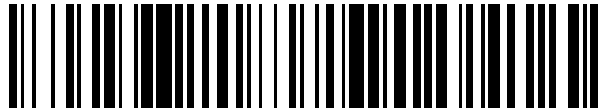


19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 835 051**

51 Int. Cl.:

**C12N 15/67** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **30.03.2017 PCT/EP2017/057592**

87 Fecha y número de publicación internacional: **05.10.2017 WO17167910**

96 Fecha de presentación y número de la solicitud europea: **30.03.2017 E 17719164 (0)**

97 Fecha y número de publicación de la concesión europea: **23.09.2020 EP 3436589**

54 Título: **Secuencias de UTR mínimas novedosas**

30 Prioridad:

**31.03.2016 EP 16163264**  
**30.06.2016 EP 16177094**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:  
**21.06.2021**

73 Titular/es:

**ETHRIS GMBH (100.0%)**  
**Semmelweisstrasse 3**  
**82152 Planegg, DE**

72 Inventor/es:

**PLANK, CHRISTIAN;**  
**RUDOLPH, CARSTEN;**  
**ANEJA, MANISH KUMAR y**  
**WEISS, LUDWIG**

74 Agente/Representante:

**ARIAS SANZ, Juan**

ES 2 835 051 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

## Secuencias de UTR mínimas novedosas

La presente invención se refiere a moléculas de ADN que pueden transcribirse para dar un ARNm que alberga secuencias de UTR novedosas que combinan las ventajas de ser extremadamente cortas y al mismo tiempo permitir altas eficiencias de traducción de moléculas de ARN que las contienen. Además, la presente invención se refiere a vectores que comprenden una molécula de ADN de este tipo y a células huésped que comprenden un vector de este tipo. Además, la presente invención se refiere a moléculas de ARN correspondientes que contienen tales UTR. Además, la presente invención se refiere a una composición farmacéutica que comprende la molécula de ARN descrita y opcionalmente un portador farmacéuticamente aceptable así como al uso de las UTR descritas para traducir una región codificante de una molécula de ARN para dar un polipéptido o una proteína codificado por dicha región codificante.

En los últimos años, el ARN mensajero (ARNm) se ha vuelto cada vez más relevante como nueva entidad farmacológica. En contraposición a productos terapéuticos génicos basados en ADN, el ARNm no necesita transportarse al interior del núcleo sino que se traduce directamente para dar una proteína en el citoplasma (J Control Release, 2011, 150:238-247, y Eur J Pharm Biopharm, 2009, 71:484-489). Esto hace que el ARNm sea más seguro para evitar posible mutagénesis por inserción, un riesgo improbable pero existente de los medicamentos génicos de ADN. Como consecuencia, están surgiendo productos terapéuticos de ARNm como alternativas prometedoras para terapias de sustitución de genes y proteínas en una amplia variedad de indicaciones médicas (J Control Release, 2011, 150:238-247; Eur J Pharm Biopharm, 2009, 71:484-489; Nat Biotech, 2011, 29:154-157, y Nat Rev Genet, 2011, 12:861-874). Sin embargo, tiene que superarse la fuerte inmunogenicidad así como la estabilidad limitada de ARNm convencional para establecer adicionalmente su aplicabilidad clínica. Con respecto a esto, la estabilidad de ARNm y en particular la tasa de traducción del ARNm es un parámetro esencial para aplicaciones médicas previstas, porque determina, por ejemplo, la dosificación y los intervalos de dosificación de fármacos de ARNm.

Varias estrategias han demostrado ser satisfactorias tanto para aumentar la estabilidad como para reducir la respuesta inmunogénica desencadenada por ARNm administrado a células u organismos. Entre estas, se encuentra la inclusión de nucleótidos químicamente modificados; Current Opinion in Drug Discovery and Development, 2007, 10:523. Kormann *et al.* han mostrado que la sustitución de tan sólo el 25 % de los residuos de uridina y citidina por 2-tiouridina y 5-metil-citidina es suficiente para aumentar la estabilidad de ARNm así como para reducir la activación de la inmunidad innata activada por ARNm administrado externamente *in vitro* (documentos WO2012/0195936 A1; WO2007024708 A2).

Además, se ha notificado que las regiones no traducidas (UTR) en ARNm desempeñan un papel fundamental en la regulación tanto de la estabilidad de ARNm como de la traducción de ARNm. Se sabe que las UTR influyen en el inicio de la traducción, elongación y terminación, así como la estabilización de ARNm y localización intracelular mediante su interacción con proteínas de unión a ARN (Briefings in Bioinformatics, 2000, 1:236-249 y Cold Spring Harbor Monograph Archive, 2007, 48:87-128). Dependiendo de los motivos específicos dentro de la UTR, puede o bien potenciar o bien reducir la renovación de ARNm (Cell. Mol. Life Sci., 2012, 69:3613-3634; Nucleic Acids Research, 2005, 33:D141-D146; Science, 2005, 309:1514-1518 y Current Protein & Peptide Science, 2012, 13:294-304). Recientemente, se han publicado datos sobre semividas de ARNm y las secuencias de UTR correspondientes (Nucleic Acids Research, 2011, 39:556-566 y Nucleic acids research, 37, e115).

Las UTR son secciones de una molécula de ARNm en el sentido de 5' del codón de iniciación y en el sentido de 3' del codón de terminación de un ARNm, es decir, secuencias que no se traducen. Estas regiones se transcriben con la región codificante y, por tanto, son exónicas dado que están presentes en el ARNm maduro. La UTR en el sentido de 5' del codón de iniciación de un ARNm se denomina 5'UTR y, una vez transcrita, alberga, entre otras cosas, secuencias que corresponden a partes (residuales en 3') del promotor así como una denominada secuencia de Kozak.

La secuencia consenso de Kozak, consenso de Kozak o secuencia de Kozak, es una secuencia que se sabe que se produce en ARNm de eucariotas y tiene el consenso (gcc)gccRccAUGG. La secuencia consenso de Kozak desempeña un papel principal en el inicio del proceso de traducción. A la secuencia se le puso el nombre de la persona que hizo que tuviera prominencia, Marilyn Kozak. Esta secuencia en una molécula de ARNm se reconoce por el ribosoma en el sitio de iniciación de la traducción, a partir del cual se codifica una proteína por esa molécula de ARNm. El ribosoma requiere esta secuencia, o una posible variación de la misma, para iniciar la traducción. La secuencia se identifica mediante la notación (gcc)gccRccAUGG que resume datos analizados por Kozak a partir de una amplia variedad de fuentes (aproximadamente 699 en total) de la siguiente manera: una letra minúscula designa la base más común en una posición en la que, no obstante, la base puede variar; las letras mayúsculas indican bases altamente conservadas, es decir la secuencia "AUGG" es constante o apenas cambia, si es lo hace alguna vez, "R" que indica que siempre se observa una purina (adenina o guanina) en esta posición (afirmando Kozak que adenina es más frecuente); y la secuencia entre paréntesis ((gcc)) es de significación incierta.

La secuencia consenso de Kozak se definió originalmente como ACCAUGG debido a un análisis de mutaciones

puntuales alrededor del codón de iniciación (AUG, definiendo A, en este contexto, la posición +1) en la traducción del gen de preproinsulina. Una mutagénesis más detallada de 699 ARNm de vertebrado dio como resultado la secuencia consenso GCCGCCACCAUGG, en la que la A en el sentido de 5' del codón de iniciación AUG en la posición -3 también podía ser una G (Nucleic Acids Res., 1987, 15 (20):8125-8148). Estudios sobre la traducción de preproinsulina y alfa-globina en células eucariotas revelaron que una purina (habitualmente A) en la posición -3 es esencial para una iniciación eficiente de la traducción y, si falta esta purina, una G en la posición +4 es esencial (J. Cell Biol., 1989, 108:229-41). La cantidad de proteína sintetizada a partir de una molécula de ARNm depende fuertemente de la secuencia del elemento de Kozak: el codón de iniciación AUG, que codifica para la metionina N-terminal de la proteína, es el más importante. Para un consenso fuerte, los nucleótidos en las posiciones +4 (G) y -3 (A o G) deben coincidir ambos con el consenso. Una secuencia consenso adecuada sólo tiene uno de estos dos sitios, mientras que una secuencia consenso débil no cumple con los requisitos en las posiciones +4 ni -3. Los dos residuos de citidina en -1 y -2 no están tan conservados (Cell, 1986, 44 (2):283-92), mientras que la G en la posición -6 es importante para la iniciación de la traducción (Br. J. Haematol., 2004, 124 (2):224-31).

Anteriormente se han dado a conocer secuencias de nucleótidos que tienen los números de registro de GenBank DQ215956, DQ215955, GU384682, GU384683, KJ144916, BC063266, BC004286, CX410129, JF819822, AY029533, BJ302563, BG188977, BC056709 y AU247853.

Además, anteriormente se ha descrito la traducción única de ARNm que contienen el elemento TISU (Nucleic Acids Research, 2011, 39(17):7598-7609)). Además, también se ha descrito un elemento de iniciación de la traducción específico para ARNm con una 5'UTR muy corta que también regula la transcripción (PLOS ONE, 2008, 3(8):e3094). Además, se ha descrito que ARNm con caperuza con un líder de un solo nucleótido se traduce de manera óptima en una eucariota primitiva, *Giardia lamblia* (Journal of Biological Chemistry, 2004, 279(15):14656-14664). Se ha descrito un análisis de secuencias no codificantes en 5' a partir de 699 ARN mensajeros de vertebrados (Nucleic Acids Research, 1987, 15(20):8125-8148). Además, se ha dado a conocer anteriormente la expresión de proteínas terapéuticas después de la administración de ARNm químicamente modificado en ratones (Nature Biotechnology, 2011, 29(2):154-157). Además, se han descrito anteriormente ácidos nucleicos novedosos, secuencias de polipéptido novedosas codificadas por estos ácidos nucleicos y usos de los mismos en el documento WO 02/074961. Además, se da a conocer un método de prueba para detectar artritis reumatoide y un kit para una prueba de artritis reumatoide en el documento WO 2012/081271.

Aunque en la técnica anterior ya se describen medios y métodos para aumentar la estabilidad de ARNm, reducir la respuesta inmunogénica desencadenada por ARNm administrado a células u organismos y aumentar la eficiencia de expresión (es decir, la eficiencia de transcripción y/o traducción), todavía hay una necesidad de mejoras, en particular con respecto a medios adicionales o alternativos para aumentar la eficiencia de expresión (es decir, la eficiencia de transcripción y/o traducción) dado que la eficiencia de expresión es un parámetro esencial para aplicaciones médicas previstas dado que determina, por ejemplo, la dosificación y los intervalos de dosificación de fármacos de ARNm y, en última instancia, determina la biodisponibilidad del producto final, es decir, el péptido o proteína codificado. Al mismo tiempo, hay una necesidad constante de reducir adicionalmente los costes de la producción de fármacos de ARNm, aumentar el rendimiento de las moléculas de ARNm producidas y aumentar el espacio disponible en la molécula de ARNm producida para el propio transgén, es decir, para la región codificante que codifica para un polipéptido deseado.

La presente invención se refiere a las realizaciones tal como se caracterizan en las reivindicaciones. Por tanto, se refiere a los siguientes puntos.

1. Una molécula de ADN, que puede transcribirse para dar un ARNm, que comprende una cadena con los siguientes elementos:

(a) una región codificante, que incluye un codón de iniciación en su extremo 5', que codifica para un polipéptido; y

(b) directamente en el sentido de 5' de dicha secuencia codificante, la secuencia:

R<sub>1</sub>-CNGCCACC (SEQ ID NO: 2), en la que el nucleótido N en la posición 2 de SEQ ID NO: 2 es T;

en la que R<sub>1</sub> es un promotor que se reconoce por una ARN polimerasa dependiente de ADN;

o que comprende la cadena complementaria,

en la que el promotor que se reconoce por una ARN polimerasa dependiente de ADN se selecciona del grupo que consiste en:

(i) TAATACGACTCACTATAGGGAGA (SEQ ID NO: 3) que se reconoce por una ARN polimerasa dependiente de ADN de T7;

(ii) AATTAACCCTCACTAAAGGGAGA (SEQ ID NO: 4) que se reconoce por una ARN polimerasa dependiente de ADN de T3;

- (iii) ATTTAGGTGACACTATAGAAG (SEQ ID NO: 5) que se reconoce por una ARN polimerasa dependiente de ADN de SP6; y
- (iv) AATTAGGGCACACTATAGGGA (SEQ ID NO: 6) que se reconoce por una ARN polimerasa dependiente de ADN de K11.
- 5 2. Un vector que comprende la molécula de ADN según el punto 1.
3. Una célula huésped que comprende el vector según el punto 2.
4. Una composición que comprende:
- la molécula de ADN según el punto 1, el vector según el punto 2 o la célula huésped según el punto 3.
5. Una molécula de ARN que comprende
- 10 (a) una región codificante, que incluye un codón de iniciación en su extremo 5', que codifica para un polipéptido; y
- (b) directamente en el sentido de 5' de dicha secuencia codificante, una UTR de la secuencia R<sub>2</sub>-CNGCCACC (SEQ ID NO: 2), en la que el nucleótido N en la posición 2 de SEQ ID NO: 2 es U,
- en la que R<sub>2</sub> es una secuencia de ARN correspondiente a la parte de una región de promotor que empieza con el nucleótido en el que una ARN polimerasa dependiente de ADN inicia la síntesis de ARN,
- 15 en la que R<sub>2</sub> se selecciona del grupo que consiste en:
- (i) GGGAGA (SEQ ID NO: 7);
- (ii) GGGAGA (SEQ ID NO: 8);
- (iii) GAAG (SEQ ID NO: 9); y
- (iv) GGGA (SEQ ID NO: 10); y
- 20 en la que la molécula de ARN comprende una cola de poli-A en el extremo 3';
- en la que la UTR tal como se define en (b) tiene una longitud máxima de 14 nucleótidos cuando R<sub>2</sub> es (i) o (ii); o
- en la que la UTR tal como se define en (b) tiene una longitud máxima de 12 nucleótidos cuando R<sub>2</sub> es (iii) o (iv).
6. La molécula de ARN según el punto 5, en la que la cola de poli-A tiene una longitud de al menos 120 nucleótidos.
7. Una molécula de ácido nucleico que codifica para la molécula de ARN según el punto 5 ó 6.
- 25 8. Un vector que comprende la molécula de ácido nucleico según el punto 7.
9. Una célula huésped que comprende el vector según el punto 8.
10. Una composición farmacéutica que comprende la molécula de ARN según el punto 5 ó 6, la molécula de ácido nucleico según el punto 7, el vector según el punto 8 o la célula huésped según el punto 9 y opcionalmente un portador farmacéuticamente aceptable.
- 30 11. La composición farmacéutica según el punto 10, para su uso en terapias basadas en ARN.
12. Un kit que comprende la molécula de ADN según el punto 1, la molécula de ARN según el punto 5 ó 6, la molécula de ácido nucleico según el punto 7, el vector según el punto 2 u 8 o la célula huésped según el punto 3 ó 9.
13. Uso de una UTR según el punto 5b, para traducir una región codificante de una molécula de ARN para dar un polipéptido o una proteína codificado por dicha región codificante.
- 35 En particular, la presente solicitud encontró sorprendentemente que es posible reducir el tamaño de la secuencia de UTR hasta una secuencia de "UTR mínima", reduciendo de ese modo los costes para la producción de fármacos de ARNm, aumentando el rendimiento de las moléculas de ARNm producidas y aumentando el espacio disponible en la molécula de ARNm producida para el propio transgén, es decir, para la región codificante que codifica para un polipéptido deseado. Además, al mismo tiempo, esta secuencia de UTR mínima conserva sorprendentemente, o
- 40 incluso mejora, la tasa de expresión con respecto a secuencias de UTR convencionales mientras que se ha encontrado que modificaciones en esta secuencia de UTR mínima incluso aumentan la tasa de expresión de la molécula de ARNm.

