

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 909 651**

21 Número de solicitud: 202031119

51 Int. Cl.:

C12Q 1/6853 (2008.01)

12

PATENTE DE INVENCION CON EXAMEN

B2

22 Fecha de presentación:

06.11.2020

43 Fecha de publicación de la solicitud:

09.05.2022

Fecha de concesión:

04.12.2023

45 Fecha de publicación de la concesión:

13.12.2023

73 Titular/es:

**UNIVERSIDAD DE GRANADA (55.0%)
Hospital Real. Avda. del Hospicio s/n
18071 Granada (Granada) ES;
SERVICIO ANDALUZ DE SALUD (35.0%) y
LÓPEZ-VIOTA GALLARDO, Julián (10.0%)**

72 Inventor/es:

**LÓPEZ-VIOTA GALLARDO, Julián;
CARAZO GALLEG0, Ángel;
REDRUELLO ROMERO, Anaïs;
LÓPEZ PÉREZ, David;
LEÓN LÓPEZ, Josefa;
MORALES SANTANA, Sonia y
VERBENI, Michela**

54 Título: **Método de fabricación de librerías de miRNAs para secuenciación paralela masiva**

57 Resumen:

Método de fabricación de librerías de mirnas para secuenciación paralela masiva.

La presente invención se refiere a un procedimiento para la fabricación de librerías de mirnas para secuenciación paralela masiva aplicando nanotecnología que permite reducir los sesgos e incrementar la eficiencia.

ES 2 909 651 B2

Aviso: Se puede realizar consulta prevista por el art. 41 LP 24/2015.
Dentro de los seis meses siguientes a la publicación de la concesión en el Boletín Oficial de la Propiedad Industrial cualquier persona podrá oponerse a la concesión. La oposición deberá dirigirse a la OEPM en escrito motivado y previo pago de la tasa correspondiente (art. 43 LP 24/2015).

DESCRIPCIÓN

MÉTODO DE FABRICACIÓN DE LIBRERÍAS DE miRNAs PARA SECUENCIACIÓN PARALELA MASIVA

5 La presente invención se encuentra dentro del campo de la biología molecular y la nanotecnología, y se refiere a un procedimiento para la fabricación de librerías de miRNAs para secuenciación paralela masiva, aplicando nanotecnología para la reducción de sesgos, el incremento de la eficiencia y la reducción de costes.

10 **ESTADO DE LA TÉCNICA ANTERIOR**

Las células eucariotas (y algunos virus) producen pequeñas moléculas de RNA no codificante (de 19 a 25 nucleótidos en sus formas maduras) que regulan la expresión de numerosos genes. En humanos, se ha estimado que el 30% del genoma está regulado por micro-RNAs.

15 Principalmente, actúan sobre RNAs mensajeros del citoplasma, reconociendo secuencias específicas de las UTRs (*untranslated regions*) a través de las cuales reducen la frecuencia de traducción y la vida media del mensajero. El funcionamiento molecular, de cierta complejidad, requiere la implicación de estructuras proteicas del citoplasma (como el RISC, *RNA-induced silencing complex*).

20 Los miRNAs tienen un papel relevante en procesos tan importantes como la proliferación celular, la apoptosis, la diferenciación o el metabolismo energético. La biogénesis de los miRNAs está sometida a un estricto control espacial y temporal. Su desregulación se asocia a la mayoría de los procesos patológicos crónicos humanos (incluyendo cáncer diabetes, 25 disfunción endotelial y enfermedades neurodegenerativas). La estabilidad del miRNA en sangre y su relevancia en procesos patológicos crónicos sugieren que, dentro de la población de miRNAs circulantes, podrían encontrarse marcadores de presencia, grado y pronóstico de enfermedad. En este sentido, existe un importante campo de investigación en cáncer (incluidos en el concepto de *biopsia líquida*) y, con menor desarrollo, en diabetes tipo II y 30 enfermedades neurodegenerativas.

Existen barreras metodológicas, no resueltas, que limitan el análisis de los miRNA, así como la existencia de sesgos no asumidos que tienen como consecuencia la publicación de resultados contradictorios y la falta de reproducibilidad de numerosos estudios. Los problemas metodológicos se originan en dos características de los miRNAs:

- 5 • **La naturaleza química.** Un tamaño extremadamente reducido que carece de estructuras comunes (como la cola poly(A) de los RNAs mensajeros). La naturaleza química supone un desafío considerable que condiciona la metodología a todos los niveles (purificación, retrotranscripción, amplificación y reacción de ligación).
- 10 • **La reducida concentración.** Los miRNAs son capaces de desarrollar su función biológica (suprimir la expresión de RNAs mensajeros concretos) a concentraciones muy reducidas. En consecuencia, son una población minoritaria dentro del conjunto del RNA celular. Este problema se exagera en los estudios de miRNA circulante, en los que se suele trabajar con cantidades muy inferiores a las obtenidas en tejido.

Existen tres grupos metodológicos para abordar el análisis de los miRNAs: (1) Análisis de
15 secuencias concretas de miRNAs mediante RT-PCR (retrotranscripción - PCR), (2) Hibridación sobre *microarrays*, y (3) miRNA-seq (secuenciación paralela masiva).

El estudio de los miRNAs circulantes como biomarcadores de cáncer tiene desafíos adicionales. Las concentraciones en sangre son mucho más reducidas que en tejido tumoral pero, además, la proporción de miRNA circulante que proviene de células neoplásicas es muy
20 variable, ya que depende del volumen de la masa tumoral y del estadio de cáncer. Por otra parte, ciertas especies de miRNA pueden provenir de exosomas liberados por poblaciones neoplásicas minoritarias pero de gran relevancia (como las células iniciadoras de tumor o *cancer stem cells*).

La secuenciación paralela masiva de miRNA (miRNA-seq) debería tener la suficiente potencia
25 como para abordar satisfactoriamente los desafíos del miRNA circulante. Sin embargo, los resultados no son reproducibles entre las diferentes metodologías y plataformas y se agravan cuando se aplican a las peculiaridades del miRNA circulante. En consecuencia, el “clásico y obsoleto” *microarray* sigue siendo el método de elección para el análisis masivo de secuencias de miRNA.

Los problemas de reproducibilidad no están originados en la secuenciación masiva en sí, sino que se originan a nivel de la fabricación de librerías de secuenciación masiva (librerías miRNA-seq).

Existen varios métodos para fabricar librerías miRNA-seq. Todos ellos utilizan reacciones de ligación en una o dos etapas. La reacción de ligación introduce sesgos, ya que la probabilidad de unir dos moléculas (DNA o RNA) depende de sus secuencias. En consecuencia, durante la fabricación de librerías mediante reacciones de ligación, puede haber una alteración significativa de frecuencias relativas dentro de la población de moléculas a secuenciar (sobrevalorando ciertas secuencias e infravalorando otras). El sesgo de ligación es un problema metodológico aún no resuelto, cuyo error es muy difícil de cuantificar.

