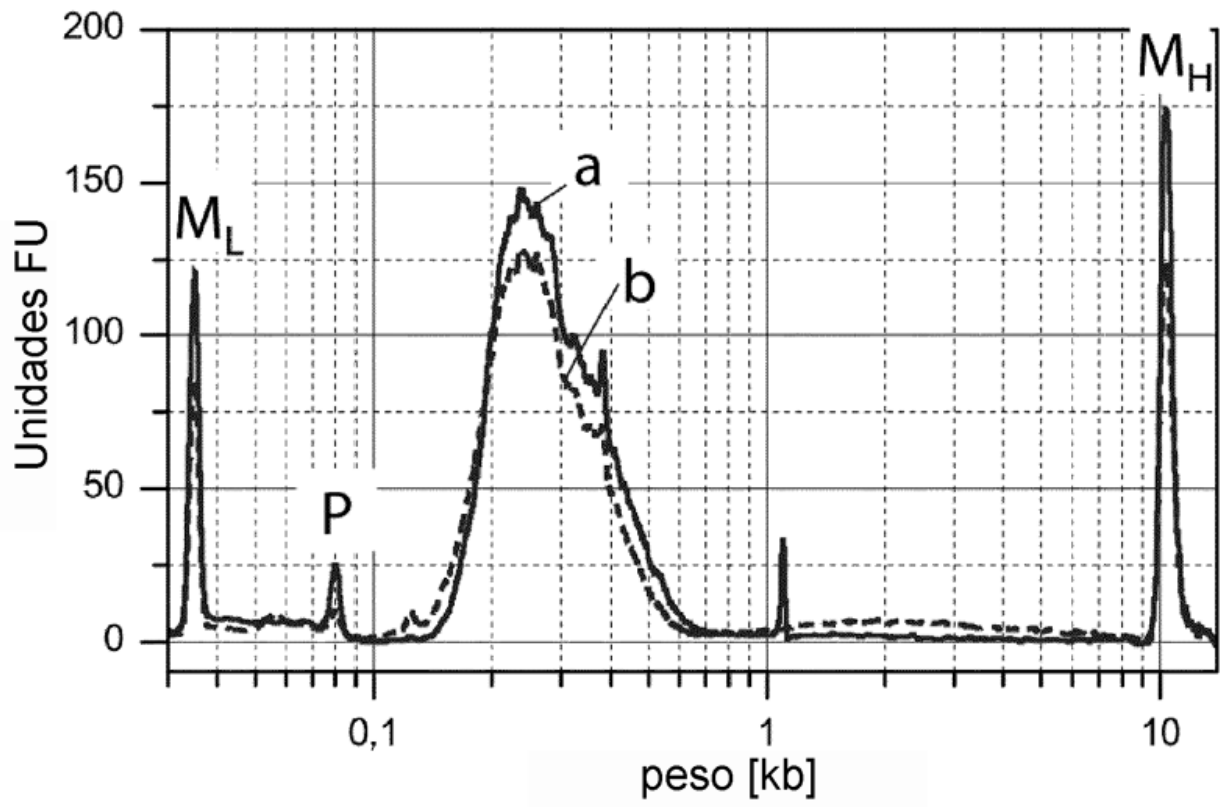


Figura 9

| SEQ ID: | Secuencia | nt | Tipo |
|---------|--|-----|------|
| 1 | A*C*T*GTAAAACGACGGCCAGTATAGTTC GCTCGGTGACCCTGGTCTTTCTGGTGCTTG TCTCACTGACCGCCTGTATGCTATCCAGA GTGAGTGCCTCTTGTCAAGAACATATAGAT AGGCTCCGTTGCTAATCTGAAGGTCGAA | 145 | |
| 2 | G*T*C*CAGCTACCTAGGTCATGTACTIONG GAGCCTATCTATATGTTCTTGACA | 51 | |
| 3 | /5Phos/CAGTGAGACAAGCACCAGAAAGACCAGGGT CA*C*C*G | 35 | |
| 4 | CAGTGAGACAAGCACCAGAAAGACCAGGGT CA*C*C*GAGCGAACTATACTGGCCGTCGTTTTACAGT | 65 | |
| 5 | G*T*C*CAGCTACCTAGGTCATGTACTIONG CCTATCTATATGTTCTTGACAAGAGGCACT CACTCTGGATAGCATAACAGGCCGGT | 85 | |
| 6 | G*T*C*CAGCTACCTAGGTCATGTACTIONG CCTATCTATATGTTCTTGACAAGAGGCACT CACTCTGGATAGCATAACAGGCCGGTCAGTG AGACAAGCACCAGAAAGACCAGGGTCACCG AGCGAACTATACTGGCCGTCGTTTTACAGT | 150 | |
| 7 | GTTGTCACCAGCATCCCTAGACCCGTACAG TGCCCACTCCCCTTCCAGTTTCCGACTGT CCCCGGCCTCCTGCGGCCATTCC AAAAAAAAAAAAAAAAAAAAAAAAAAAA | 111 | |
| 8 | /5Biosg/TCCCTACACGACGCTCTTCCGATC TTTTTTTTTTTTTTTTTTTTV | 44 | |
| 9 | AGATCGGAAGAGCACACGTCTGA/3InvdT/ | 24 | |
| 10 | TTCAGACGTGTGCTCTTCCGATCTNNNNNNNNNN | 36 | |
| 11 | AATGATACGGCGACCACCGAGATCTACACT CTTCCCTACACGACGCTCTTCCGATCT | 58 | |
| 12 | CAAGCAGAAGACGGCATAACGAGATGTGACT GGAGTTCAGACGTGTGCTCTTCCGATCT | 58 | |

Figura 10