Este hallazgo conduce a proporcionar las realizaciones tal como se caracterizan en las reivindicaciones, en particular a proporcionar moléculas de ADN que permiten la producción de moléculas de ARN que albergan una

secuencia de "UTR mínima" de este tipo así como a proporcionar las moléculas de ARN correspondientes.

En un primer aspecto, se describen moléculas correspondientes a nivel de ADN mientras que más adelante, en un segundo aspecto, se describen moléculas correspondientes a nivel de ARN.

5 Por tanto, en un primer aspecto, se da a conocer una molécula de ADN, que puede transcribirse para dar un ARNm, en la que dicha molécula de ADN comprende una cadena con los siguientes elementos:

(a) una región codificante, que incluye un codón de iniciación en su extremo 5', que codifica para un polipéptido; y

(b) directamente en el sentido de 5' de dicha secuencia codificante, una secuencia seleccionada del grupo que consiste en:

(b1) R<sub>1</sub>-CGCCACC (SEQ ID NO: 1);

10 o una secuencia en la que en dicha secuencia la C en la posición 6 de SEQ ID NO: 1 está sustituida por una A y la C en la posición 7 de SEQ ID NO: 1 está sustituida por una G; y/o la A en la posición 5 de SEQ ID NO: 1 está sustituida por una G; y

(b2) R<sub>1</sub>-CNGCCACC (SEQ ID NO: 2), en la que el nucleótido N en la posición 2 de SEQ ID NO: 2 es un nucleótido seleccionado del grupo que consiste en T, G, C o A;

15 o una secuencia en la que en dicha secuencia la C en la posición 7 de SEQ ID NO: 2 está sustituida por una A y la C en la posición 8 de SEQ ID NO: 2 está sustituida por una G; y/o la A en la posición 6 de SEQ ID NO: 2 está sustituida por una G,

en la que R<sub>1</sub> es un promotor que se reconoce por una ARN polimerasa dependiente de ADN;

o que comprende la cadena complementaria.

20 Una secuencia de ADN se denomina "sentido" si su secuencia es la misma que la de una copia de ARN mensajero que se traduce para dar una proteína. La secuencia en la cadena opuesta, complementaria, se denomina secuencia "antisentido". La molécula de ADN de la presente invención se define en (a) y (b), anteriormente, mediante referencia a la cadena sentido mientras que la cadena antisentido, complementaria, correspondiente puede determinarse fácilmente por el experto dadas las reglas de apareamiento de bases.

25 La molécula de ADN de la presente invención es una molécula de ADN que puede transcribirse para dar una molécula de ARNm. La transcripción es la primera etapa de la expresión génica, en la que se copia un segmento particular de una molécula de ADN en una molécula de ARNm mediante la enzima ARN polimerasa. Durante la transcripción, se lee una secuencia de ADN por una ARN polimerasa, lo cual produce una cadena de ARN antiparalela, complementaria, denominada transcrito primario.

30 Sólo una de las dos cadenas de ADN sirve como molde para la transcripción. La cadena antisentido de ADN se lee por una ARN polimerasa dependiente de ADN desde el extremo 3' hasta el extremo 5' durante la transcripción (3' → 5'). El ARN complementario se crea en el sentido opuesto, en el sentido de 5' → 3', coincidiendo con la secuencia de la cadena sentido con la excepción del cambio de uracilo por timina. Esta direccionalidad se debe a que la ARN polimerasa sólo puede añadir nucleótidos al extremo 3' de la cadena de ARNm en crecimiento. La cadena sentido

35 no de molde de ADN se denomina cadena codificante, porque su secuencia es la misma que el transcrito de ARN recién creado (excepto por la sustitución de uracilo por timina). Esta es la cadena que se usa por convenio y en el contexto de la presente invención cuando se presenta una secuencia de ADN.

La molécula de ADN de la presente invención puede ser bicatenaria o monocatenaria o parcialmente bicatenaria y parcialmente monocatenaria.

40 Una molécula de ADN tal como se da a conocer en el presente documento comprende dos módulos principales (también denominados "puntos"), es decir, (a) una región codificante que codifica para un polipéptido y que incluye un codón de iniciación en su extremo 5', y (b) directamente en el sentido de 5' de dicha región codificante, una secuencia tal como se definió en (b1) o (b2) anteriormente en el presente documento. Una molécula de ADN de este tipo, cuando se transcribe, conduce a un ARNm con una secuencia de UTR extremadamente corta que confiere las

45 ventajas anteriormente descritas.

Además, la molécula de ADN tal como se da a conocer en el presente documento comprende preferiblemente una secuencia que, cuando se transcribe para dar ARNm, da como resultado una UTR en el sentido de 3' de la región codificante. Por tanto, la molécula de ADN de la presente invención alberga preferiblemente una región codificante así como secuencias que, tras la transcripción, dan como resultado regiones no traducidas (UTR) (en 5' y 3') en la

50 molécula de ARNm producida.

El término "región codificante que incluye un codón de iniciación en su extremo 5'" tal como se usa según la presente invención se refiere a una secuencia de ADN que está compuesta por codones, que se transcriben para

dar una molécula de ARNm por una ARN polimerasa dependiente de ADN en la que una molécula de ARNm correspondiente puede decodificarse y traducirse para dar proteínas por el ribosoma según la información proporcionada por el “código genético”. Las regiones codificantes comienzan habitualmente con un codón de iniciación en su extremo 5' y terminan con un codón de terminación. En general, el codón de iniciación es un triplete ATG (correspondiente a un triplete AUG a nivel de ARN) y el codón de terminación es TAA, TAG o TGA (correspondiente a UAA, UAG o UGA a nivel de ARN). Además de codificar para proteínas, porciones de regiones codificantes pueden servir como secuencias reguladoras en el pre-ARNm como potenciadores de corte y empalme exónicos o silenciadores de corte y empalme exónicos. La región codificante de un gen que codifica para un polipéptido o una proteína tal como se usa según la presente invención también se conoce como secuencia codificante o CDS (de secuencia de ADN codificante) y es esa porción del ADN o ARN de un gen, compuesta por exones, la que codifica para un polipéptido o proteína. La región codificante en ARNm está flanqueada por la región no traducida en 5' (5'UTR) y la región no traducida en 3' (3'UTR) que también son partes de los exones. Además, las moléculas de ARNm pueden comprender además una denominada caperuza en 5' y una cola de poli-A. La caperuza en 5', la 5'UTR, la 3'UTR y la cola de poli-A son regiones de una molécula de ARNm que no se traducen para dar una proteína.

El término “región no traducida” o “UTR” tal como se usa según la presente invención se refiere a secciones de un ARNm en el sentido de 5' del codón de iniciación y en el sentido de 3' del codón de terminación que no se traducen, y, por tanto, se denominan región no traducida en cinco prima (5'UTR) y región no traducida en tres prima (3'UTR), respectivamente. Estas regiones se transcriben con la región codificante y por tanto son exónicas dado que están presentes en el ARNm maduro.

Tal como se usa en la presente invención, la región no traducida en 3' (3'-UTR) se refiere a la sección de ARN mensajero (ARNm) que sigue inmediatamente al codón de terminación de la traducción. La 3'UTR puede comprender regiones reguladoras dentro de la región no traducida en 3' que se sabe que influyen en la poliadenilación y estabilidad del ARNm. Muchas 3'-UTR también contienen elementos ricos en AU (ARE). Además, la 3'-UTR puede contener preferiblemente la secuencia AAUAAA que dirige la adición de varios cientos de residuos de adenina denominados cola de poli(A) al extremo del transcrito de ARNm. La región no traducida en 5' (5'UTR) (también conocida como secuencia líder o ARN líder) es la región de un ARNm que está directamente en el sentido de 5' del codón de iniciación. La 5'UTR comienza en el sitio de iniciación de la transcripción y termina un nucleótido (nt) antes del codón de iniciación (habitualmente AUG en el ARNm) de la región codificante. En eucariotas, la longitud de la 5'UTR es generalmente de desde 100 hasta varios miles de nucleótidos de longitud, pero algunas veces también se producen UTR más cortas en eucariotas.

En la presente invención, la secuencia entre el promotor y la región codificante (tal como se definió en (b1) o (b2), anteriormente), es extremadamente corta y conduce, tras la transcripción, a una molécula de ARNm con una secuencia de UTR “mínima” muy corta.

Un módulo de la molécula de ADN, es decir, “una región codificante que incluye un codón de iniciación en su extremo 5' que codifica para un polipéptido” (módulo (a)) no está particularmente limitado y puede ser cualquier región codificante deseada que tiene que expresarse en una célula dada. Por tanto, este módulo puede ser una región codificante que codifica para un polipéptido deseado, es decir, el producto final deseado. La presente invención no está limitada con respecto a la “región codificante que incluye un codón de iniciación en su extremo 5' que codifica para un polipéptido” dado que la naturaleza de la región codificante depende del producto deseado que tiene que producirse en la célula. Tal región codificante también puede ser una secuencia de nucleótidos que difiere de una secuencia natural conocida y contiene mutaciones (es decir, mutaciones puntuales, mutación por inserción, deleciones y combinaciones de las mismas). Además, una región codificante de este tipo puede ser parcialmente, o en el alcance completo, una secuencia de codones optimizados derivada a partir de la secuencia natural que va a usarse como módulo (a). La optimización de codones es una técnica que maximiza la expresión de proteína aumentando la eficiencia de traducción del ARNm derivado a partir de un gen de interés. Se sabe que los genes naturales no usan los codones disponibles de manera aleatoria, sino que muestran una determinada preferencia por codones particulares para el mismo aminoácido. Por tanto, debido a la degeneración del código genético (un aminoácido puede codificarse por varios codones) transformando la secuencia de nucleótidos de un gen de interés en un conjunto de codones preferidos de la misma u otra especie.

Tal como se menciona, el módulo (a) no está particularmente limitado y puede ser cualquier región codificante deseada que tiene que expresarse en una célula dada. Por tanto, en el contexto de la presente invención, debe entenderse que “región codificante” significa cualquier molécula de poli-desoxirribonucleótido que, si se introduce en una célula, puede transcribirse para dar una molécula de ARNm que puede traducirse para dar un polipéptido/proteína o fragmento del mismo. Los términos “polipéptido” y “proteína” en este caso abarcan cualquier clase de secuencia de aminoácidos, es decir, cadenas de dos o más aminoácidos que están unidos, cada uno, mediante enlaces peptídicos y también incluyen péptidos y proteínas de fusión.

En una realización preferida, la “región codificante que incluye un codón de iniciación en su extremo 5' que codifica para un polipéptido” contiene una secuencia de desoxirribonucleótidos que codifica para un polipéptido/proteína o fragmento del mismo cuya función en la célula o en las inmediaciones de la célula se necesita o es beneficiosa, por ejemplo, una proteína cuya falta o forma defectuosa es un factor desencadenante para una enfermedad o una

afección, cuya provisión puede moderar o prevenir una enfermedad o una afección, o una proteína que puede fomentar un proceso que es beneficioso para el organismo, en una célula o sus inmediaciones. La región codificante puede contener la secuencia para la proteína completa o una variante funcional de la misma. Además, la secuencia de desoxirribonucleótidos de la región codificante puede codificar para una proteína que actúa como factor, inductor, regulador, estimulador o enzima, o como fragmento funcional del mismo, en la que esta proteína es una cuya función es necesaria con el fin de remediar un trastorno, en particular un trastorno metabólico o con el fin de iniciar procesos *in vivo* tales como la formación de nuevos vasos sanguíneos, tejidos, etc. En este caso, se entiende que variante funcional significa un fragmento que, en la célula, puede emprender la función de la proteína cuya función en la célula se necesita o cuya falta o forma defectuosa es patógena.

En una realización preferida, la “región codificante que incluye un codón de iniciación en su extremo 5’ que codifica para un polipéptido” codifica para un polipéptido o proteína terapéutica o farmacéuticamente activo que tiene un efecto terapéutico o preventivo. Como tal, la molécula de ADN de la presente invención que puede transcribirse para dar un ARNm que comprende dicha “región codificante que incluye un codón de iniciación en su extremo 5’ que codifica para un polipéptido” puede usarse en terapia de ácido nucleico y aplicaciones relacionadas. En este contexto, según la invención, la transcripción y traducción de una molécula de ADN de la presente invención para dar un ARNm y adicionalmente para dar un polipéptido o una proteína puede estar destinada a compensar o complementar la expresión de genes endógenos, en particular en casos en los que un gen endógeno es defectuoso o está silenciado, conduciendo a un producto ausente, insuficiente o defectuoso o disfuncional de expresión génica tal como es el caso con muchas enfermedades metabólicas o hereditarias tales como fibrosis quística, hemofilia o distrofia muscular por nombrar unas pocas. La transcripción y traducción de una molécula de ADN de la presente invención para dar un ARNm y adicionalmente para dar un polipéptido o una proteína también puede estar destinada a hacer que el producto de la expresión interaccione o interfiera con cualquier proceso celular endógeno tal como la regulación de la expresión génica, transducción de señales y otros procesos celulares. La transcripción y traducción de una molécula de ADN de la presente invención para dar un ARNm y adicionalmente para dar un polipéptido o una proteína también puede estar destinada a dar lugar a una respuesta inmunitaria en el contexto del organismo en el que reside o se hace que resida una célula transfectada o transducida. Los ejemplos son la modificación genética de células presentadoras de antígenos tales como células dendríticas con el fin de hacer que presenten un antígeno con fines de vacunación. Otro ejemplo es la transcripción y traducción de una molécula de ADN de la presente invención para dar un ARNm y adicionalmente para dar un polipéptido o una proteína en la que dicha región codificante codifica para citocinas. Esto puede ser deseable, por ejemplo, en tumores con el fin de provocar una respuesta inmunitaria específica de tumor. Además, la transcripción y traducción de una molécula de ADN de la presente invención para dar un ARNm y adicionalmente para dar un polipéptido o una proteína también puede estar destinada a generar *in vivo* o *ex vivo* células genéticamente modificadas de manera transitoria para terapias celulares tales como células T modificadas o células precursoras o madre u otras para medicina regenerativa.

En otras realizaciones preferidas, la “región codificante que incluye un codón de iniciación en su extremo 5’ que codifica para un polipéptido” puede codificar para una proteína que desempeña un papel en procesos de crecimiento y angiogénesis, que son necesarios, por ejemplo, en la regeneración controlada y pueden formarse entonces específicamente mediante introducción de la molécula de ARN según la invención. Esto puede ser útil, por ejemplo, en procesos de crecimiento o para el tratamiento de defectos óseos, defectos tisulares y en el contexto de implantación y trasplantes.

Tal como se menciona, la molécula de ADN y, en particular, la molécula de ARN transcrita de manera correspondiente de la presente invención que comprende una “región codificante que incluye un codón de iniciación en su extremo 5’ que codifica para un polipéptido” puede usarse de manera apropiada en cualquier caso en el que tiene que proporcionarse al organismo un polipéptido o una proteína, que estaría presente de manera natural en el organismo pero que no está presente o está presente en forma deficiente o en una cantidad demasiado pequeña debido a enfermedades o defectos génicos. Se conocen proteínas y los genes que codifican para las mismas, cuya deficiencia o defecto está vinculado a una enfermedad. La versión intacta respectiva de la región codificante que codifica para el polipéptido o proteína intacto puede usarse según la presente invención.