Además, en miRNA, se incrementan considerablemente los sesgos asociados a la reacción de ligación, debido a dos razones:

1.- Error de muestreo incrementado. Las reacciones de ligación tienen eficiencias muy reducidas, lo que unido a la baja concentración del miRNA (especialmente en sangre) puede dificultar la cuantificación de las variantes miRNA de menor frecuencia.

2.- Alteración de la probabilidad de ligación. En moléculas (RNA o DNA) de reducido tamaño se incrementa considerablemente la influencia de la secuencia sobre la probabilidad de ligación. Pero además, este fenómeno adquiere una especial relevancia cuando se ligan moléculas de RNA, probablemente debido a la adquisición de diferentes estructuras secundarias que modulan la velocidad de reacción.

Por tanto, es necesario desarrollar técnicas potentes para realizar un análisis genético preciso. Lo suficientemente preciso como para que el miRNA circulante pueda reflejar en su composición tumores incipientes o subpoblaciones tumorales de relevancia clínica.

DESCRIPCIÓN DE LAS FIGURAS

Figura 1. Elongación de hebras de cadena simple unidas covalentemente por el extremo 5' a una nanopartícula magnética.

Figura 2. Descripción de la *herramienta coloidal* usada en la fabricación de librerías miRNA-seq.

Figura 3. La captura específica del material genético sobre la superficie de las herramientas coloidales.

Figura 4. Retrotranscripción sobre las partículas.

Figura 5. Bloqueo de la partícula.

5 **Figura 6.** Elongación del segundo adaptador

Figura 7. Finalización de la librería.

Figura 8. Electroforesis en gel de agarosa que muestra el tamaño de las librerías de miRNA. Calle 1, patrón de tamaños (las dos bandas inferiores corresponden a los tamaños de 100 y 200 pb). Calles 2 y 3, librerías de miRNA (a partir de secuencias de RNA sintéticas). Calle 4, banco (se ha seguido todo el proceso en paralelo, pero sin añadir RNA).

Figura 9. Muestra 3 secuencias que forman parte del Ejemplo 1, en las que en sombreado aparecen las secuencias procedentes del plásmido pGEM que contiene una librería miRNA-seq insertada en su sitio polylinker; en azul, las secuencias de los dos adaptadores de secuenciación masiva de la plataforma Ion S5; y en negrita, el cDNA de miRNA sintético.

15

DESCRIPCIÓN DE LA INVENCION

Los autores de la presente invención han desarrollado una nueva metodología para creación de librerías de miRNA. La nueva metodología aplica la nanotecnología al proceso lo que permite añadir (realmente elongar) adaptadores mediante DNA polimerasa, evitando la reacción de ligación y los problemas asociados a ella (reducción de eficiencia y sesgos).

20

PROCEDIMIENTO DE LA INVENCION

Por tanto, un **primer aspecto** de la invención se refiere a un procedimiento para obtener una librería de secuenciación masiva con el ADN complementario, cDNA, de una población de miRNA de interés, que comprende:

25

a) Capturar los miRNA de interés sobre partículas magnéticas unidas a oligonucleótidos que portan en su mitad 3' una secuencia de polinucleótido de purina de entre 18 y 20 bases, preferiblemente poly(T) de entre 18 y 20 timinas, y en su mitad 5' la secuencia de uno de los adaptadores de secuenciación masiva, unidos los oligonucleótidos a la superficie de las

partículas magnéticas mediante un enlace covalente por su extremo 5', mediante un proceso que comprende tratar la población de miRNAs con una *poly(purina) polymerase*, preferiblemente *poly(A) polymerase*, de forma que adquieran un extremo 3' de entre 20 y 30 nucleótidos de adeninas (*tailing* de RNA).

- 5 b) Llevar a cabo una reacción de retrotranscripción,
- c) Realizar un lavado alcalino de las partículas para eliminar los sustratos de la reacción de retrotranscripción (deshibridar los miRNA y dejar su secuencia complementaria en los extremos de los oligos) y eliminar cualquier molécula de ácido nucleico no unida covalentemente a las partículas magnéticas,
- 10 d) Bloquear los oligonucleótidos (unidos a las partículas magnéticas) que no han adquirido una secuencia de cDNA en el extremo 3' añadiendo un nucleótido terminador,
- e) Realizar una elongación del DNA unido a las partículas del extremo 3' con una cola de nucleótidos, preferiblemente de guaninas (polyG),
- f) Elongar el segundo adaptador de secuenciación masiva en la cola de nucleótidos,
- 15 preferiblemente cola de polyG, mediante una plantilla de elongación que consta de:
 - en el extremo 5', una cola de 20 citosinas,
 - en el extremo opuesto, una secuencia complementaria al segundo adaptador de secuenciación masiva.

Quedando la secuencia de los miRNAs de interés entre los dos adaptadores,

- 20 g) Realizar una reacción de polimerización del DNA, preferiblemente una PCR estándar usando cebadores (*primers*) frente a los extremos de los dos adaptadores.

La etapa (a) consiste en la captura específica del material genético sobre la superficie de las herramientas coloidales.

- La partícula porta en su superficie un oligonucleótido unido covalentemente por su extremo 5' (en el ejemplo de la **figura 2**, mediante un enlace amida). Es recomendable utilizar un *linker* de 10-20 carbonos (-CH₂)_n entre el enlace amida y el oligonucleótido, con la intención de minimizar impedimentos estéricos sobre la superficie de la partícula. La densidad óptima de oligonucleótido por unidad de partícula debe de ponerse a punto para cada combinación de partícula magnética y oligonucleótido. En el ejemplo descrito en el presente texto, la densidad
- 25 fue de 0,5 nM (nanoMoles, nMoles) de oligonucleótido por mg de partícula. En el ejemplo, el
- 30

enlace amida se creó incubando oligonucleótidos y partículas durante 12-14 horas, con agitación estocástica, en una disolución de 250 mM (microMoles, mMoles) de 3-(*Dimethylaminopropyl*)-*N'*-ethylcarbodiimide (EDAC), 1 M de NaCl y tampón MES 100 mM a pH 5.

5 Por tanto, en una realización preferida de este aspecto de la invención, la partícula magnética se caracteriza por que:

I) tiene un núcleo magnético,

II) tiene la superficie recubierta con compuestos orgánicos con grupos de carácter ácido expuestos que le aportan carga negativa,

10 III) es estable a PH alcalino y ácido, en un rango amplio entre pH 2 y 14

IV) tiene un reducido coeficiente de sedimentación y una reducida agregación,

V) tienen un tamaño de entre 100 nm (nanómetros) y 2000 nm, preferiblemente entre 700 nm y 1500 nm, y más preferiblemente de aproximadamente 800 nm,

VI) no inhibe la Taq polimerasa y puede usarse en reacciones de PCR, y

15 VII) La partícula es estable a temperaturas de hasta 100°C.

Idealmente, el tamaño de la cola poly(A) que se añade al extremo 3' de los miRNAs es de 20-30 nucleótidos. El tiempo de reacción y la cantidad de enzima pueden ajustarse para que la cola poly(A) esté en el rango deseado. Para finalizar la reacción, basta con calentar 10 minutos a 65°C con el objetivo de desnaturalizar la enzima.