Se conocen numerosos trastornos genéticos, provocados por la mutación de un único gen, y son candidatos para enfoques terapéuticos de ARNm. Los trastornos provocados por mutaciones en un único gen, tales como fibrosis quística, hemofilia y muchos otros, pueden ser dominantes o recesivos con respecto a la probabilidad de que aparezca un determinado rasgo en la progenie. Mientras que un alelo dominante manifiesta un fenotipo en individuos que sólo tienen una copia del alelo, para un alelo recesivo el individuo debe tener dos copias, una de cada progenitor, para manifestarse. En cambio, los trastornos poligénicos están provocados por dos o más genes y la manifestación de la enfermedad respectiva es con frecuencia fluida y está asociada a factores del entorno. Los ejemplos de trastornos poligénicos son hipertensión, nivel aumentado de colesterol, cáncer, trastornos neurodegenerativos, enfermedad mental y otros. También en estos casos, un ARNm terapéutico que representa uno o más de estos genes puede ser beneficioso para esos pacientes. Además, un trastorno genético puede no haberse transmitido por los genes de los progenitores, sino que puede estar provocado por nuevas mutaciones. También en estos casos, un ARNm terapéutico que representa la secuencia génica correcta puede ser beneficioso para los pacientes.

Un catálogo en línea que tiene actualmente 22.993 entradas de genes humanos y trastornos genéticos junto con sus

genes respectivos y una descripción de sus fenotipos está disponible en la página web de ONIM (Online Mendelian Inheritance in Man) (<http://onim.org>); secuencias de cada uno están disponibles a partir de la base de datos de Uniprot (<http://www.uniprot.org>). Como ejemplos no limitativos, la siguiente tabla 1 indica algunas enfermedades congénitas y el/los gen(es) correspondiente(s). Debido al alto grado de interacción de rutas de señalización celular, la mutación de un determinado gen provoca una multitud de síntomas patógenos, de los cuales sólo se indica uno característico en la tabla 1.

La proteína terapéutica puede elegirse de las proteínas celulares indicadas en la tabla 1. Por tanto, la molécula de ADN puede codificar para una proteína celular terapéutica, en la que la proteína terapéutica codificada es una indicada en la tabla 1 o un homólogo de la misma.

La proteína terapéutica puede elegirse de las proteínas secretadas indicadas en la tabla 1. Por tanto, la molécula de ADN puede codificar para una proteína de fusión terapéutica, en la que la proteína terapéutica codificada o un homólogo de la misma es una indicada en la tabla 1 y la segunda proteína es un péptido señal que permite la secreción de la proteína terapéutica. Un péptido señal es una secuencia de aminoácidos corta, normalmente de 5-30 aminoácidos de longitud, presente en el extremo N-terminal de dicha proteína terapéutica y que lleva la proteína de fusión hacia la ruta secretora de la célula a través de determinados orgánulos (es decir, el retículo endoplasmático, el aparato de Golgi o los endosomas). Por tanto, tal proteína de fusión se secreta a partir de la célula o a partir de un orgánulo celular o se inserta en una membrana celular (por ejemplo, proteínas transmembrana de extensión múltiple) en un compartimento celular o en la superficie de la célula.

Por tanto, la “región codificante que incluye un codón de iniciación en su extremo 5’ que codifica para un polipéptido” (módulo (a)) puede codificar para, pero no se limita a, los siguientes genes que provocan, predisponen o protegen frente a enfermedades. Los ejemplos no limitativos de tales trastornos que pueden tratarse (o prevenirse) incluyen aquellos en los que dicho polipéptido, proteína o péptido se selecciona del grupo que consiste en los expuestos en la siguiente tabla 1.

La “región codificante que incluye un codón de iniciación en su extremo 5’ que codifica para un polipéptido” puede transcribirse y traducirse para dar una proteína de longitud parcial o completa que comprende actividad celular a un nivel igual o superior al de la proteína nativa. En algunas realizaciones, la “región codificante que incluye un codón de iniciación en su extremo 5’ que codifica para un polipéptido” codifica para un polipéptido, proteína o péptido terapéutico o farmacéuticamente activo que tiene un efecto terapéutico o preventivo, en el que dicho polipéptido, proteína o péptido se selecciona del grupo que consiste en los expuestos en la siguiente tabla 1. La “región codificante que incluye un codón de iniciación en su extremo 5’ que codifica para un polipéptido” puede usarse para expresar una proteína de longitud parcial o completa con actividad celular a un nivel igual o inferior al de la proteína nativa. Esto puede permitir el tratamiento de enfermedades para las que puede estar indicada la administración de una molécula de ARN.

Tabla 1: Ejemplos no limitativos de genes humanos y trastornos genéticos

Enfermedad	Patología	Gene, herencia
Enfermedades de la sangre		
Anemia de Fanconi	Anemia y neutropenia, evidencia de que se ve afectado un mecanismo de reparación del ADN	FANCA, autosómica, recesiva
Hemofilia A	Hemorragia anómala	Factor VIII de coagulación, cromosoma X, recesiva
Hemofilia B	Hemorragia anómala	Factor IX de coagulación, cromosoma X, recesiva
Esferocitosis hereditaria (diversos tipos)	Eritrocitos con forma esférica (esferocitos)	Anquirina (ANK1)
Hemoglobinuria paroxística nocturna	Anemia y presencia de sangre en la orina	PIG-A, cromosoma X
Porfiria cutánea tardía	Sobreproducción de hemo, sobrecarga de hierro	Uroporfirinógeno descarboxilasa (UROD), autosómica, recesiva
Inmunodeficiencia combinada grave (SCID)	Debido a síntesis de ADN alterada, deficiencia inmunitaria grave en la inmunidad humoral y celular	Adenosina desaminasa, autosómica, recesiva, IL-2R- $\gamma$ , JAK3, (IL-7R- $\alpha$ , RAG1/2, Artemis, CD3 $\delta$ , CD3 $\epsilon$
Anemia de células falciformes	Hemoglobina anómala (HbS)	$\beta$ -Hemoglobina (HB), autosómica, recesiva
Talasemia (forma $\alpha$ y $\beta$ )	Falta de hemoglobina $\alpha$ o $\beta$ dando como resultado anemia	Deleción de HBA1 y/o HBA2,
Enfermedad de Von Willebrand (tres tipos)	Hemorragia anómala, hemorragia similar a hemofilia A y B	Formas autosómicas dominante y recesiva



ES 2 835 051 T3

conocidos, el tipo III es el más grave)		
Cáncer		
Melanoma maligno	Mutación de P16 conduce a proliferación descontrolada de fibroblastos	Inhibidor de cinasa dependiente de ciclina 2 (CDKN2)
Neurofibromatosis (2 tipos)	Tumores benignos en nervios auditivos conduce a sordera	NF1, NF2, autosómica, dominante
Sordera (oído)		
Sordera	Pérdida de audición	Sordera 1A (DFNB1), autosómica, recesiva
Síndrome de Pendred	Pérdida de audición	Pendrin (PDS), autosómica, recesiva
Corazón		
Ataxia telangiectasia	Reparación alterada del daño de ADN	ATM,
Aterosclerosis	Aumento de colesterol en sangre	apoE,
Síndrome de LQT (QT prolongado)	Defecto de canales de potasio	LQT1 y otros genes
Síndrome de Von-Hippel Lindau	Crecimiento anómalo de vasos sanguíneos, puede conducir a cáncer	VHL, autosómica, dominante
Síndrome de Williams Beuren	Deleción de elastina da como resultado defectos vasculares, estenosis aórtica supraavalvular	Deleción de genes de elastina y LIM cinasa
Trastornos metabólicos y enfermedades de almacenamiento de glucógeno		
Adrenoleucodistrofia	Transporte y metabolismo de ácidos grasos alterados	ABCD1, cromosoma X
Alcaptonuria	Defecto de metabolismo de nitrógeno, la orina se vuelve oscura cuando se expone a oxígeno	Oxidasa homogentísica, autosómica, recesiva
Diabetes tipo I	Producción de insulina alterada	IDDM1, IDDM2, GCK, ...
Galactosemia	Trastorno del metabolismo de galactosa	Gen de galactosa-1-fosfato uridiltransferasa (GALT), autosómica, recesiva
Enfermedad de Gauche	Alteración del metabolismo de grasas	Glucocerebrosidasa
Mala absorción de glucosa galactosidasa	Transporte alterado de glucosa y galactosa fuera de la luz intestinal que da como resultado diarrea	SGLT1, autosómica, recesiva
Enfermedad de almacenamiento de glucógeno tipo I, enfermedad de Von-Gierke	Acumulación de glucosa en el hígado y el riñón	Glucosa-6-fosfatasa, autosómica, recesiva
Enfermedad de almacenamiento de glucógeno tipo II, enfermedad de Pompe	Acumulación de glucógeno en el hígado, corazón, músculo esquelético, cardiomegalia	$\alpha$ -1-Glucosidasa, autosómica, recesiva
Enfermedad de almacenamiento de glucógeno tipo III, enfermedad de Cori	Acumulación de glucógeno en el hígado, corazón, músculo esquelético, hepatomegalia	Enzima de desramificación, autosómica, recesiva
Enfermedad de almacenamiento de glucógeno tipo V, enfermedad de McArdle	No puede usar glucógeno en células musculares	Fosforilasa muscular, autosómica, recesiva
Glucosa-6-fosfato deshidrogenasa	Incapacidad para mantener glutatión conduce a anemia hemolítica	G6PD, cromosoma X, recesiva
Hemocromatosis hereditaria (4 tipos)	Exceso de hierro en el organismo (esp. el hígado) debido a absorción excesiva de hierro en el intestino	Hemocromatosis (HFE)

ES 2 835 051 T3

Homocistinuria	Defecto del metabolismo de nitrógeno	Defecto de cistación sintetasa, autosómica, recesiva
Síndrome de Lesh Nyhan	Acumulación de ácido úrico que conduce a gota, piedras de ureato y pérdida muscular	HPRT1, cromosoma X
Cetoaciduria de cadena ramificada	Defecto del metabolismo de aminoácidos conduce a la acumulación de $\alpha$ -cetoácidos y muerte en los primeros meses si no se trata	alfa-deshidrogenasa de cadena ramificada (BCKDH)
Síndrome de Menkes	Capacidad reducida de absorber cobre, conduce a muerte en la infancia si no se trata	ATP7A, cromosoma X, recesiva
Obesidad	Aumento del peso corporal	Poligénico, niveles elevados de leptina pueden desempeñar un papel
Fenilcetonuria	Incapacidad para descomponer fenilalanina para dar tirosina conduce a retraso mental	Fenilalanina hidroxilasa (PAH), autosómica, recesiva
Enfermedad de Tangier	Niveles reducidos de lipoproteínas de alta densidad en plasma	Gen de casete de unión a ATP 1 (ABCA1)
Síndrome de Zellweger (conduce a muerte en lactantes)	Altos niveles de hierro y cobre en la sangre	PXR1 (receptor en la superficie de peroxisomas)
Enfermedad de Wilsons	Acumulación de cobre en el cerebro y el hígado	ATP7B (ATPasa tipo P), autosómica, recesiva
<b>Sistema musculoesquelético</b>		
Acondroplasia	Talla baja con cabeza grande debido a una lenta proliferación de condrocitos	Receptor de factor de crecimiento de fibroblastos 3 (FGF3R),
Síndrome de Charcot-Marie-Tooth y su forma más grave síndrome de Dejerine-Sottas	Degeneración de los músculos en las articulaciones	Diferentes formas provocadas por diferentes mutaciones génicas, autosómica, recesiva y de cromosoma X
Síndrome de Cockayne (2 tipos)	Envejecimiento prematuro y talla baja, pérdida de reparación de ADN "sobre la marcha"	Proteína de reparación por escisión complementaria cruzada de grupo 8 (ERCC8)
Displasia condroectodérmica	Malformación de huesos y polidactilia	EVC, autosómica, recesiva
Displasia diastrófica (DTD)	Malformación en las manos, defecto de transportador de sulfato	Gen de DTDST
Distrofia muscular de Duchenne	Agrandamiento de tejido muscular con consiguiente pérdida de función	DMD, cromosoma X, recesiva
Fibrodysplasia osificante progresiva	Formación de huesos heterotópicos	NOG, BMP, autosómica, dominante
Ataxia de Friedreich	Agrandamiento del corazón y pérdida progresiva de coordinación muscular	Frataxina, autosómica, recesiva
Hipofosfatasa	Producción de una versión anómala de fosfatasa alcalina que afecta al proceso de mineralización	ALPL, autosómica, recesiva
Síndrome de Marfan	Trastorno de tejido conjuntivo debido a deficiencia de fibrilina	Fibrilina 1 (FBN), autosómica, dominante
Distrofia miotónica (aparición durante la juventud)	Defecto de proteína cinasa en células de músculo esquelético	Proteína cinasa de distrofia miotónica (DMPK), autosómica, dominante
Osteogénesis imperfecta (diversos tipos)	Defecto en la formación de colágeno tipo I conduce a múltiples fracturas después del nacimiento	COL1A1, COL1A2
Síndrome de Prader-Willi	Reducción del tono muscular y retraso mental	SNRPN (ribonucleoproteína pequeña N) delecionada debido a una deleción en el cromosoma 15
<b>Neuronas y cerebro</b>		
Enfermedad de Alzheimer	Aumento de la producción de amiloide, incapacidad progresiva para recordar	Poligénico, PS1, PS2, ...

	hechos	
Esclerosis lateral amiotrófica (ALS) (diversas formas)	Degeneración progresiva de células neuronales motoras (defecto en la eliminación radicales de superóxido)	Superóxido dismutasa 1 (SOD1), diversos genes implicados
Síndrome de Angelman	Retraso mental con risas inadecuadas	Impresión genómica en el cromosoma 15
Piruvato deshidrogenasa	Defectos neurológicos si no se trata	Piruvato deshidrogenasa, autosómica, recesiva
Enfermedad de Refsum	Acumulación de ácido fitánico conduce a neuropatía periférica	Pitavil-CoA hidroxilasa (PHYH), autosómica, recesiva
Síndrome de Rett	Retraso mental con interrupción del desarrollo entre 6 y 18 meses de edad	Proteína de unión a metil-CpG 2 (MECP2), cromosoma X, dominante
Enfermedad de Tay-Sachs (diversas formas de gravedad)	Descomposición alterada del gangliósido GM2 conduce a daño neurológico	HEXA ( $\beta$ -hexosaminidasa A), autosómica, recesiva
Enfermedad de LaFora	Forma agresiva de epilepsia	EPM2A, autosómica, recesiva
Temblor esencial (formas variables)	Temblores incontrolables	ETM1, ETM2, autosómica, dominante
Síndrome del cromosoma X frágil	Falta de proteína de unión a ARN de FMR1, retraso mental	El gen de FMR1 no se expresa debido a una amplificación de CGG en la región 5'UTR
Enfermedad de Huntington	Demencia progresiva con aparición en la edad adulta	HTT (huntingtina), autosómica, dominante
Intestino		
Síndrome de Bartter (3 tipos)	Enfermedad renal	Gen de canal de cloruro de B (CLCNKB) del riñón, autosómica, recesiva
Enfermedad de riñón poliquístico (2 tipos)	Enfermedad renal	PDK1, PDK2, autosómica, dominante, también se conoce una forma autosómica, recesiva (ARPKD)
Pulmón		
Alfa-1-antitripsina	Defecto de alveolos debido a liberación descontrolada de elastasa	SERPINA1, autosómica, codominante
Asma	Trastorno inflamatorio crónico de las vías respiratorias	Poligénico
Fibrosis quística	Mucosa excesivamente viscosa debido al transporte defectuoso de iones Cl <sup>-</sup>	CFTR (regulador de conductancia de membrana de fibrosis quística), autosómica, recesiva
Disfunción del metabolismo de tensoactivos (diversos tipos)	Recién nacidos con peso corporal normal, pero todos ellos que no logran inflar	Transportador de casete de unión a ATP (ABCA3)
Discinesia ciliar primaria	Mucosa excesivamente viscosa debido a función de cilios defectuosa/ausente	DNAI1, CCNO, CCDC40, entre otros
Enfermedades de almacenamiento lisosómico		
Enfermedad de Fabry	Más que otros, lesiones cutáneas debido a la acumulación de trihexósido de ceramida	$\alpha$ -Galactosidasa A, cromosoma X, recesiva
Enfermedad de Gaucher tipo-I: forma adulta (esperanza de vida normal con tratamiento), tipo-II: forma infantil (muerte antes de 1 año de edad), tipo III: forma juvenil (comienzo al inicio de la infancia, menos grave que el tipo II)	Acumulación de glucocerebrósidos (gangliósidos, esfingolípidos)	Glucocerebrosidasa, autosómica, recesiva,
Síndrome de Hunter	Acumulación de mucopolisacáridos	L-iduronosulfato sulfatasa, cromosoma X,

		recesiva
Síndrome de Hurler (muerte antes de los 10 años de edad)	Acumulación de mucopolisacáridos	$\alpha$ -L-iduronidasa, autosómica, recesiva
Enfermedad de Niemann-Pick (tres formas distintas A, B, C)	Defecto en la liberación de colesterol a partir de lisosomas, acumulación de esfingomielinina	Esfingomielinasa, autosómica, recesiva
Enfermedad de Tay-Sachs (muerte antes de 4 años de edad)	Acumulación de gangliósido G <sub>M2</sub> en células neuronales	Hexosaminidasa A, autosómica, recesiva
<b>Piel</b>		
Albinismo	Defecto de metabolismo del nitrógeno	Deficiencia de tirosinasa, autosómica, recesiva
Albinismo, oculocutáneo, tipo II	Biosíntesis reducida de pigmento melanina	OCA2, autosómica, recesiva
Síndrome de Ehlers-Danlos (diversos tipos)	Hernia diafragmática, habitual, desprendimiento de retina	Diversos defectos en la síntesis de colágeno
Epidermolisis ampollosa (diversos tipos incluyendo EB simple, EB juntural, EB distrófica y síndrome de Kindler)	Defectos en el mantenimiento de la estabilidad estructural de queratinocitos o adhesión del queratinocito a la dermis subyacente	Epidermolisis ampollosa macular tipo (EBM), epidermolisis ampollosa 3 progresiva (EBR3), epidermolisis ampollosa 4 pseudojuntural (EBR4), desmoplacina (DSP), placofilina 1 (PKP1), creatina (KRT5, KRT14), plectina (PLEC), ITGA6, subunidad de integrina (ITGB4), subunidades de laminina (LAMA3, LAMP3, LAMB3, LAMC2), colágeno (COL17A1, COL7A1 (autosómica, dominante), FERMT1, autosómica, recesiva
Enfermedad de Hartnup	Defecto en la captación de triptófano en el tracto gastrointestinal, piel sensible a la luz	SLC6A19, autosómica, recesiva
Telangiectasia hemorrágica hereditaria, síndrome de Osler-Weber-Rendu	Telangiectasia de la piel y membranas mucosas	Endoglina (ENG), autosómica, dominante
Hipercolesterolemia, familiar	Aumento del colesterol en suero unido a lipoproteína de baja densidad, acumulación en la piel y arteriosclerosis	Receptor de lipoproteínas de baja densidad (LDLR), apolipoproteína B (APOB), autosómica, dominante
Xerodermia pigmentosa	Defecto en la piel y melanoma debido a exposición a UV	Defecto de reparación del ADN, autosómica, recesiva
Calvicie de patrón masculino	Conversión alterada de testosterona en dihidrotestosterona en la piel	5- $\alpha$ -reductasa
<b>Enfermedades hepáticas genéticas</b>		
Trastornos del metabolismo de los aminoácidos	Alteraciones en el proceso de múltiples etapas que descompone el aminoácido tirosina y fenilalanina	FAH, TAT, HPD, autosómica, recesiva
Beta-talasemia intermedia	Escasez de glóbulos rojos maduros	HBB, autosómica, recesiva
Síndrome de Crigler-Najjar	Deficiencia en la glucuronidación en la que la bilirrubina se vuelve soluble en agua	UGT1A1, autosómica, recesiva
Trastornos de la oxidación de ácidos grasos	Deficiencia en el procesamiento de ácidos grasos de cadena larga y ácidos grasos de cadena muy larga dando como resultado apatía e hipoglucemia	HADHA, ACADVL, autosómica, recesiva
Trastornos del metabolismo de la fructosa	Gluconeogénesis alterada que provoca hipoglucemia	FBP1, ALDOB, autosómica, recesiva
Galactosemia	Deficiencia en el procesamiento de galactosa	GALT, GALK1, GALE, autosómica, recesiva

Enfermedades de almacenamiento de glucógeno	Descomposición alterada de 6-fosfato de glucosa y glucógeno conduce a acumulación de glucógeno así como moléculas de glucógeno anómalas que provocan daño celular	G6PC, SLC37A4, AGL, GBE1, autosómica, recesiva
Trastorno de la biosíntesis de hemo	Disminución de uroporfirinógeno descarboxilasa dando como resultado acumulación de compuestos denominados porfirinas que provocan niveles tóxicos en el hígado	UROD, autosómica, dominante, ALAS2, ligada al cromosoma X, dominante, ALAD, autosómica, recesiva
Trastornos del metabolismo (transporte) de lípidos	Escasez de proteína funcional, que impide el movimiento de colesterol y otros lípidos, conduciendo a su acumulación en células	NPC1, NPC2, autosómica, recesiva, LDLR, autosómica, dominante
Trastornos del metabolismo de metales	Trastornos en el almacenamiento y transporte de hierro y cobre dando como resultado acumulación en tejidos y órganos	ATP7B, HAMP, HFE, HFE2, autosómica, recesiva
Trastornos de ácidos orgánicos (acidurias / acidemias)	Descomposición alterada de varios elementos estructurales de proteínas (aminoácidos), determinados lípidos y colesterol	BCKDHA, BCKDHB, y DBT, PCCA y PCCB, MUT, MMAA, MMAB, MMADHC, MCEE, IVD, MCCC1 o MCCC2, autosómica, recesiva
Hiperoxaluria primaria tipo 1	Descomposición alterada de glioxilato conduciendo a daño renal	AGXT, GRHPR, autosómica, recesiva
Colestasis intrahepática familiar progresiva	Acumulación de ácidos biliares en células hepáticas provocando daño hepático	ATP8B1, autosómica, recesiva
Trastorno de la actividad de trombocitos	Falta de actividad enzimática altera el equilibrio habitual entre hemorragia y coagulación	ADAMTS13, autosómica, recesiva
Trastornos del ciclo de urea	Trastorno del ciclo de urea que provoca una forma de hiperamonemia	OTC (trastorno ligado al cromosoma X), CPS1, ASS1 y SLC25A13, ASL, autosómica, recesiva

La tabla 1 anterior muestra ejemplos de genes en los que un defecto conduce a una enfermedad que puede tratarse con la molécula de ARN transcrita a partir de la molécula de ADN de la presente invención, en los que la molécula de ADN (y la molécula de ARN transcrita de manera correspondiente) comprende una "región codificante que incluye un codón de iniciación en su extremo 5' que codifica para un polipéptido" que codifica para una versión intacta de la proteína o un fragmento funcional de la misma del gen defectuoso anteriormente dado a conocer.

Pueden mencionarse enfermedades hereditarias que afectan, por ejemplo, a los pulmones, tales como deficiencia de SPB (proteína tensoactiva B), deficiencia de ABCA3, fibrosis quística y deficiencia de  $\alpha$ 1-antitripsina, o que afectan a proteínas plasmáticas (por ejemplo hemocromatosis congénita (deficiencia de hepcidina), púrpura trombocitopénica trombótica (TPP, deficiencia de ADAMTS 13) y provocan defectos de coagulación (por ejemplo hemofilia a y b) y defectos del complemento (por ejemplo deficiencia de proteína C), defectos inmunitarios tales como por ejemplo SCID (provocado por mutaciones en diferentes genes tales como: RAG1, RAG2, JAK3, IL7R, CD45, CD3 $\delta$ , CD3 $\epsilon$ ) o por deficiencias debidas a falta de adenosina desaminasa por ejemplo (ADA-SCID), granulomatosis séptica (por ejemplo provocada por mutaciones del gen gp-91-phox, el gen p47-phox, el gen p67-phox o el gen p33-phox) y enfermedades de almacenamiento tales como enfermedad de Gaucher, enfermedad de Fabry, enfermedad de Krabbe, MPS I, MPS II (síndrome de Hunter), MPS VI, enfermedad de almacenamiento de glucógeno tipo II o mucopolisacaridosis.

Otros trastornos para los que puede ser útil la presente invención que comprende una "región codificante que incluye un codón de iniciación en su extremo 5' que codifica para un péptido" incluyen trastornos tales como atrofia muscular espinal relacionada con SMN1 (SMA); esclerosis lateral amiotrófica (ALS); galactosemia relacionada con GALT; fibrosis quística (CF); trastornos relacionados con SLC3A1 incluyendo cistinuria; trastornos relacionados con COL4A5 incluyendo síndrome de Alport; deficiencias de galactocerebrosidasa; adrenomieloparpatía y adrenoleucodistrofia ligada al cromosoma X; ataxia de Friedreich; enfermedad de Pelizaeus-Merzbacher; esclerosis tuberosa relacionada con TSC1 y TSC2; síndrome de Sanfilippo B (MPS IIIB); cistinosis relacionada con CTNS; los trastornos relacionados con FMR1 que incluyen síndrome del cromosoma X frágil, síndrome de temblores/ataxia ligado al cromosoma X frágil y síndrome de insuficiencia ovárica prematura del cromosoma X frágil; síndrome de Prader-Willi; telangiectasia hemorrágica hereditaria (AT); enfermedad de Niemann-Pick tipo C1; las enfermedades relacionadas con lipofuscinosis neuronales ceroides incluyendo lipofuscinosis neuronal ceroides juvenil (JNCL), enfermedad de Batten juvenil, enfermedad de Santavuori-Haltia, enfermedad de Jansky-Bielschowsky, y deficiencias de PTT-1 y TPP1; ataxia infantil relacionada con EIF2B1, EIF2B2, EIF2B3, EIF2B4 y EIF2B5 con hipomielinización del sistema nervioso central/sustancia blanca evanescente; ataxia episódica tipo 2 relacionada con CACNA1A y CACNB4; los trastornos relacionados con MECP2 incluyendo síndrome de Rett clásico, síndrome de PPM-X y

encefalopatía neonatal grave relacionada con MECP2; síndrome de Rett atípico relacionado con CDKL5; enfermedad de Kennedy (SBMA); arteriopatía autosómica dominante cerebral relacionada con Notch-3 con infartos subcorticales y leucoencefalopatía (CADASIL); trastornos de convulsiones relacionados con SCN1A y SCN1B; los trastornos relacionados con polimerasa G que incluyen síndrome de Alpers-Huttenlocher, neuropatía atáxica sensorial relacionada con POLG, disartria y oftalmoparesia, y oftalmoplegia externa progresiva autosómica dominante y recesiva con deleciones de ADN mitocondrial; hipoplasia suprarrenal ligada al cromosoma X; agammaglobulinemia ligada al cromosoma X; enfermedad de Fabry; y enfermedad de Wilson.

En todas estas enfermedades, una proteína, por ejemplo una enzima, es defectuosa, lo cual puede tratarse mediante tratamiento con el ARN transcrito a partir de la molécula de ADN tal como se da a conocer en el presente documento, lo cual hace que la proteína codificada por el gen defectuoso o un fragmento funcional de la misma esté disponible. Las terapias de sustitución de transcrito/terapias de sustitución de enzima no afectan al defecto genético subyacente, sino que aumentan la concentración de la enzima para la que el paciente presenta deficiencia. Como ejemplo, en la enfermedad de Pompe, la terapia de sustitución de transcrito/terapia de sustitución de enzima sustituye la enzima lisosómica deficiente alfa-glucosidasa ácida (GAA).

Por tanto, ejemplos no limitativos de proteínas que pueden codificarse por la “región codificante que incluye un codón de iniciación en su extremo 5’ que codifica para un polipéptido” del módulo (a) son eritropoyetina (EPO), hormona del crecimiento (somatotropina, hGH), regulador de la conductancia transmembrana de fibrosis quística (CFTR), factores de crecimiento tales como GM-SCF, G-CSF, MPS, proteína C, hepcidina, ABCA3 y proteína tensioactiva B. Ejemplos adicionales de enfermedades que pueden tratarse con el ARN tal como se da a conocer en el presente documento son hemofilia A/B, enfermedad de Fabry, CGD, ADAMTS13, enfermedad de Hurler, A- $\gamma$ -globulinemia mediada por el cromosoma X, inmunodeficiencia relacionada con adenosina desaminasa y síndrome de dificultad respiratoria en el recién nacido, que está asociado con SP-B. De manera particularmente preferible, la “región codificante que incluye un codón de iniciación en su extremo 5’ que codifica para un polipéptido” de la molécula de ADN contiene la secuencia para proteína tensioactiva B (SP-B) o para eritropoyetina. Ejemplos adicionales de proteínas que pueden codificarse por la “región codificante que incluye un codón de iniciación en su extremo 5’ que codifica para un polipéptido” de la molécula de ADN según la invención son factores de crecimiento tales como hormona del crecimiento humana hGH, BMP-2 o factores de angiogénesis.

Alternativamente, los ácidos nucleicos pueden codificar para anticuerpos de longitud completa o anticuerpos más pequeños (por ejemplo, cadenas tanto pesadas como ligeras) para conferir inmunidad a un sujeto.

La “región codificante que incluye un codón de iniciación en su extremo 5’ que codifica para un polipéptido” puede codificar para un anticuerpo monoclonal o policlonal funcional, que puede ser útil para seleccionar como diana y/o inactivar una diana biológica (por ejemplo, una citocina estimulante tal como factor de necrosis tumoral). De manera similar, la “región codificante que incluye un codón de iniciación en su extremo 5’ que codifica para un polipéptido” puede codificar, por ejemplo, para anticuerpos anti-factor nefrótico funcionales útiles para el tratamiento de glomerulonefritis membranoproliferativa tipo II o síndrome urémico hemolítico agudo, o alternativamente puede codificar para anticuerpos anti-factor de crecimiento endotelial vascular (VEGF) útiles para el tratamiento de enfermedades mediadas por VEGF, tales como cáncer.

El módulo (a), es decir, la “región codificante que incluye un codón de iniciación en su extremo 5’ que codifica para un polipéptido”, puede ser una región codificante que codifica para un polipéptido o una proteína que puede usarse en tecnologías de edición del genoma. La edición del genoma es un tipo de ingeniería genética en el que se inserta, delecióna o sustituye ADN en el genoma de un organismo usando nucleasas. Estas nucleasas crean roturas bicatenarias (DSB) específicas del sitio en ubicaciones deseadas en el genoma. Las roturas bicatenarias inducidas se reparan mediante unión de extremos no homólogos o recombinación homóloga, dando como resultado mutaciones dirigidas en el genoma, “editando” de ese modo el genoma. En la técnica se conocen numerosos sistemas de edición del genoma que usan diferentes polipéptidos o proteínas, es decir, por ejemplo, el sistema de CRISPR-Cas, meganucleasas, nucleasas de dedos de cinc (ZFN) y nucleasas basadas en efector similar a activador de transcripción (TALEN). Se revisan métodos para modificación por ingeniería del genoma en Trends in Biotechnology, 2013, 31 (7), 397-405.