20 Una vez realizada la reacción de *tailing*, se añaden las herramientas coloidales que capturan la población de miRNAs gracias a una secuencia de timinas (idealmente entre 18 y 20) en cadena simple. El oligonucleótido (DNA) unido a la partícula porta en su mitad 3' la secuencia poly(T) y en su mitad 5' la secuencia de uno de los adaptadores de secuenciación masiva (figura 1). Dicho oligo está unido por su extremo 5' a la superficie de las partículas mediante
25 un enlace covalente del tipo amida.

Las herramientas coloidales pueden capturar a los miRNAs en el mismo tampón en el que se realizó la reacción de *tailing* (es decir, no es necesario purificar el producto de reacción mediante procedimientos químicos). La captura (mediante hibridación entre la cola poly(A) de los miRNAs y la secuencia poly(T) de las partículas) se realiza durante 2-4 horas bajo
30 agitación estocástica y a una temperatura de 55°C.

En el ejemplo de la invención, tras la creación del enlace amida, las partículas magnéticas son sedimentadas sobre un imán, lavadas 2 veces con NaOH 200 mM e incubadas en la misma disolución alcalina durante 30 minutos. El incremento brusco de pH provocado por la disolución de NaOH tiene una doble función:

- 5 • Hacer que los oligonucleótidos que permanezcan covalentemente unidos a las partículas estén en forma de cadena simple (la hebra complementaria no unida covalentemente es eliminada durante los lavados).
- Eliminar de la superficie de la partícula productos indeseados de reacción (ésteres de acilisourea) y restaurar grupos carboxilos en aquellos radicales que no hayan
- 10 realizado un enlace amida con los oligonucleótidos.

Una vez realizada la incubación alcalina, se reequilibra el pH mediante dos lavados en tampón Tris-HCl 100 mM a pH 7,4 y se resuspenden las partículas en tampón Tris-HCl 10 mM a pH 7,4.

En el ejemplo de la invención, la secuencia del oligonucleótido unido covalentemente a la

15 partícula consta de:

- En su mitad 3', 18 timinas.
- En su mitad 5', la secuencia de uno de los adaptadores de secuenciación masiva (en el ejemplo, el Tc-P1 de la plataforma Ion S5).

Para finalizar, la partícula se hibrida con un oligonucleótido complementario de la secuencia

20 Tc-P1 (en un tampón adecuado) y se lava en Tris-HCl 10 mM a pH 7,4 para eliminar el exceso de oligo no hibridado. Este paso, que es opcional, mejora la eficiencia de la captura.

En la figura 3 se representa la reacción de tailing de miRNAs y la captura específica por las herramientas coloidales.

25 En la etapa (b) o retrotranscripción sobre las partículas la unión entre el oligonucleótido de las partículas y la población de miRNAs es utilizada para el cebado (*priming*) de la reacción de retrotranscripción. Por esta razón, se recomienda que la longitud de la cola poly(T) del oligonucleótido unido a las partículas sea inferior a la cola poly(A) de los miRNAs (de esta forma se asegura que los extremos 3' de los oligonucleótido de las partículas permanezcan

30 hibridados a la cola poly(A) de los miRNAs, lo que incrementa la eficiencia de la retrotranscripción).

Las partículas unidas a miRNAs se resuspenden en medio de reacción de retrotranscripción (puede utilizarse cualquier kit de retrotranscripción comercial) que se desarrolla (de forma estándar) incubando durante 30-60 minutos a 42°C. Es conveniente, justo antes de añadir la retrotranscriptasa, calentar a 70-80°C durante 5-10 minutos para eliminar estructuras secundarias de RNA que puedan disminuir la eficiencia de la retrotranscripción.

Tras la reacción de retrotranscripción, se realiza un lavado alcalino de las partículas (etapa (c)), con una doble función:

- Eliminar los sustratos de la reacción de retrotranscripción.
- Eliminar cualquier molécula de ácido nucleico no unida covalentemente a las partículas.

Como resultado, las partículas magnéticas llevan unidas covalentemente en 5' moléculas de DNA de cadena simple que portan:

- En 5' la secuencia de uno de los adaptadores de secuenciación masiva (Tc-P1 de Ion S5 en el ejemplo mostrado) seguido de una secuencia poly(T).
- En 3' el DNA copia (cDNA) de la población de miRNAs.

En la **figura 4** se muestra la reacción de retrotranscripción y el resultado de los lavados alcalinos.

En la etapa (d), bloqueo de la partícula. Los oligonucleótidos unidos a la herramienta coloidal deben de estar en exceso para incrementar la eficiencia de la captura y retrotranscripción de miRNAs. En consecuencia, tras las etapas (a) – (c), una proporción elevada de oligos no llega a elongar (a adquirir una secuencia cDNA en 3'). Estos oligonucleótidos sin elongar deben de ser bloqueados para que no interfieran en las etapas finales del proceso. El bloqueo de la partícula es un requisito esencial.

El bloqueo se realiza añadiendo selectivamente un nucleótido terminador (dideoxitimina) en el extremo 3' de los oligonucleótidos que no han adquirido secuencias cDNA. Para ello, las partículas se resuspenden en un medio con tampón de reacción de PCR, Taq-polimerasa (puede usarse cualquier Taq-polimerasa, u otra DNA-polimerasa termoestable, de origen comercial) un oligonucleótido (que hemos denominado plantilla de elongación) y dideoxitimina-trifosfato (ddTTP) a una concentración en torno a 0,2 mM.

Los dideoxi-nucleótidos (como dideoxi-timina) carecen de grupo 3' hidroxilo y, en consecuencia, no pueden seguir incorporando nuevos nucleótidos mediante un enlace fosfodiéster (este es, por ejemplo, el fundamento de la secuenciación Sanger).

La plantilla de elongación consta de:

- 5 • En su extremo 5', una cola de 21 adeninas (en cualquier caso, un número superior a la cola de timinas de los oligos de las herramientas coloidales).
- En el extremo opuesto, una secuencia complementaria al adaptador de secuenciación masiva.
- 10 • Opcionalmente, un dideoxinucleótido en el extremo 3', que impida que la propia plantilla pueda elongar.

Mediante esta plantilla de elongación, se incorpora una dideoxitimina únicamente en aquellos oligonucleótidos de la partícula que no hayan adquirido una secuencia cDNA.

La reacción de bloqueo se realiza siguiendo un protocolo de varios ciclos:

- Desnaturalización a 95°C
- 15 • Hibridación de la plantilla de elongación (en torno a 60°C)
- Elongación de los extremos 3' hibridados a la plantilla de elongación (a la temperatura óptima de la DNA-polimerasa, usualmente 72-74°C)

Dando varios ciclos (5-15 ciclos), puede incrementarse la proporción de moléculas elongadas, llegando virtualmente al 100 %. Esta es la razón de que las plantillas de elongación tengan inactivadas su extremo 3' (mediante un dideoxi-nucleótido u otro nucleótido terminador). Para que la única elongación posible sea la de los extremos 3' de las hebras de DNA unidas covalentemente a la partícula.