Por tanto, la “región codificante que incluye un codón de iniciación en su extremo 5’ que codifica para un polipéptido” puede contener una secuencia de desoxirribonucleótidos que codifica para un polipéptido o una proteína de la familia de proteínas Cas (proteína asociada a CRISPR), preferiblemente Cas9 (proteína asociada a CRISPR 9). Pueden usarse proteínas de la familia de proteínas Cas, preferiblemente Cas9, en métodos basados en CRISPR/Cas9 y/o tecnologías de edición del genoma de CRISPR/Cas9. Se revisan sistemas de CRISPR-Cas para regulación, direccionamiento y edición del genoma en Nat. Biotechnol., 2014, 32(4):347-355.

La “región codificante que incluye un codón de iniciación en su extremo 5’ que codifica para un polipéptido” también puede contener una secuencia de desoxirribonucleótidos que codifica para una meganucleasa. Las meganucleasas son endodesoxirribonucleasas que, a diferencia de endodesoxirribonucleasas “convencionales”, reconocen un sitio de reconocimiento grande (por ejemplo, una secuencia de ADN bicatenaria de 12 a 40 pares de bases). Como resultado, el sitio respectivo sólo se produce unas pocas veces, preferiblemente sólo una vez, en cualquier genoma dado. Por tanto, se considera que las meganucleasas son las enzimas de restricción que se producen de manera

natural más específicas y, por consiguiente, son herramientas adecuadas en tecnologías de edición del genoma.

La "región codificante que incluye un codón de iniciación en su extremo 5' que codifica para un polipéptido" también puede contener una secuencia de desoxirribonucleótidos que codifica para una nucleasa de dedos de cinc (ZFN). Las ZFN son enzimas de restricción artificiales generadas fusionando un dominio de unión a ADN de dedos de cinc a un dominio de escisión de ADN. Los dominios de dedos de cinc pueden modificarse por ingeniería para seleccionar como diana secuencias de ADN deseadas específicas y esto permite a las nucleasas de dedos de cinc seleccionar como diana secuencias únicas dentro de genomas complejos. Aprovechando la maquinaria de reparación del ADN endógena, pueden usarse ZFN para alterar de manera precisa el genoma de organismos superiores y, por tanto, son herramientas adecuadas en tecnologías de edición del genoma.

La "región codificante que incluye un codón de iniciación en su extremo 5' que codifica para un polipéptido" puede contener una secuencia de desoxirribonucleótidos que codifica para una nucleasa de efector similar a activador de transcripción (TALEN). Las TALEN son enzimas de restricción que pueden modificarse por ingeniería para cortar secuencias de ADN específicas. Las TALEN son proteínas de fusión en las que un dominio de unión a ADN de efector TAL se fusiona a un dominio de escisión de ADN de una nucleasa. Los efectores similares a activador de transcripción (TALE) pueden modificarse por ingeniería para unirse prácticamente a cualquier secuencia de ADN deseada. Por tanto, cuando se combina con una nucleasa, puede cortarse ADN en ubicaciones deseadas específicas.

La molécula de ADN de la presente invención comprende como segundo módulo (b) una secuencia que está ubicada directamente en el sentido de 5' de la secuencia codificante.

Más específicamente, la molécula de ADN tal como se da a conocer en el presente documento comprende un módulo (b) directamente en el sentido de 5' de dicha secuencia codificante, en la que dicho módulo (b) es una secuencia seleccionada del grupo que consiste en:

(b1) R<sub>1</sub>-CGCCACC (SEQ ID NO: 1);

o una secuencia en la que en dicha secuencia la C en la posición 6 de SEQ ID NO: 1 está sustituida por una A y la C en la posición 7 de SEQ ID NO: 1 está sustituida por una G; y/o la A en la posición 5 de SEQ ID NO: 1 está sustituida por una G; y

(b2) R<sub>1</sub>-CNGCCACC (SEQ ID NO: 2), en la que el nucleótido N en la posición 2 de SEQ ID NO: 2 es un nucleótido seleccionado del grupo que consiste en T, G, C o A;

o una secuencia en la que en dicha secuencia la C en la posición 7 de SEQ ID NO: 2 está sustituida por una A y la C en la posición 8 de SEQ ID NO: 2 está sustituida por una G; y/o la A en la posición 6 de SEQ ID NO: 2 está sustituida por una G,

en la que R<sub>1</sub> es un promotor que se reconoce por una ARN polimerasa dependiente de ADN.

Tal como se da a conocer en el presente documento, las secuencias según el punto (b) anteriormente en el presente documento no están particularmente limitadas a las secuencias específicas anteriores sino que también pueden referirse a (una) secuencia(s) que muestra(n) (una) adición/adiciones de nucleótido(s) en comparación con tales secuencias, en la(s) que el/los nucleótido(s) adicional(es) puede(n) añadirse en el extremo 5' de R<sub>1</sub> en la(s) secuencia(s) anteriormente descrita(s). El/los nucleótido(s) adicional(es) comprende(n) cadenas de polinucleótido de hasta 0 (sin cambios), 1, 2, 3, 4, 5, 6, 7, 8, 9 ó 10 nucleótidos, preferiblemente de hasta 20 nucleótidos. Más preferiblemente, se añaden 11, 12, 13, 14, 15, 16, 18 ó 19 nucleótidos en el extremo 5'. Incluso más preferiblemente se añaden hasta 30 nucleótidos en el extremo 5'.

Dado que la adición de nucleótidos en el sentido de 5' del promotor R<sub>1</sub> no cambiará las propiedades funcionales anteriores de la(s) UTR de la invención, la adición de los nucleótidos también puede tener una longitud de hasta 40, 50, 60, 70, 80, 90 o incluso 100 nucleótidos o incluso más, hasta 200, 300, 400 ó 500 nucleótidos.

Tal como se mencionó anteriormente, una molécula de ADN bicatenaria comprende dos cadenas antiparalelas en la que una cadena se denomina cadena "sentido" si su secuencia es la misma que la de una copia de ARN mensajero que se traduce para dar una proteína. La secuencia en la cadena opuesta, complementaria, se denomina secuencia "antisentido". Por tanto, la molécula de ADN de la presente invención no sólo se refiere a la molécula de ADN anterior que corresponde a un ARNm que comprende una cadena con los elementos anteriores (a) y (b) sino también a una molécula de ADN que comprende la cadena complementaria, es decir, la cadena antisentido que puede transcribirse para dar ARNm. Esta cadena complementaria de la molécula de ADN de la presente invención se define por referencia a la cadena antisentido que puede determinarse fácilmente dadas las reglas de apareamiento de bases.

La molécula de ADN de la presente invención también comprende en el módulo (b) un promotor R<sub>1</sub> que se reconoce por una ARN polimerasa dependiente de ADN. Dicho promotor R<sub>1</sub> está directamente unido a la secuencia restante definida en el punto (b1) o (b2), es decir, sin la aparición de ningún nucleótido intermedio.

Tal como se da a conocer en el presente documento, el promotor (y variantes del mismo) puede usarse siempre que una ARN polimerasa dependiente de ADN correspondiente pueda reconocer la secuencia respectiva. En la técnica se conocen numerosas ARN polimerasas (también conocidas como ARN polimerasas dependientes de ADN y con frecuencia abreviadas como RNAP o RNAPol). Estas enzimas pueden producir el ARN transcrito primario. Tal como se expuso anteriormente, las ARN polimerasas dependientes de ADN pueden sintetizar cadenas de ARN usando ADN como moldes en un proceso denominado transcripción. Una ARN polimerasa dependiente de ADN inicia la transcripción en secuencias de ADN específicas conocidas como promotores. Después, produce una cadena de ARN que es complementaria a la cadena de ADN de molde. El proceso de añadir nucleótidos a la cadena de ARN se conoce como elongación. Por tanto, en el contexto de la presente divulgación, preferiblemente el término "reconocer" no sólo significa que la ARN polimerasa dependiente de ADN puede detectar/unirse específicamente a su secuencia de promotor correspondiente  $R_1$ . Este término también se refiere a la capacidad de la ARN polimerasa dependiente de ADN para iniciar la transcripción y para producir después una molécula de ARN durante la elongación.

El experto puede determinar, mediante métodos conocidos en la técnica, si una ARN polimerasa dependiente de ADN dada puede reconocer un promotor respectivo. Además, usando métodos bien conocidos para la evaluación de interacciones de proteína/ADN, puede identificarse una secuencia de promotor  $R_1$  correspondiente (desconocida) de una ARN polimerasa dependiente de ADN dada y viceversa.

Por tanto, la capacidad de una ARN polimerasa dependiente de ADN para reconocer/unirse a su promotor  $R_1$  y, preferiblemente, la capacidad para iniciar la transcripción pueden determinarse mediante métodos conocidos en la técnica tal como se describe, por ejemplo, en Journal of Biological Chemistry, 1993, 268(26):19299-19304, mientras que el descubrimiento de numerosas ARN polimerasas dependientes de ADN se revisa en Journal of Biological Chemistry, 2005, 280(52):42477-42485).

El promotor  $R_1$  que se reconoce por una ARN polimerasa dependiente de ADN es un promotor de bacteriófago.

En la técnica se sabe que una ARN polimerasa dependiente de ADN de T7 reconoce la secuencia TAATACGACTCACTATAGGGAGA (SEQ ID NO: 3), la ARN polimerasa dependiente de ADN de T3 reconoce la secuencia AATTAACCCTCACTAAAGGGAGA (SEQ ID NO: 4), la ARN polimerasa dependiente de ADN de SP6 reconoce la secuencia ATTTAGGTGACACTATAGAAG (SEQ ID NO: 5) y la ARN polimerasa dependiente de ADN de K11 reconoce la secuencia AATTAGGGCACACTATAGGGA (SEQ ID NO: 6).

Tal como se da a conocer en el presente documento, en el contexto de la presente invención, de hecho, puede usarse cualquier promotor (y variantes del mismo) siempre que una ARN polimerasa dependiente de ADN correspondiente, preferiblemente ARN polimerasa dependiente de ADN de bacteriófago, pueda reconocer la secuencia respectiva.

En la presente invención,  $R_1$  se selecciona del grupo que consiste en:

(i) TAATACGACTCACTATAGGGAGA (SEQ ID NO: 3) o una secuencia que muestra de 1 a 6 sustituciones en comparación con SEQ ID NO: 3 y que se reconoce por una ARN polimerasa dependiente de ADN de T7;

(ii) AATTAACCCTCACTAAAGGGAGA (SEQ ID NO: 4) o una secuencia que muestra de 1 a 6 sustituciones en comparación con SEQ ID NO: 4 y que se reconoce por una ARN polimerasa dependiente de ADN de T3;

(iii) ATTTAGGTGACACTATAGAAG (SEQ ID NO: 5) o una secuencia que muestra de 1 a 6 sustituciones en comparación con SEQ ID NO: 5 y que se reconoce por una ARN polimerasa dependiente de ADN de SP6; y

(iv) AATTAGGGCACACTATAGGGA (SEQ ID NO: 6) o una secuencia que muestra de 1 a 6 sustituciones en comparación con SEQ ID NO: 6 y que se reconoce por una ARN polimerasa dependiente de ADN de K11.

Tal como también se da a conocer en el presente documento, la secuencia puede ser una secuencia que muestra de 1 a 3, 4 ó 5 sustituciones siempre que la secuencia correspondiente todavía pueda reconocerse por la ARN polimerasa dependiente de ADN de T7, T3, SP6 y K11, respectivamente. Tal como se da a conocer en el presente documento, la secuencia puede ser una secuencia que muestra de 1 a 2 sustituciones siempre que la secuencia correspondiente todavía pueda reconocerse por la ARN polimerasa dependiente de ADN de T7, T3, SP6 y K11, respectivamente. Lo más preferiblemente, la secuencia puede ser una secuencia que muestra 1 sustitución siempre que la secuencia correspondiente todavía pueda reconocerse por la ARN polimerasa dependiente de ADN de T7, T3, SP6 y K11, respectivamente.

Tal como se da a conocer en el presente documento, las secuencias de promotor  $R_1$  que se reconocen por una ARN polimerasa dependiente de ADN no están particularmente limitadas a ninguna de las secuencias de SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5 o SEQ ID NO: 6 o secuencias que muestran de 1 a 6 sustituciones en comparación con las mismas, sino que también pueden ser secuencias que muestran de 1 a 7, 8, 9, 10, 11 ó 12 sustituciones siempre que la secuencia correspondiente todavía pueda reconocerse por la ARN polimerasa dependiente de ADN de T7, T3, SP6 y K11, respectivamente.



- 5 Tal como se da a conocer en el presente documento, a partir de la(s) sustitución/sustituciones anterior(es) en las secuencias de TAATACGACTCACTATAGGGAGA (SEQ ID NO: 3), AATTAACCCTCACTAAAGGGAGA (SEQ ID NO: 4), ATTTAGGTGACACTATAGAAG (SEQ ID NO: 5) o AATTAGGGCACACTATAGGGA (SEQ ID NO: 6), se excluyen las sustituciones en los 5 nucleótidos "CACTA" en las posiciones 11 a 12 en las secuencias anteriores de SEQ ID NO: 3 a 6 dado que estos 5 nucleótidos están conservados entre las cuatro secuencias.
- 10 Tal como se da a conocer en el presente documento, a partir de la(s) sustitución/sustituciones anterior(es) en las secuencias de SEQ ID NO: 3 a 6, se excluye una sustitución en el nucleótido "T" en la posición 4 en las secuencias anteriores de SEQ ID NO: 3 a 6 dado que este nucleótido está conservado entre las cuatro secuencias.
- Tal como se da a conocer en el presente documento, a partir de la(s) sustitución/sustituciones anterior(es) en las secuencias de SEQ ID NO: 3 a 6, se excluye una sustitución en el nucleótido "A" en la posición 5 en las secuencias anteriores de SEQ ID NO: 3 a 6 dado que este nucleótido está conservado entre las cuatro secuencias.
- 15 Tal como se da a conocer en el presente documento, a partir de la(s) sustitución/sustituciones anterior(es) en las secuencias de SEQ ID NO: 3 a 6, se excluye una sustitución en el nucleótido "G" en la posición 18 en las secuencias anteriores de SEQ ID NO: 3 a 6 dado que este nucleótido está conservado entre las cuatro secuencias.
- La capacidad de una ARN polimerasa dependiente de ADN de T7, T3, SP6 y K11 para reconocer/unirse a su promotor R<sub>1</sub> puede determinarse mediante métodos conocidos en la técnica tal como se expuso anteriormente.
- 20 Tal como se da a conocer en el presente documento, la molécula de ADN puede ser una molécula de ADN que comprende un módulo (b1) directamente en el sentido de 5' de dicha secuencia codificante, en la que en dicho módulo (b1) el nucleótido N en la posición 2 de SEQ ID NO: 2 es un nucleótido seleccionado del grupo que consiste en T, G o C y en la que el nucleótido N no es una A.
- Tal como se da a conocer en el presente documento, dicho nucleótido N en la posición 2 de SEQ ID NO: 2 puede ser T.
- 25 Tal como se da a conocer en el presente documento, la molécula de ADN puede ser una molécula de ADN en la que el nucleótido que sigue directamente en el sentido de 3' del codón de iniciación no es el nucleótido G. La molécula de ADN de la presente invención puede ser una molécula de ADN en la que el nucleótido que sigue directamente en el sentido de 3' del codón de iniciación es un nucleótido seleccionado del grupo que consiste en A, T y C.
- 30 La molécula de ADN de la presente invención también puede ser una molécula de ADN que comprende un módulo (b1) tal como se definió anteriormente, en la que dicho módulo (b1) es una secuencia en la que la C en la posición 6 de SEQ ID NO: 1 está sustituida por una A y la C en la posición 7 de SEQ ID NO: 1 está sustituida por una G; y/o la A en la posición 5 de SEQ ID NO: 1 está sustituida por una G y en la que el nucleótido que sigue directamente en el sentido de 3' del codón de iniciación es un nucleótido seleccionado del grupo que consiste en A, T y C.
- 35 Tal como también se da a conocer en el presente documento, la molécula de ADN puede ser una molécula de ADN que comprende un módulo (b2) tal como se definió anteriormente, en la que dicho módulo (b2) es una secuencia en la que la C en la posición 7 de SEQ ID NO: 2 está sustituida por una A y la C en la posición 8 de SEQ ID NO: 2 está sustituida por una G; y/o la A en la posición 6 de SEQ ID NO: 2 está sustituida por una G y en la que el nucleótido que sigue directamente en el sentido de 3' del codón de iniciación es un nucleótido seleccionado del grupo que consiste en A, T y C.
- 40 En biología molecular y genética, en el sentido de 5' y en el sentido de 3' se refieren ambos a una posición relativa en una molécula de ADN. En el contexto de la presente invención, en el sentido de 5' es hacia el extremo 5' de la cadena sentido de la molécula de ADN y en el sentido de 3' es hacia el extremo 3' de la molécula.
- 45 Por consiguiente, en la presente invención, la secuencia definida en el punto (b), anterior, está ubicada directamente en el sentido de 5' de la región codificante según el punto (a), de manera más específica, directamente en el sentido de 5' del codón de iniciación de la región codificante. Por tanto, "directamente en el sentido de 5'" en este contexto significa que no hay ningún nucleótido adicional entre la secuencia según el punto (b) y la secuencia codificante que se inicia con un codón de iniciación. Por tanto, la región codificante que se inicia con un codón de iniciación está inmediatamente adyacente a la secuencia según el punto (b) anteriormente en el presente documento.
- Las moléculas de ADN de la presente invención pueden generarse/sintetizarse de manera recombinante (por ejemplo, en un sistema *in vivo* o uno *in vitro*) o de manera sintética (por ejemplo, mediante una reacción PCR o en una reacción química) mediante métodos conocidos por el experto en la técnica.
- 50 La molécula de ADN de la presente invención es preferiblemente una molécula de ácido nucleico recombinante, es decir, está compuesta por elementos que no se producen en la naturaleza en esta combinación. La molécula de ácido nucleico de la invención puede ser sintética o semisintética.
- La molécula de ADN puede estar presente en forma de secuencias de ADN fusionadas de módulos (a) y (b) (definidos en los puntos (a) y (b), de manera respectiva, anteriormente) es decir, una molécula de ADN (de fusión)

que se forma combinando al menos dos secuencias de nucleótidos que contienen dichos módulos. Normalmente, tal como se explicará en más detalle a continuación, esto puede lograrse clonando un ADNc en un vector de expresión lo cual permite la transcripción al interior de la molécula de ARN. Por consiguiente, la molécula de ADN de la presente invención puede ser una secuencia de ADN fusionada, es decir, una molécula química que se forma uniendo dos o más polinucleótidos mediante el grupo fosfato de un nucleótido unido al carbono en 3' en otro nucleótido, formando un enlace fosfodiéster entre los extremos respectivos de un módulo y el extremo de otra molécula. De esta manera, moléculas de ADN que contienen dichos al menos dos módulos se unen entre sí en forma de una molécula de ADN. Una vez clonada en marco, una molécula de ADN recombinante de este tipo puede transcribirse entonces para dar su secuencia de ácido nucleico de ARN correspondiente que codifica para dicha molécula de proteína, polipéptido o enzima.

Una molécula de ADN según la presente invención puede introducirse en un vector, preferiblemente un vector de expresión, mediante técnicas de biología molecular convencionales (véase, por ejemplo Sambrook *et al.*, Molecular Cloning, A laboratory manual, 2ª ed., 1989). El término "vector" tal como "vector de expresión" o "vector de clonación" en el sentido de la presente invención se entiende como una unidad circular, bicatenaria, de ADN que se replica dentro de una célula independientemente del ADN cromosómico y que se usa como vehículo para transportar material genético al interior de una célula, en la que puede replicarse y/o expresarse (es decir, transcribirse para dar ARN y traducirse para dar una secuencia de aminoácidos). Un vector que contiene ADN foráneo se denomina ADN recombinante. El propio vector es generalmente una secuencia de ADN que consiste normalmente en un inserto (por ejemplo, una molécula de ácido nucleico/molécula de ADN de la presente invención) y una secuencia más grande que sirve como "estructura principal" del vector. Los plásmidos en el sentido de la presente invención se encuentran lo más frecuentemente en bacterias y se usan en investigación de ADN recombinante para transferir genes entre células y, como tal, son una subpoblación de "vectores" tal como se usa en el sentido de la presente invención.

Resulta evidente para el experto en la técnica que pueden añadirse secuencias reguladoras adicionales a la molécula de ADN de la invención. Por ejemplo, pueden emplearse potenciadores de la transcripción y/o secuencias que permiten la expresión inducida. Un sistema inducible adecuado es, por ejemplo, expresión génica regulada por tetraciclina tal como se describe, por ejemplo, por Gossen y Bujard, Proc. Natl. Acad. Sci. USA 89 (1992), 5547-5551) y Gossen, Trends Biotech. 12 (1994), 58-62, o un sistema de expresión génica inducible por dexametasona tal como se describe, por ejemplo por Crook, EMBO J. 8 (1989), 513-519.

La presente invención también se refiere a un vector, preferiblemente un vector de expresión, que comprende la molécula de ADN de la presente invención.

El vector de la presente invención puede ser, por ejemplo, un plásmido, cósmido, virus, bacteriófago u otro vector usado por ejemplo convencionalmente en ingeniería genética, y puede comprender genes adicionales tales como genes de marcador que permiten la selección de dicho vector en una célula huésped adecuada y en condiciones adecuadas.

La molécula de ADN tal como se da a conocer en el presente documento también contiene preferiblemente señal de poli-A, garantizando la terminación de la transcripción y estabilización del transcrito mediante adición de una cola de poli-A.

Las moléculas de ADN y los vectores de la invención pueden diseñarse para su introducción directa o para su introducción mediante liposomas, vectores virales (por ejemplo, adenovirales, retrovirales), electroporación, sistemas balísticos (por ejemplo, pistola génica) u otros sistemas de suministro al interior de la célula. Adicionalmente, puede usarse un sistema baculoviral como sistema de expresión en eucariotas para las moléculas de ácido nucleico de la invención.

La presente invención también se refiere a una célula huésped que comprende el vector de la presente invención. Por tanto, la presente invención se refiere a un huésped transfectado o transformado con el vector de la invención o a un huésped no humano que porta el vector de la presente invención, es decir a una célula huésped o un huésped que se modifica genéticamente con una molécula de ADN según la invención o con un vector que comprende una molécula de ADN de este tipo. El término "genéticamente modificado" significa que la célula huésped o el huésped comprende, además de su genoma natural, una molécula de ADN o vector según la invención que se introdujo en la célula o el huésped o en uno de sus predecesores/progenitores. La molécula de ADN o el vector puede estar presente en la célula huésped o el huésped genéticamente modificado o bien como molécula independiente fuera del genoma, preferiblemente como molécula con capacidad para replicación, o bien puede integrarse de manera estable en el interior del genoma de la célula huésped o el huésped. La transformación de la célula huésped con un vector según la invención puede llevarse a cabo mediante métodos convencionales, tal como se describe, por ejemplo, en Sambrook y Russell (2001), Molecular Cloning: A Laboratory Manual, CSH Press, Cold Spring Harbor, NY, EE.UU.; Methods in Yeast Genetics, A Laboratory Course Manual, Cold Spring Harbor Laboratory Press, 1990. La célula huésped se cultiva en medios de nutrientes que cumplen los requisitos de la célula huésped particular usada, en particular con respecto al valor de pH, temperatura, concentración de sal, aireación, antibióticos, vitaminas, oligoelementos, etc.

La célula huésped puede ser cualquier célula procariota o eucariota. Las células procariotas adecuadas son aquellas

usadas generalmente para clonación tales como *E. coli* o *Bacillus subtilis*. Además, las células eucariotas comprenden, por ejemplo, células fúngicas o de animal. Los ejemplos de células fúngicas adecuadas son células de levadura, preferiblemente las del género *Saccharomyces* y lo más preferiblemente las de la especie *Saccharomyces cerevisiae*. Células de animales adecuadas son, por ejemplo, células de insectos, células de vertebrados, preferiblemente células de mamíferos, tales como por ejemplo HEK293, NSO, CHO, COS-7, MDCK, U2-OSHela, NIH3T3, MOLT-4, Jurkat, PC-12, PC-3, IMR, NT2N, Sk-n-sh, CaSki, C33A. Pueden obtenerse líneas celulares adecuadas adicionales conocidas en la técnica a partir de depósitos de líneas celulares tales como, por ejemplo, la Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH (DSMZ, colección alemana de microorganismos y cultivos celulares) o la American Type Culture Collection (ATCC, colección americana de cultivos tipo). Según la presente invención, se prevé además que las células/cultivos celulares primarios pueden funcionar como células huésped. Dichas células se derivan en particular a partir de insectos (tales como insectos de las especies *Drosophila* o *Blatta*) o mamíferos (tales como ser humano, cerdo, ratón o rata). Dichas células huésped también pueden comprender células de y/o derivadas de líneas celulares tales como líneas celulares de neuroblastoma. Las células primarias anteriormente mencionadas se conocen bien en la técnica y comprenden, entre otras, astrocitos primarios, cultivos espinales (mixtos) o cultivos de hipocampo.

La presente invención también se refiere a una composición que comprende la molécula de ADN de la presente invención, el vector de la presente invención o la célula huésped de la presente invención.

En un segundo aspecto, se da a conocer una molécula de ARN que comprende

(a) una región codificante, que incluye un codón de iniciación en su extremo 5', que codifica para un polipéptido; y  
 (b) directamente en el sentido de 5' de dicha secuencia codificante, una UTR seleccionada del grupo que consiste en:

(b1) una UTR de la secuencia

R<sub>2</sub>-CGCCACC (SEQ ID NO: 1),

o una secuencia en la que en dicha secuencia de UTR la C en la posición 6 de SEQ ID NO: 1 está sustituida por una A y la C en la posición 7 de SEQ ID NO: 1 está sustituida por una G; y/o la A en la posición 5 de SEQ ID NO: 1 está sustituida por una G; y

(b2) una UTR de la secuencia

R<sub>2</sub>-CNGCCACC (SEQ ID NO: 2), en la que el nucleótido N en la posición 2 de SEQ ID NO: 2 es un nucleótido seleccionado del grupo que consiste en U, G, C o A, o una secuencia en la que en dicha secuencia de UTR la C en la posición 7 de SEQ ID NO: 2 está sustituida por una A y la C en la posición 8 de SEQ ID NO: 2 está sustituida por una G; y/o la A en la posición 6 de SEQ ID NO: 2 está sustituida por una G,

en la que R<sub>2</sub> es una secuencia de ARN correspondiente a la parte de una región de promotor que empieza con el nucleótido en el que una ARN polimerasa dependiente de ADN inicia la síntesis de ARN.

Una molécula de ácido ribonucleico (ARN) tal como se usa según la presente invención se refiere a una molécula polimérica que se ensambla como una cadena de los nucleótidos denominados G, A, U y C. Cada nucleótido en ARN contiene un azúcar de ribosa, con carbonos numerados de 1' a 5'. Una base nitrogenada está unida a la posición 1', en general, adenina (A), citosina (C), guanina (G) o uracilo (U). En una molécula de ARN polimérica, un grupo fosfato está unido a la posición 3' de una ribosa y a la posición 5' de la siguiente. Por tanto, los nucleótidos en una molécula de ARN polimérica están unidos de manera covalente entre sí, en los que el grupo fosfato de un nucleótido se une al carbono en 3' en el nucleótido posterior, formando así un enlace fosfodiéster. Por consiguiente, una cadena de ARN tiene un extremo 5' y un extremo 3', denominados de ese modo por los carbonos en el anillo de ribosa. Por convenio, en el sentido de 5' y en el sentido de 3' se refieren a la dirección de 5' a 3' en la que tiene lugar la transcripción de ARN. Preferiblemente, la molécula de ARN es una molécula de ARN mensajero (ARNm). ARNm es una gran familia de moléculas de ARN que transportan información genética desde el ADN hasta el ribosoma, en el que especifican la secuencia de aminoácidos de los productos de proteína de expresión génica. Tras la transcripción de ARNm transcrito primario (conocido como pre-ARNm) por ARN polimerasa, se traduce ARNm procesado, maduro, para dar un polímero de aminoácidos: una proteína, tal como se resume en el dogma central de la biología molecular. Como en el ADN, la información genética de ARNm está en la secuencia de nucleótidos, que se dispone en codones que consisten en tres bases cada uno. Cada codón codifica para un aminoácido específico, excepto por los codones de terminación, que terminan la síntesis de proteínas.

La molécula de ARN tal como se da a conocer en el presente documento comprende dos módulos principales tal como se definió en los puntos (a) y (b) anteriormente. Además, la molécula de ARN de presente invención comprende una UTR en su extremo 3'. Por tanto, la molécula de ARN de la presente invención se asemeja, con respecto a su estructura, a una molécula de ARNm "normal" que se produce en la naturaleza, que alberga una región codificante así como regiones no traducidas (UTR) (en 5' y 3') y, opcionalmente, una cola de poli-A.

El término “región codificante que incluye un codón de iniciación en su extremo 5’” tal como se usa según la presente invención se refiere a una secuencia que está compuesta por codones, que se decodifican y traducen para dar proteína por el ribosoma según la información proporcionada por el código genético. Las regiones codificantes comienzan habitualmente con un codón de iniciación en su extremo 5’ y terminan con un codón de terminación. En general, el codón de iniciación es un triplete AUG y el codón de terminación es UAA, UAG o UGA. Además de ser codificantes para proteínas, las porciones de regiones codificantes pueden servir como secuencias reguladoras en el pre-ARNm como potenciadores de corte y empalme exónicos o silenciadores de corte y empalme exónicos. La región codificante de un gen que codifica para un polipéptido o una proteína tal como se usa según la presente invención también se conoce como secuencia codificante o CDS (de secuencia de ADN codificante) y es esa porción del ADN o ARN de un gen, compuesta por exones, que codifica para un polipéptido o proteína. La región codificante en ARNm está flanqueada por la región no traducida en 5’ (5’UTR) y la región no traducida en 3’ (3’UTR) que también son partes de los exones. Además, las moléculas de ARNm pueden comprender además una denominada caperuza en 5’ y una cola de poli-A. La caperuza en 5’, la 5’UTR, la 3’UTR y la cola de poli-A son regiones de una molécula de ARNm que no se traducen para dar una proteína.