Finalmente, se realizan dos lavados alcalinos de partícula, seguido de un reequilibrado de pH y un resuspensión en tampón a pH 7,4. Este lavado alcalino elimina el medio de la reacción de bloqueo y la plantilla de elongación usada para bloquear específicamente oligos sin secuencia cDNA (etapa (d)) y únicamente deja sobre la partícula hebras de cadena simple unidas covalentemente en 5'.

El proceso de bloqueo de los oligonucleótidos de la partícula no elongados se muestra en la figura 5.

En las etapas (e) y (f) se realiza la elongación del segundo adaptador. Mediante una transferasa terminal, se realiza una reacción de *tailing* de DNA que añade una cola de guaninas (poly(G)) a los extremos 3' (elongables) del DNA de cadena simple unido a las partículas. El *tailing* de DNA, a diferencia del *tailing* de RNA (etapa (a)) puede añadir cualquier tipo de nucleótido al extremo 3' de moléculas de DNA (tanto de cadena doble como simple). La reacción de *tailing* de DNA, realizada en presencia únicamente de dGTP, añade una cola de poly(G).

Para la reacción de *tailing* de DNA, las partículas se resuspenden en presencia del medio de reacción adecuado (sirve cualquier transferasa terminal comercial) y dGTP a 0,2 mM. Idealmente, el tamaño de la cola poly(G) que se añade al extremo 3' del DNA unido a las partículas es de 15-20 nucleótidos, no superando nunca los 20. El tiempo de reacción y la cantidad de enzima pueden ajustarse para que la cola poly(G) esté en el rango deseado. Para finalizar la reacción, basta con calentar 10 minutos a 65°C con el objetivo de desnaturalizar la enzima.

Tras el *tailing* de DNA, las partículas magnéticas se lavan dos veces en un tampón adecuado (con PBS, TBS u otro tampón similar) con la intención de eliminar los sustratos de la reacción.

La cola poly(G) permite elongar el segundo adaptador de secuenciación masiva mediante una plantilla de elongación que hibride específicamente en la cola poly (G). En nuestro ejemplo, esta segunda plantilla de elongación consta de:

- En su extremo 5', una cola de 20 citosinas (razón por la que advertimos que la cola poly(G) añadida durante la reacción de *tailing* no debe de ser superior a 20).
- En el extremo opuesto, una secuencia complementaria al segundo adaptador de secuenciación masiva. En nuestro ejemplo añadimos el adaptador A (de Ion S5), la secuencia Key de S5 y una secuencia identificadora de muestra (barcode 1 de S5).
- Opcionalmente, un dideoxinucleótido en el extremo 3', que impida que la propia plantilla pueda elongar.

La reacción de elongación se realiza mediante el mismo procedimiento descrito en la etapa (b), con la salvedad de que en lugar de dideoxi-timina-trifosfato se añade una mezcla de los 4 nucleótidos trifosfato (dATP, dTTP, dGTP, dCTP) a 0,2 mM. A continuación, se realizan dos lavados alcalinos (y reequilibrado de pH) para eliminar restos del medio de reacción y hebras de DNA no unidas covalentemente a las partículas.

Estos lavados alcalinos eliminan el medio de la reacción de elongación y la plantilla de elongación usada para añadir la secuencia del segundo adaptador, dejando únicamente sobre la partícula hebras de cadena simple unidas covalentemente en 5'.

En el ejemplo, hemos mostrado un *tailing* con guaninas que hibrida a la cola poly(C) de una
5 plantilla de elongación. Nuestro diseño no está limitado a esta combinación y pueden utilizarse otros extremos de nucleótidos complementarios.

El proceso de elongación del segundo adaptador de muestra en la figura 6.

En la etapa (g) se produce la finalización de la librería. Para finalizar, se realiza una PCR
estándar con las partículas y usando cebadores (*primers*) frente a los extremos de los dos
10 adaptadores. Se elimina la partícula (sedimentándola sobre un imán) y se purifica el producto de PCR (figura 7). En el caso de librerías fabricadas a partir de RNA tisular, en el que no se hayan eliminado los RNAs de mayor tamaño (ribosómicos y mensajeros) es necesario realizar una purificación selectiva de tamaños (por debajo de 200 pb). En el mercado existen varios kits que purifican DNA en función del tamaño, aunque se pueden utilizar técnicas alternativas
15 basadas en una purificación tras separación por electroforesis.

El resultado es una librería de secuenciación masiva con el cDNA de la población de miRNA flanqueado por dos homopolímeros (A/T y G/C) de unos 10-20 pares de bases cada uno. En la figura 8 se muestran algunas secuenciaciones realizadas durante las pruebas de concepto.

20 KIT O DISPOSITIVO DE LA INVENCION

Otro **aspecto** de la invención se refiere a un kit o dispositivo, de ahora en adelante kit o dispositivo de la invención, que comprende los elementos necesarios para llevar a cabo el procedimiento de la invención.

25 USOS DE LA INVENCION

Otro **aspecto** de la invención se refiere al uso del procedimiento de la invención, o del kit o dispositivo de la invención, para el análisis de alta resolución de las poblaciones de miRNAs en una muestra biológica. Preferiblemente la muestra biológica es sangre, y aún más preferiblemente plasma.

El análisis de miRNAs en sangre, más efectivo y menos costoso, tendría varias aplicaciones para el diagnóstico de numerosas enfermedades. Por ejemplo, pero sin limitarnos, permitiría el empleo de miRNAs circulantes como biomarcadores de cáncer, y más concretamente, de cáncer de mama.

- 5 Por tanto, otro aspecto de la invención se refiere al uso del procedimiento de la invención, o del kit o dispositivo de la invención, para el diagnóstico de una enfermedad. Más preferiblemente la enfermedad es el cáncer. En una realización particular, la enfermedad es el cáncer de mama.

10 **Definiciones**

Los ácidos nucleicos o polinucleótidos para la secuenciación incluyen, pero no se limitan a, ácidos nucleicos tales como ADN, ARN o PNA (ácido nucleico peptídico), variantes o fragmentos de los mismos y/o concatámeros de los mismos. Los polinucleótidos pueden ser de una secuencia conocida o desconocida, de naturaleza natural o artificial y pueden ser de cualquier fuente (por ejemplo, eucariotas o procariotas). Los polinucleótidos pueden derivarse naturalmente, producirse recombinantemente o sintetizarse químicamente. Los polinucleótidos concatamerizados pueden contener subunidades o análogos de los mismos que pueden o no pueden producirse en la naturaleza, o subunidades modificadas. Pueden usarse métodos como se describe en este documento para determinar una secuencia de un polinucleótido. La longitud del ácido nucleico diana para la secuenciación puede variar. Por ejemplo, el ácido nucleico para la secuenciación puede incluir al menos 10, al menos 20, al menos 30, al menos 40, al menos 50, al menos 100, al menos 200, al menos 500, al menos 1.000, al menos 10.000, al menos 100.000, al menos 1.000.000, al menos 10.000.000 nucleótidos. El polinucleótido para la secuenciación puede ser de origen genómico o fragmentos o variantes de los mismos. La cadena de ácido nucleico para la secuenciación puede ser de cadena sencilla y puede o no derivarse de una molécula de ácido nucleico bicatenario. Las moléculas de cadena sencilla también pueden producirse, por ejemplo, mediante métodos y tecnologías de síntesis química o in vitro. Las realizaciones como se describen en la presente memoria descriptiva no están limitadas por los métodos preparatorios del ácido nucleico y cualquier número de métodos puede ser practicado por los expertos en la técnica para proporcionar una composición para uso en los métodos descritos. Por ejemplo, en la secuencia mediante metodologías de síntesis a menudo se genera una biblioteca que comprende los ácidos nucleicos diana, y después se secuencia una parte de la biblioteca de ADN.