El término “región no traducida” o “UTR” tal como se usa según la presente invención se refiere a secciones del ARNm en el sentido de 5’ del codón de iniciación y en el sentido de 3’ del codón de terminación que no se traducen, y, por tanto, se denominan región no traducida en cinco prima (5’UTR) y región no traducida en tres prima (3’UTR), respectivamente. Estas regiones se transcriben con la región codificante y por tanto son exónicas dado que están presentes en el ARNm maduro.

Tal como se usa en la presente invención, la región no traducida en 3’ (3’-UTR) se refiere a la sección de ARN mensajero (ARNm) que sigue inmediatamente al codón de terminación de la traducción. La 3’UTR puede comprender regiones reguladoras dentro de la región no traducida en 3’ que se sabe que influyen en la poliadenilación y estabilidad del ARNm. Muchas 3’-UTR también contienen elementos ricos en AU (ARE). Además, la 3’-UTR puede contener preferiblemente la secuencia AAUAAA que dirige la adición de varios cientos de residuos de adenina denominados cola de poli(A) al extremo del transcrito de ARNm.

Tal como se usa en la presente invención, la región no traducida en 5’ (5’UTR) (también conocida como secuencia líder o ARN líder) es la región de un ARNm que está directamente en el sentido de 5’ del codón de iniciación. La 5’UTR comienza en el sitio de iniciación de la transcripción y termina un nucleótido (nt) antes del codón de iniciación (habitualmente AUG) de la región codificante. En eucariotas, la longitud de la 5’UTR es generalmente de desde 100 hasta varios miles de nucleótidos de longitud, pero algunas veces también se producen UTR más cortas en eucariotas.

En la presente invención, la 5’UTR es extremadamente corta dado que un objetivo de la presente invención es proporcionar una secuencia de UTR mínima.

Una molécula de ARN de la presente invención también puede contener una cola de poli-A. Una cola de poli-A es una secuencia larga de nucleótidos de adenina (con frecuencia varios cientos) añadidos al extremo 3’ del pre-ARNm mediante un proceso denominado poliadenilación. Esta cola fomenta la exportación desde el núcleo y la traducción, y protege al ARNm frente a la degradación. La poliadenilación es la adición de una cola de poli(A) a un ARN mensajero. La cola de poli(A) consiste en múltiples monofosfatos de adenosina; dicho de otro modo, es un tramo de ARN que sólo tiene bases de adenina. En eucariotas, la poliadenilación es parte del proceso que produce ARN mensajero (ARNm) maduro para su traducción.

Un módulo de la molécula de ARN, es decir, “una región codificante que incluye un codón de iniciación en su extremo 5’ que codifica para un polipéptido” (módulo (a)) no está particularmente limitado y puede ser cualquier región codificante deseada que tiene que expresarse en una célula dada. Con respecto a las realizaciones preferidas del término “una región codificante que incluye un codón de iniciación en su extremo 5’ que codifica para un polipéptido” (módulo (a)) se aplica lo mismo, cambiando lo que sea necesario, a la molécula de ARN de la presente invención que tal como se ha expuesto anteriormente en el contexto de la molécula de ADN de la presente invención.

La molécula de ARN tal como se da a conocer en el presente documento comprende un módulo (b) directamente en el sentido de 5’ de dicha secuencia codificante, en la que dicho módulo (b) es una UTR seleccionada del grupo que consiste en:

(b1) una UTR de la secuencia

R<sub>2</sub>-CGCCACC (SEQ ID NO: 1),

o una secuencia en la que en dicha secuencia de UTR la C en la posición 6 de SEQ ID NO: 1 está sustituida por una A y la C en la posición 7 de SEQ ID NO: 1 está sustituida por una G; y/o la A en la posición 5 de SEQ ID NO: 1 está sustituida por una G; y

(b2) una UTR de la secuencia

R<sub>2</sub>-CNGCCACC (SEQ ID NO: 2), en la que el nucleótido N en la posición 2 de SEQ ID NO: 2 es un nucleótido seleccionado del grupo que consiste en U, G, C o A, o una secuencia en la que en dicha secuencia de UTR la C en la posición 7 de SEQ ID NO: 2 está sustituida por una A y la C en la posición 8 de SEQ ID NO: 2 está sustituida por una G; y/o la A en la posición 6 de SEQ ID NO: 2 está sustituida por una G,

- 5 en la que R<sub>2</sub> es una secuencia de ARN correspondiente a la parte de una región de promotor que empieza con el nucleótido en el que una ARN polimerasa dependiente de ADN inicia la síntesis de ARN.

Tal como se da a conocer en el presente documento, puede usarse cualquier secuencia de ARN que corresponde a la parte de una región de promotor que empieza con el nucleótido en el que una ARN polimerasa dependiente de ADN inicia la síntesis de ARN. El experto está fácilmente en posición de determinar las partes de una región de promotor que empieza con el nucleótido a partir del cual una ARN polimerasa dependiente de ADN inicia la síntesis de ARN. Esta secuencia de ARN R<sub>2</sub> es la secuencia de un promotor que corresponde a la parte de un promotor que se transcribe, es decir, que está realmente presente en el transcrito una vez transcrito.

El promotor R<sub>2</sub> es una secuencia de ARN correspondiente a la parte de una región de promotor que empieza con el nucleótido en el que una ARN polimerasa dependiente de ADN derivada de bacteriófago inicia la síntesis de ARN.

- 15 El promotor R<sub>2</sub> es una secuencia de ARN correspondiente a la parte de una región de promotor que empieza con el nucleótido en el que una ARN polimerasa dependiente de ADN de T7, ARN polimerasa dependiente de ADN de T3, ARN polimerasa dependiente de ADN de SP6 o una ARN polimerasa dependiente de ADN de K11 inicia la síntesis de ARN.

Con el fin de ilustrar esto, como ejemplos no limitativos, R<sub>2</sub> es la secuencia subrayada en las siguientes secuencias de promotor de TAATACGACTCACTATAGGGAGA (SEQ ID NO: 3; es decir, el promotor reconocido por la ARN polimerasa dependiente de ADN de T7), AATTAACCCCTCACTAAAGGGAGA (SEQ ID NO: 4; es decir, el promotor reconocido por la ARN polimerasa dependiente de ADN de T3), ATTTAGGTGACACTATAGAAG (SEQ ID NO: 5; es decir, el promotor reconocido por la ARN polimerasa dependiente de ADN de SP6) y AATTAGGGCACACTATAGGGA (SEQ ID NO: 6; es decir, el promotor reconocido por la ARN polimerasa dependiente de ADN de K11). Las secuencias subrayadas corresponden a la parte del promotor respectivo en el que una ARN polimerasa dependiente de ADN inicia la síntesis de ARN y, por consiguiente, que está realmente presente en la molécula de ARN (es decir, en el transcrito) una vez transcrito.

La(s) secuencia(s) de UTR que tiene(n) cualquiera de las sustituciones anteriores en comparación con una UTR de la secuencia R<sub>2</sub>-CGCCACC (SEQ ID NO: 1) o en comparación con una UTR de la secuencia R<sub>2</sub>-CNGCCACC (SEQ ID NO: 2) puede(n) dar como resultado una molécula de ARN que muestra la misma eficiencia de traducción o una similar, preferiblemente superior, que una molécula de ARN que comprende una UTR de la secuencia R<sub>2</sub>-CGCCACC (SEQ ID NO: 1) y una molécula de ARN que comprende una UTR de la secuencia R<sub>2</sub>-CNGCCACC (SEQ ID NO: 2), respectivamente. La eficiencia de traducción de una molécula de ARN dada que comprende una UTR tal como se describe en el presente documento puede determinarse por el experto mediante métodos conocidos en la técnica y tal como se describe a continuación.

La eficiencia de traducción es la tasa de traducción de ARNm para dar polipéptidos o proteínas dentro de células. La eficiencia de traducción de un ARNm dado se mide como el número de proteínas o polipéptidos que se traducen por ARNm por unidad de tiempo. La traducción es el proceso en el que ribosomas celulares crean proteínas y lo conoce bien el experto. En resumen, en la traducción, ARN mensajero (ARNm) que se produce mediante transcripción a partir de ADN se decodifica por un ribosoma para producir una cadena de aminoácidos específica o un polipéptido o una proteína.

Por tanto, la eficiencia de traducción de una molécula de ARN dada que alberga una secuencia de UTR modificada con cualquiera de las sustituciones anteriores es preferiblemente la misma o superior en comparación con la eficiencia de traducción del mismo ARN dado pero que alberga una UTR de R<sub>2</sub>-CGCCACC (SEQ ID NO: 1) o R<sub>2</sub>-CNGCCACC (SEQ ID NO: 2) tal como se definió anteriormente en el presente documento, respectivamente. Por consiguiente, el número de proteínas o polipéptidos codificados por la región codificante de la molécula de ARN que alberga una secuencia de UTR modificada con cualquiera de las sustituciones anteriores que se traducen por ARN por unidad de tiempo es al menos el mismo que el, o es, preferiblemente, superior al, número de proteínas o polipéptidos codificados por la región codificante de la molécula de ARN que alberga una UTR de R<sub>2</sub>-CGCCACC (SEQ ID NO: 1) o R<sub>2</sub>-CNGCCACC (SEQ ID NO: 2) tal como se definió anteriormente en el presente documento, respectivamente, que se traducen por ARN por unidad de tiempo.

La eficiencia de traducción, en el contexto de la presente invención, es preferiblemente la tasa de ARNm traducido para dar proteína dentro de una célula en un determinado punto de tiempo con respecto a la cantidad de ARNm que codifica para la proteína respectiva en dicha célula en el mismo punto de tiempo. Por tanto, la eficiencia de traducción es el cociente del ARNm traducido para dar proteína dentro de una célula en un determinado punto de tiempo y la cantidad de ARNm que codifica para la proteína respectiva. Ambos parámetros, es decir, el ARNm traducido para dar una proteína así como la cantidad de ARNm que codifica para la proteína respectiva, pueden determinarse mediante métodos conocidos en la técnica. Como ejemplos no limitativos, la cantidad de ARNm

5 traducido para dar proteína dentro de una célula puede determinarse, por ejemplo, tal como se determina mediante  
 10 citometría de flujo (FC) mientras que la cantidad de ARNm que codifica para la proteína respectiva puede medirse,  
 por ejemplo, mediante qPCR. La(s) UTR tal como se da(n) a conocer en el presente documento según el punto (b)  
 15 anteriormente en el presente documento no está(n) particularmente limitada(s) a las secuencias específicas  
 anteriores sino que también puede(n) referirse a (una) secuencia(s) de UTR que comprende(n) una secuencia que  
 muestra (una) adición/adiciones de nucleótido(s) en comparación con tales secuencias, en la(s) que el/los  
 nucleótido(s) adicional(es) puede(n) añadirse en el extremo 5' de la(s) UTR anteriormente descrita(s). El/los  
 20 nucleótido(s) adicional(es) comprende(n) cadenas de polinucleótido de hasta 0 (sin cambios), 1, 2, 3, 4, 5, 6, 7, 8, 9  
 ó 10 nucleótidos, preferiblemente de hasta 20 nucleótidos. Más preferiblemente, se añaden 11, 12, 13, 14, 15, 16, 18  
 ó 19 nucleótidos en el extremo 5'. Incluso más preferiblemente se añaden hasta 30 nucleótidos en el extremo 5'.

15 A la vista del fundamento de que es probable que la adición de nucleótidos no cambie las propiedades funcionales  
 anteriores de la(s) UTR respectiva(s), la adición de los nucleótidos también puede tener una longitud de hasta 40,  
 50, 60, 70, 80, 90 o incluso 100 nucleótidos o incluso más, hasta 200, 300, 400 ó 500 nucleótidos siempre que estas  
 secuencias tengan una capacidad similar (en cuanto a la eficiencia de traducción anteriormente descrita) a las UTR  
 20 definidas en el punto (b) anteriormente en el presente documento.

Tal como se da a conocer en el presente documento, la UTR según el punto (b1) anteriormente en el presente  
 documento tiene una longitud máxima de 11, 12 ó 13 nucleótidos. Preferiblemente, la UTR según el punto (b1)  
 25 anteriormente en el presente documento tiene una longitud máxima de 13 nucleótidos si R<sub>2</sub> es GGGAGA (SEQ ID  
 NO: 7) o GGGAGA (SEQ ID NO: 8).

20 Preferiblemente, la UTR según el punto (b1) anteriormente en el presente documento tiene una longitud máxima de  
 11 nucleótidos si R<sub>2</sub> es GAAG (SEQ ID NO: 9) o GGGA (SEQ ID NO: 10).

La UTR según el punto (b2) anteriormente en el presente documento tiene una longitud máxima de 12, 13 ó 14  
 nucleótidos. La UTR según el punto (b2) anteriormente en el presente documento tiene una longitud máxima de 14  
 nucleótidos si R<sub>2</sub> es GGGAGA (SEQ ID NO: 7) o GGGAGA (SEQ ID NO: 8).

25 La UTR según el punto (b2) anteriormente en el presente documento tiene una longitud máxima de 12 nucleótidos si  
 R<sub>2</sub> es GAAG (SEQ ID NO: 9) o GGGA (SEQ ID NO: 10).

Las moléculas de ARN de la presente invención que contienen la(s) UTR anteriormente descrita(s) pueden  
 generarse/sintetizarse de manera recombinante (por ejemplo, en un sistema *in vivo* o uno *in vitro*) o de manera  
 30 sintética mediante métodos conocidos por el experto en la técnica.

30 La transcripción *in vitro* de ARN requiere habitualmente un molde de ADN lineal que contiene una región de  
 promotor bicatenaria en la que se une la ARN polimerasa dependiente de ADN e inicia la síntesis de ARN mientras  
 que la región codificante puede ser bicatenaria o monocatenaria. En el caso en el que el molde de ADN lineal  
 contiene una región codificante monocatenaria, la cadena antisentido (es decir, la cadena que se lee por la  
 35 polimerasa dependiente de ADN) de la región codificante es parte del molde. Las ARN polimerasas dependientes de  
 ADN comunes son la polimerasa de T7, la polimerasa de T3, la polimerasa de SP6 y la polimerasa de K11. La  
 secuencia completa de sus promotores respectivos se muestra en SEQ ID NO: 3 a 6.

Los moldes de transcripción para una transcripción *in vitro* incluyen, por ejemplo, moldes de ADNc sintetizados a  
 partir de un precursor de ARN, moldes generados mediante PCR, oligonucleótidos químicamente sintetizados y  
 40 constructos de plásmidos. Muchos vectores de clonación de plásmidos ampliamente usados albergan promotores de  
 polimerasa de fagos ubicados a cada lado del sitio de clonación múltiple que permiten la transcripción de cualquier  
 cadena de una secuencia de nucleótidos insertada en el sitio de clonación múltiple. Los vectores de clonación  
 habitualmente usados incluyen, por ejemplo, los vectores pCRII de Invitrogen, pGEM de Promega y pBluescript de  
 Stratagene. La familia de vectores pTRIPLEscript de Ambion contienen los tres promotores de polimerasa de fagos  
 45 en tándem (en el mismo lado del sitio de clonación múltiple), permitiendo usar cualquiera de las tres polimerasas,  
 SP6, T7 o T3.

Las moléculas de ARN de la presente invención pueden producirse de manera recombinante en sistemas *in vivo*  
 mediante métodos conocidos por el experto en la técnica.