Por "unido operativamente" quiere decirse que dos estructuras químicas se unen entre sí de tal manera que permanecen unidas a través de las diversas manipulaciones a las que se espera que se sometan. Normalmente, el resto funcional y el oligonucleótido codificante se unen covalentemente a través de un grupo de unión apropiado. Por ejemplo, el grupo de unión
 5 puede ser un resto bifuncional con un sitio de unión para el oligonucleótido codificante y un sitio de unión para el resto funcional.

La unión entre el extremo 5' del oligonucleótido y la superficie de la partícula debe de realizarse mediante un enlace covalente. Preferiblemente existen dos opciones: un enlace amida (tal como se muestra en los ejemplos de la invención) o enlaces basados en grupos tiol
 10 como el enlace disulfuro.

Los métodos que se describen en la presente memoria descriptiva no están limitados por ningún método de preparación de muestras de secuenciación particular y las alternativas serán fácilmente evidentes para cualquier experto en la técnica y se consideran dentro del alcance de la presente descripción.

15 El término "herramienta coloidal", en esta memoria es sinónimo de "partículas magnéticas unidas a oligonucleótidos".

EJEMPLOS DE LA INVENCION

EJEMPLO 1

Para analizar el funcionamiento de la metodología, se fabricaron librerías miRNA-seq a partir
 20 de RNA sintético (en la prueba utilizamos una colección de 10 moléculas sintéticas correspondientes a secuencias miRNA humanas) a concentraciones muy reducidas (del orden de pmoles y fmoles, para tratar de emular las concentraciones existentes

Tras finalizar las librerías de miRNA, se realizó una prueba de concepto mediante secuenciación Sanger. La librería de miRNA está compuesta por una colección de secuencias
 25 diferentes, por lo que no es un sustrato adecuado para la secuenciación Sanger.

Realizamos una purificación y ligación de las librerías miRNA en el plásmido pGEM-T (sistema comercial de clonación de amplicones), seguido de transformación en E. coli (cepa JM109). A continuación, las bacterias se crecieron en un medio selectivo con ampicilina (el plásmido pGEM aporta resistencia a ampicilina). Únicamente aquellos clones de bacterias que
 30 asimilaron un plásmido fueron capaces de crecer en el medio selectivo y formar colonias en agar con medio de cultivo (las colonias son clones que portan el plásmido pGEM-T con una única versión de inserto).

Tras la clonación, se repicaron y crecieron una veintena colonias, a las que se les extrajo el DNA plasmídico, que se secuenció por el método de Sanger utilizando primers específicos que flanquean el punto de inserción. La secuencia correspondiente a los insertos tuvo un tamaño medio de 128 pares de bases.

- 5 El inserto se une al plásmido abierto mediante una reacción de ligación que, de no tener sesgos, tendría dos opciones en igual frecuencia (sentido o antisentido, respecto a la secuencia plasmídica). Sin embargo, en todas las secuenciaciones de nuestra prueba, el inserto estuvo en dirección sentido. Este dato evidencia la existencia de fuertes sesgos de ligación, que se incrementan al trabajar con secuencias relativamente cortas.
- 10 A continuación (Figura 9), se muestran 3 ejemplos de secuenciación Sanger realizados en la prueba de concepto, en las que se muestra:

Secuencia 1. Orientación sentido. Contiene el DNA copia de miR-18a (en negrita)

Secuencia 2. Orientación sentido. Contiene el DNA copia de miR-26b (en negrita)

Secuencia 3. Orientación sentido. Contiene el DNA copia de miR-135b (en negrita)

- 15 • En sombreado, se muestran las secuencias procedentes del plásmido pGEM que contiene una librería miRNA-seq insertada en su sitio polylinker.
- En azul, las secuencias de los dos adaptadores de secuenciación masiva de la plataforma Ion S5:
- 20 ○ ctcacccctgcgtgtctccgactcag**ctaaggtaac**gat - Adaptador A en la hebra sentido (contiene el barcode-1 **ctaaggtaa**). La secuenciación masiva comenzaría a partir del barcode.
- atcacccgactgcccatagagagg - Adaptador TcP1 en la hebra antisentido.
- En negrita, el cDNA de miRNA sintético.

- 25 Se observa además que las colas poly(C/G) y poly(A/T) tienen tamaños variables debido a la propia naturaleza de las reacciones de tailing y al proceso de elongación sobre partícula.

Secuencia 1. Orientación sentido. Contiene el DNA copia de miR-18a (en negrita)

5' _gggcccgcgctgcgatgctcccgccgcatggccgcgggattccatctcatccctgcgtgtctccgactcag**ctaaggtaa**
cgatcccccccccccc**ctaagggtgcatctagtgcagatac**aaaaaaaaaaaaaaaaaaaaaatcacccgactaccata
30 gagaggaatcactagtgcggccgcctgcaggtcgaccatatgggagagctcccaacgcgtt_3'

Secuencia 2. Orientación sentido. Contiene el DNA copia de miR-26b (en negrita)

5' _gggcccgcgctgcatgctcccgccgcatggccgcgggattccatctcatccctgctgtctccgactcag**ctaaggtaa**
cgatcccccccccccccc**ttcaagtaattcaggataggta**aaaaaaaaaaaaaaaaaaaaaatcaccgactgccata
gagaggaatcactagtgcggccgcctgcaggtcgaccatatgggagagctccaacgcgtt_3'

5 **Secuencia 3.** Orientación sentido. Contiene el DNA copia de miR-135b (en negrita)

5' _gggcccgcgctgcatgctcccgccgcatggccgcgggattccatctcatccctgctgtctccgactcag**ctaaggtaa**
cgatcccccccccc**tatggcttttcattcctatgtg**aaaaaaaaaaaaaaaaaaaaaatcaccgactgccatagag
aggaatcactagtgcggccgcctgcaggtcgaccatatgggagagctccaacgcgtt_3'

10 **EJEMPLO 2. Validación del procedimiento mediante el análisis de miRNA circulante como biomarcador en cáncer de mama.**

El estudio consta de dos cohortes:

• Una cohorte de 200 mujeres con cáncer de mama, de las que se recogerán muestras de suero de forma prospectiva en el Hospital San Cecilio de Granada y en el Complejo
15 Hospitalario de Jaén. Las muestras serán preservadas y gestionadas por el Biobanco Andaluz (se adjuntan documentos de disponibilidad y cesión de muestras por Biobanco). En caso de que haya terapia neoadyuvante, previa a la cirugía, se obtendrán muestras de suero, antes y después del tratamiento.

• Una cohorte control de 30 mujeres sanas, sin síndrome metabólico y sin presencia de cáncer
20 en su historial clínico. La cohorte control deberá ser similar en edad a la cohorte de pacientes con cáncer de mama. Las muestras se encuentran recogidas retrospectivamente y criopreservadas en el Biobanco Andaluz.

El estudio incluye muestras controles de mujeres sanas y enfermas de cáncer de mama en diferentes estadios de progresión (estadios 0, I, II, III y IV) y con diferentes fenotipos (Luminal
25 A, Luminal B, HER2 y triple negativo). En aquellas mujeres sometidas a terapia neoadyuvante, la cohorte constará de muestras de sangre obtenidas antes y después del tratamiento.

Se evaluará la composición, de alta resolución, de los miRNAs circulantes, analizando los siguientes parámetros:

- Capacidad de identificar la presencia de cáncer mama.
- 30 • Capacidad para identificar los diferentes estadios de progresión del cáncer de mama.

- Capacidad para identificar los principales fenotipos del cáncer de mama y su correlación con la expresión de los principales marcadores clínicos de fenotipo: ER (*Estrogen Receptor*), PR (*Progesteron Receptor*), HER2 (*Human Epidermal Growth Factor Receptor 2*) y Ki67 (marcador nuclear de proliferación).
- 5 • Capacidad de pronóstico de recidiva, supervivencia y erradicación del cáncer. En este parámetro se tendrá en cuenta la evolución de las poblaciones de miRNAs circulantes tras la terapia neoadyuvante.

REIVINDICACIONES

1.- Un procedimiento para obtener una librería de secuenciación masiva con el cDNA de una población de miRNA de interés, que comprende:

- 5 a) Capturar los miRNA de interés sobre partículas magnéticas unidas a oligonucleótidos que portan en su mitad 3' una secuencia poly(T) de entre 18 y 20 timinas, y en su mitad 5' la secuencia de uno de los adaptadores de secuenciación masiva, unidos los oligonucleótidos a la superficie de las partículas magnéticas mediante un enlace covalente por su extremo 5', mediante un proceso que comprende tratar la población de miRNAs con una *poly(A)* *polymerase*, de forma que adquieran un extremo 3' de entre 20 y 30 nucleótidos de adeninas (tailing de RNA).
10
- b) Llevar a cabo una reacción de retrotranscripción,
- c) Realizar un lavado alcalino de las partículas para eliminar los sustratos de la reacción de retrotranscripción (deshibridar los miRNA y dejar su secuencia complementaria en los extremos de los oligos) y eliminar cualquier molécula de ácido nucleico no unida
15 covalentemente a las partículas,
- d) Bloquear los oligonucleótidos que no han adquirido una secuencia de cDNA en el extremo 3' añadiendo un nucleótido terminador,
- e) Realizar una elongación del DNA unido a las partículas del extremo 3' con una cola de guaninas (polyG),
20
- f) Elongar el segundo adaptador de secuenciación masiva en la cola de polyG mediante una plantilla de elongación que consta de:
 - en el extremo 5', una cola de 20 citosinas,
 - en el extremo opuesto, una secuencia complementaria al segundo adaptador de secuenciación masiva.
- 25 g) Realizar una PCR estándar usando cebadores (primers) frente a los extremos de los dos adaptadores.

2.- El procedimiento según la reivindicación 1, donde después de cada etapa se realiza una precipitación y un lavado con tampón adecuado.

- 3.- El procedimiento según cualquiera de las reivindicaciones 1-2, donde la elongación de la etapa 5 se realiza con una transferasa terminal
- 4.- El procedimiento según cualquiera de las reivindicaciones 1-3, donde la cola de polyG de la etapa 5 debe ser de entre 15-20 nucleótidos de Guanina
- 5 5.- El procedimiento según cualquiera de las reivindicaciones 1-4, donde la plantilla de elongación del paso f) puede tener un dideoxinucleótido en el extremo 3', que impida que la propia plantilla pueda elongar.
- 6.- El procedimiento según cualquiera de las reivindicaciones 1-5, donde en caso de librerías fabricadas a partir de RNA tisular, en el que no se hayan eliminado los RNAs de mayor tamaño (ribosómicos y mensajeros) es necesario realizar una purificación selectiva de tamaños (por debajo de 200 pb) tras el paso g)
- 10 7.- El procedimiento según cualquiera de las reivindicaciones 1-6, donde se emplean otros extremos de nucleótidos complementarios.
- 8.- Uso del procedimiento según cualquiera de las reivindicaciones 1-7, para la identificación de miRNAs de interés en una muestra biológica.
- 15 9.- Uso del procedimiento, según la reivindicación anterior, donde la muestra biológica es sangre.
- 10.- Uso del procedimiento, según la reivindicación 8, donde la muestra biológica es plasma.
- 11.- Uso del procedimiento según cualquiera de las reivindicaciones 1-7, para el diagnóstico, pronóstico o respuesta al tratamiento de una enfermedad.
- 20 12.- Uso del procedimiento según cualquiera de las reivindicaciones 1-7, para el diagnóstico, pronóstico o respuesta al tratamiento del cáncer.

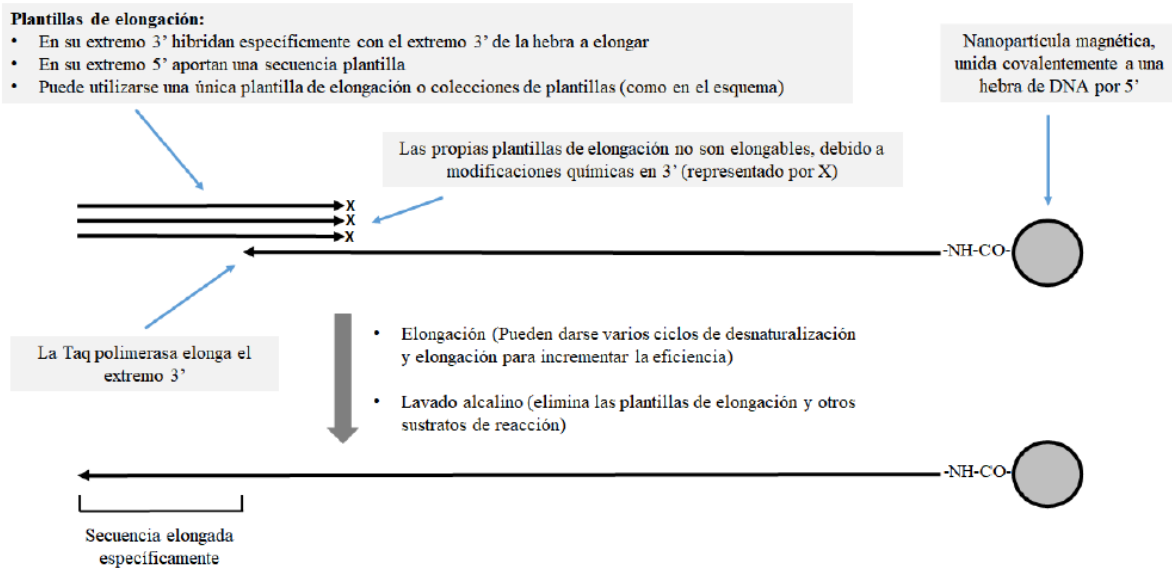


Figura 1

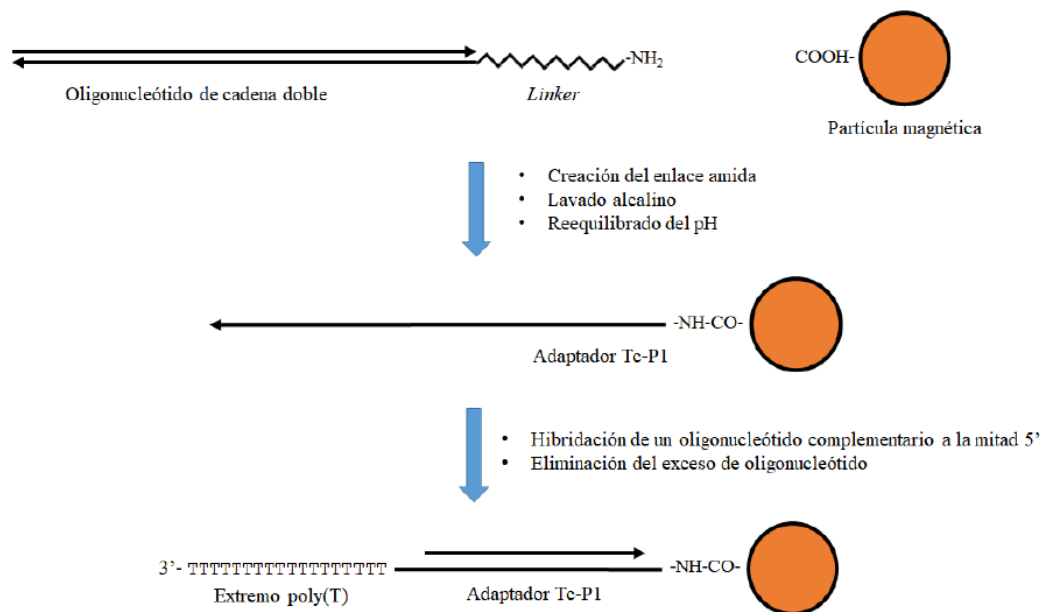


Figura 2

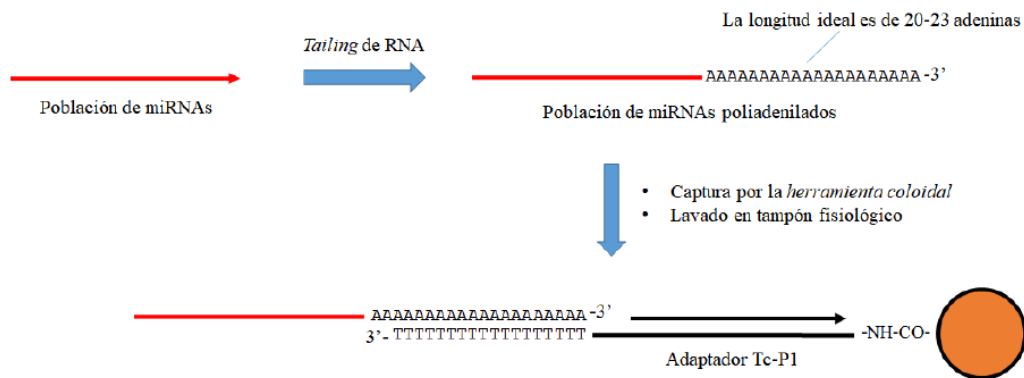


Figura 3

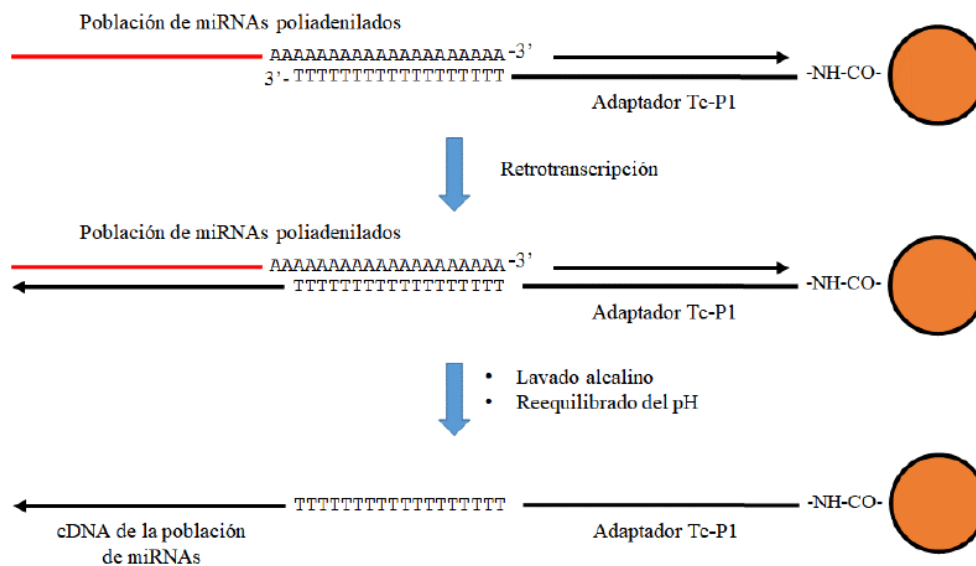


Figura 4

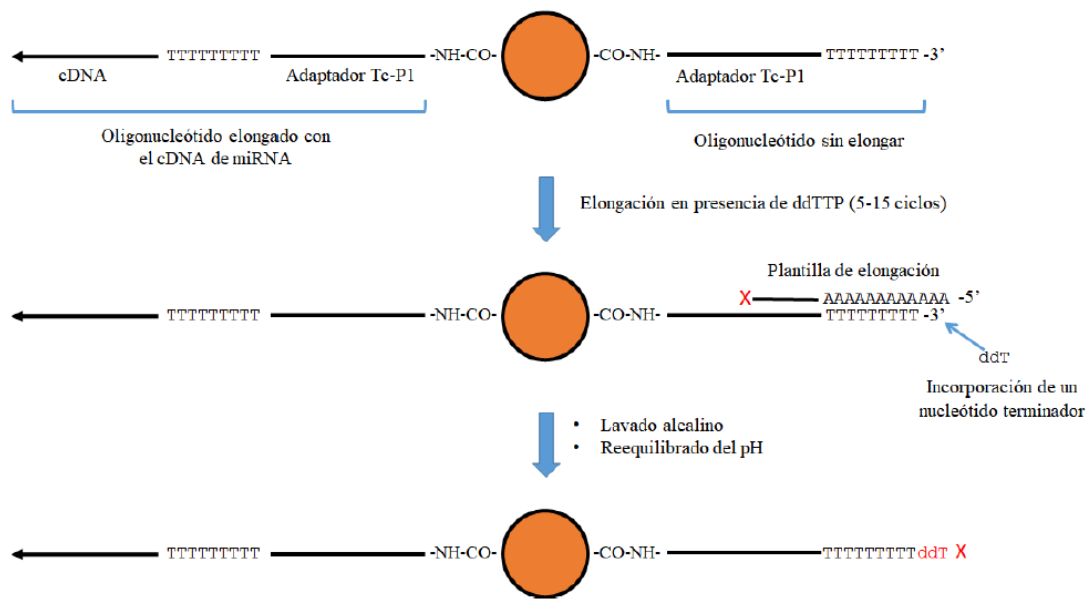


Figura 5

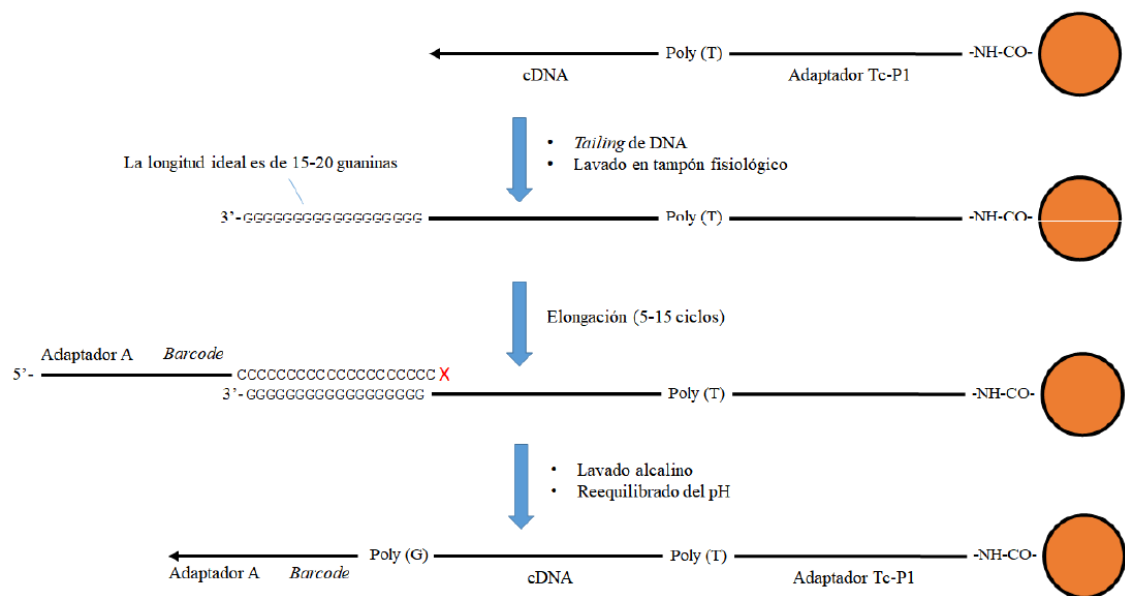


Figura 6

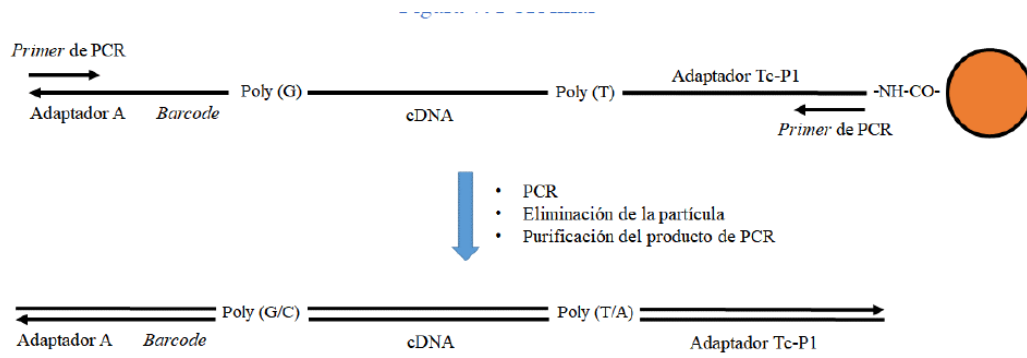


Figura 7

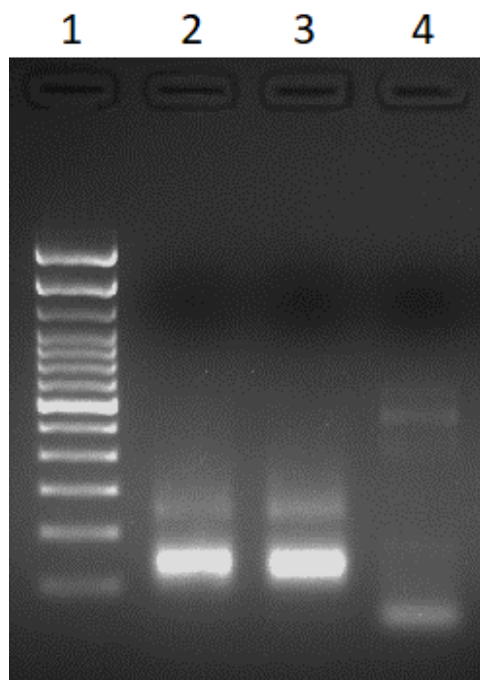


Figura 8

Secuencia 1. Orientación sentido. Contiene el DNA copia de *miR-18a* (en negrita)

5' _gggcccgcgctcgcatgctcccggccgcatggccgcgggattccatctcatccctgcgtgtctc
cgactcag**ctaaggtaac**gatcccccccccccccc**taagggtgatctagtgcagatac**aaaaaaaaa
aaaaaaaaaaaaaatcaccgactacccatagagaggaatcactagtgcggccgcctgcaggtcgaccat
atgggagagctcccaacgcgtt_3'

Secuencia 2. Orientación sentido. Contiene el DNA copia de *miR-26b* (en negrita)

5' _gggcccgcgctcgcatgctcccggccgcatggccgcgggattccatctcatccctgcgtgtctc
cgactcag**ctaaggtaac**gatcccccccccccccc**ttcaagtaattcaggataggt**aaaaaaaaa
aaaaaaaaaaaaaatcaccgactgcccatagagaggaatcactagtgcggccgcctgcaggtcgaccat
atgggagagctcccaacgcgtt_3'

Secuencia 3. Orientación sentido. Contiene el DNA copia de *miR-135b* (en negrita)

5' _gggcccgcgctcgcatgctcccggccgcatggccgcgggattccatctcatccctgcgtgtctc
cgactcag**ctaaggtaac**gatcccccccccccccc**tatggcttttcattcctatgtga**aaaaaaaaaaaa
aaaaaaaaaaaaaatcaccgactgcccatagagaggaatcactagtgcggccgcctgcaggtcgaccata
tgggagagctcccaacgcgtt_3'

Figura 9