Alternativamente, las moléculas de ARN de la presente invención pueden producirse en un sistema *in vitro* usando,  
 por ejemplo, un sistema de transcripción *in vitro*. Se conocen habitualmente sistemas de transcripción *in vitro* y  
 50 habitualmente requieren un molde de ADN lineal purificado que contiene una secuencia de ADN "que codifica para"  
 la molécula de ARN en la que dicha secuencia de ADN está bajo el control de un promotor apropiado. Además, un  
 sistema de transcripción *in vitro* también requiere habitualmente trifosfatos de ribonucleósidos, un sistema de  
 tampón que incluye DTT e iones de magnesio, y una ARN polimerasa apropiada que proporciona la actividad  
 55 enzimática para la transcripción *in vitro* de la secuencia de ADN para dar una molécula de ARN correspondiente de  
 la presente invención.

Además, las moléculas de ARN pueden sintetizarse químicamente, por ejemplo, mediante síntesis química  
 convencional en un sintetizador de secuencias de nucleótidos automatizado usando un soporte de fase sólida y

técnicas convencionales o mediante síntesis química de las secuencias de ADN respectivas y posterior transcripción *in vitro* o *in vivo* de las mismas.

Según lo anterior, la presente invención proporciona moléculas de ARN/moléculas de poli(ácido ribonucleico), preferiblemente moléculas de poli(ácido ribonucleico) modificadas, en las que un módulo de dicha molécula de ARN, es decir, "una región codificante que incluye un codón de iniciación en su extremo 5'" (módulo (a)), codifica para un polipéptido. Los términos ácido nucleico y polinucleótido se usan de manera intercambiable e incluyen cualquier compuesto y/o sustancia que comprende un polímero de nucleótidos. El término nucleótido incluye desoxinucleótidos y ribonucleótidos. Los términos ácido ribonucleico y polirribonucleótido se usan de manera intercambiable y, en determinadas realizaciones, incluyen cualquier compuesto y/o sustancia que comprende un polímero de nucleótidos en el que más del 50 % de los nucleótidos son ribonucleótidos. En determinadas realizaciones, los polirribonucleótidos comprenden un polímero de nucleótidos en el que más del 60 %, el 70 %, el 75 %, el 80 %, el 90 %, más del 95 %, más del 99 % o el 100 % de los nucleótidos son ribonucleótidos. Los polirribonucleótidos en los que uno o más nucleótidos son nucleótidos modificados pueden denominarse polirribonucleótidos modificados. Sin embargo, el término polirribonucleótidos puede incluir polirribonucleótidos modificados.

La secuencia de las moléculas de ARN/polirribonucleótidos puede derivarse, por ejemplo, a partir de cualquier ácido nucleico adecuado que comprende la información genética de un gen de interés. Los ejemplos de ácidos nucleicos incluyen ADN genómico, ARN o ADNc de cualquier célula bacteriana o arquea que comprende el/los gen(es) de interés. Los polinucleótidos pueden derivarse a partir de ácidos nucleicos que portan genes mutados y polimorfismos. Una molécula de ARN/polirribonucleótido de la presente invención comprende una secuencia que no está particularmente limitada y puede comprender, como módulo A, cualquier región codificante deseada que se expresa en una célula dada. En una realización preferida, dicha secuencia puede ser una región codificante que codifica para un polipéptido/proteína deseado tal como se expuso anteriormente. Preferiblemente, en línea con lo anterior, la molécula de ARN/polirribonucleótido comprende además una secuencia no traducida posicionada en el sentido de 5' (5') del codón de iniciación del módulo A, una secuencia no traducida posicionada en el sentido de 3' (3') del codón de terminación del módulo A, o tanto una secuencia no traducida posicionada en el sentido de 5' (5') del codón de iniciación del módulo A como una secuencia no traducida posicionada en el sentido de 3' (3') del codón de terminación del módulo A. En una realización preferida, una molécula de ARN/polirribonucleótido de la presente invención puede ser una molécula de ARN/polirribonucleótido modificado.

Además de los cuatro ribonucleótidos clásicos, concretamente, adenosina, guanosina, citidina y uridina, existen numerosos análogos de cada una de estas nucleobases. Algunas veces a lo largo de, y en la, bibliografía, estos análogos, o moléculas de ARN/polirribonucleótidos que incluyen uno o más de estos análogos, se denominan como modificados (por ejemplo, nucleótidos modificados o ribonucleótidos modificados). Algunos análogos difieren de las nucleobases canónicas anteriores, pero todavía pueden existir en la naturaleza. Otros análogos no se producen de manera natural. Se contempla cualquier tipo de análogo.

Tal como se da a conocer en el presente documento, las moléculas de ARN/polirribonucleótidos pueden comprender análogos de nucleótidos (por ejemplo, el polirribonucleótido comprende un polirribonucleótido modificado). A continuación se proporcionan análogos de nucleótidos a modo de ejemplo (por ejemplo, análogos de U; análogos de C; análogos de A; análogos de G). Además, una molécula de ARN/polirribonucleótido u otro ácido nucleico de la divulgación también puede comprender (de manera adicional o alternativa) modificaciones en la estructura principal de fosfodiéster o en la unión entre nucleobases. Los ácidos nucleicos a modo de ejemplo que pueden formar parte o la totalidad de una molécula de ARN/polirribonucleótido de la divulgación incluyen, pero no se limitan a, ácidos ribonucleicos (ARN), ácidos desoxirribonucleicos (ADN), ácidos nucleicos de treosa (ANT), ácidos nucleicos de glicol (ANG), ácidos nucleicos peptídicos (ANP), ácidos nucleicos bloqueados (ANB, incluyendo ANB que tiene una configuración beta-D-ribo, alfa-ANB que tiene una configuración alfa-L-ribo (un diastereómero de ANB), 2'-amino-ANB que tiene una funcionalización 2'-amino, y 2'-amino-alfa-ANB que tiene una funcionalización 2'-amino) o híbridos de los mismos.

Tal como se da a conocer en el presente documento, una modificación puede estar en uno o más nucleósidos o la estructura principal de la molécula de ácido nucleico/polinucleótido. Tal como se da a conocer en el presente documento, una modificación puede estar tanto en un nucleósido como en una unión de estructura principal. Tal como se da a conocer en el presente documento, puede introducirse por ingeniería una modificación en un polinucleótido *in vitro*.

Tal como se da a conocer en el presente documento, un ribonucleótido/nucleótido modificado también puede sintetizarse tras la transcripción mediante modificación covalente de los ribonucleótidos/nucleótidos clásicos/naturales.

Una molécula de ARN/polirribonucleótido tal como se da a conocer en el presente documento puede ser una molécula de ARN/polirribonucleótido modificado y puede comprender análogos de purinas y/o análogos de pirimidinas. Tal como se da a conocer en el presente documento, una molécula de ARN/polirribonucleótido modificado de la presente invención comprende un análogo de pirimidina, tal como un análogo de uridina y/o un análogo de citidina. Tal como se da a conocer en el presente documento, una molécula de ARN/polirribonucleótido

modificado de la presente invención comprende un análogo de uridina y un análogo de citidina. Tal como se da a conocer en el presente documento, la molécula de ARN/polirribonucleótido modificado no comprende análogos de adenosina y/o análogos de guanosina. Tal como se da a conocer en el presente documento, la molécula de ARN/polirribonucleótido comprende un único tipo de análogo de uridina y un único tipo de análogo de citidina (por ejemplo, un tipo de análogo, no una única molécula de análogo, el único análogo puede estar presente a cualquiera de varios porcentajes descritos en el presente documento). Tal como se da a conocer en el presente documento, la molécula de ARN/polirribonucleótido comprende más de un tipo de análogo de uridina y/o citidina y, opcionalmente y si están presentes, uno o más análogos de adenosina y/o guanosina (o ninguno de los dos o ambos).

En algunos casos, una uridina modificada (por ejemplo, análogo de uridina) se selecciona de 2-tiouridina, 5'-metiluridina, pseudouridina, 5-yodouridina (I5U), 4-tiouridina (S4U), 5-bromouridina (Br5U), 2'-metil-2'-desoxiuridina (U2'm), 2'-amino-2'-desoxiuridina (U2'NH<sub>2</sub>), 2'-azido-2'-desoxiuridina (U2'N<sub>3</sub>) y 2'-fluoro-2'-desoxiuridina (U2'F). En algunos casos, una citidina modificada (por ejemplo, análogo de citidina) se selecciona de 5-metilcitidina, 3-metilcitidina, 2-tio-citidina, 2'-metil-2'-desoxicitidina (C2'm), 2'-amino-2'-desoxicitidina (C2'NH<sub>2</sub>), 2'-fluoro-2'-desoxicitidina (C2'F), 5-yodocitidina (I5C), 5-bromocitidina (Br5C) y 2'-azido-2'-desoxicitidina (C2'N<sub>3</sub>). Obsérvese que cuando se hace referencia a análogos, lo anterior también se refiere a análogos en su forma de 5'-trifosfato. Tal como se da a conocer en el presente documento, el análogo de citidina puede ser 5-yodocitidina y el análogo de uridina es 5-yodouridina.

Tal como se da a conocer en el presente documento, la molécula de ARN/polirribonucleótido puede ser una molécula de ARN/polirribonucleótido modificado. En algunos casos, la molécula de ARN/polirribonucleótido modificado es al menos el 25 % más estable en comparación con una molécula de ARN/polirribonucleótido no modificado (o sin modificar). En algunos casos, la molécula de ARN/polirribonucleótido modificado puede ser al menos el 30 % más estable, al menos el 35 % más estable, al menos el 40 % más estable, al menos el 45 % más estable, al menos el 50 % más estable, al menos el 55 % más estable, al menos el 60 % más estable, al menos el 65 % más estable, al menos el 70 % más estable, al menos el 75 % más estable, al menos el 80 % más estable, al menos el 85 % más estable, al menos el 90 % más estable o al menos el 95 % más estable en comparación con una molécula de ARN/polirribonucleótido no modificado.

Tal como se da a conocer en el presente documento, la estabilidad puede medirse *in vivo*. Tal como se da a conocer en el presente documento, la estabilidad puede medirse *in vitro*. Tal como se da a conocer en el presente documento, la estabilidad puede cuantificarse midiendo la semivida del polirribonucleótido.

Una molécula de ARN/polirribonucleótido tal como se da a conocer en el presente documento puede tener nucleótidos que se han modificado de la misma forma o bien una mezcla de diferentes nucleótidos modificados. Los nucleótidos modificados pueden tener modificaciones que se producen de manera natural o que no se producen de manera natural en ARN mensajero. Puede usarse una mezcla de diversos nucleótidos modificados. Por ejemplo, uno o más nucleótidos modificados dentro de una molécula de ARN/polirribonucleótido pueden tener modificaciones naturales, mientras que otra parte tiene modificaciones que no se encuentran de manera natural en ARNm. Adicionalmente, algunos nucleótidos modificados pueden tener una modificación de base, mientras que otros nucleótidos modificados tienen una modificación de azúcar. De la misma manera, es posible que todas las modificaciones sean modificaciones de base o todas las modificaciones sean modificaciones de azúcar o cualquier mezcla adecuada de las mismas. En algunos casos, la estabilidad de la molécula de ARN/polirribonucleótido modificado puede optimizarse de manera selectiva cambiando la naturaleza de bases modificadas dentro del polirribonucleótido modificado.

Tabla 2: Ejemplos no limitativos de análogos de U

Nombre	Modificación de base (posición 5')	Modificación de azúcar (posición 2')	De manera natural en ARNm
5-metiluridina (m5U)	CH <sub>3</sub>	-	No
5-yodouridina (I5U)	I	-	No
5-bromouridina (Br5U)	Br	-	No
2-tiouridina (S2U)	S (en la posición 2)	-	No
4-tiouridina (S4U)	S (en la posición 4)	-	No
2'-metil-2'-desoxiuridina (U2'm)	-	CH <sub>3</sub>	Sí
2'-amino-2'-desoxiuridina (U2'NH <sub>2</sub> )	-	NH <sub>2</sub>	No
2'-azido-2'-desoxiuridina (U2'N <sub>3</sub> )	-	N <sub>3</sub>	No
2'-fluoro-2'-desoxiuridina (U2'F)	-	F	No

Tabla 3: Ejemplos no limitativos de análogos de C

Nombre	Modificación de base (posición 5')	Modificación de azúcar (posición 2')	De manera natural en ARNm
5-metilcitidina (m5C)	CH <sub>3</sub>	-	Sí
5-yodocitidina (I5C)	I	-	No
5-bromocitidina (Br5C)	Br	-	No



2-tiocitidina (S2C)	S (en la posición 2)	-	No
2'-metil-2'-desoxicitidina (C2'm)	-	CH <sub>3</sub>	Sí
2'-amino-2'-desoxicitidina (C2'NH <sub>2</sub> )	-	NH <sub>2</sub>	No
2'-azido-2'-desoxicitidina (C2'N <sub>3</sub> )	-	N <sub>3</sub>	No
2'-fluoro-2'-desoxicitidina (C2'F)	-	F	No

Tabla 4: Ejemplos no limitativos de análogos de A

Nombre	Modificación de base (posición 5')	Modificación de azúcar (posición 2')	De manera natural en ARNm
N6-metiladenosina (m6A)	CH <sub>3</sub> (en la posición 6)	-	Sí
N1-metiladenosina (m1A)	CH <sub>3</sub> (en la posición 1)	-	No
2'-O-metiladenosina (A2'm)	-	CH <sub>3</sub>	Sí
2'-amino-2'-desoxiadenosina (A2'NH <sub>2</sub> )	-	NH <sub>2</sub>	No
2'-azido-2'-desoxiadenosina (A2'N <sub>3</sub> )	-	N <sub>3</sub>	No
2'-fluoro-2'-desoxiadenosina (A2'F)	-	F	No

Tabla 5: Ejemplos no limitativos de análogos de G

Nombre	Modificación de base (posición 5')	Modificación de azúcar (posición 2')	De manera natural en ARNm
N1-metilguanosina (m1G)	CH <sub>3</sub> (en posición 1)	-	No
2'-O-metilguanosina (G2'm)	-	CH <sub>3</sub>	Sí
2'-amino-3'-desoxiguanosina (G2'NH <sub>2</sub> )	-	NH <sub>2</sub>	No
2'-azido-2'-desoxiguanosina (G2'N <sub>3</sub> )	-	N <sub>3</sub>	No
2'-fluoro-2'-desoxiguanosina (G2'F)	-	F	No

5 Tal como también se da a conocer en el presente documento, un análogo (por ejemplo, un nucleótido modificado) puede seleccionarse del grupo que comprende ribonucleósido de piridin-4-ona, 5-yodouridina, 5-yodocitidina, 5-aza-uridina, 2'-amino-2'-desoxicitidina, 2'-fluor-2'-desoxicitidina, 2-tio-5-aza-uridina, 2-tiouridina, 4-tio-pseudouridina, 2-tio-pseudouridina, 5-hidroxiuridina, 3-metiluridina, 5-carboximetil-uridina, 1-carboximetil-pseudouridina, 5-propinil-uridina, 1-propinil-pseudouridina, 5-taurinometiluridina, 1-taurinometil-pseudouridina, 5-taurinometil-2-tio-uridina, 1-taurinometil-4-tio-uridina, 5-metil-uridina, 1-metil-pseudouridina, 4-tio-l-metil-pseudouridina, 2-tio-l-metil-pseudouridina, 1-metil-l-desaza-pseudouridina, 2-tio-1-metil-l-desaza-pseudouridina, dihidrouridina, dihidropseudouridina, 2-tio-dihidrouridina, 2-tio-dihidropseudouridina, 2-metoxiuridina, 2-metoxi-4-tio-uridina, 4-metoxi-pseudouridina, 4-metoxi-2-tio-pseudouridina, 5-aza-citidina, pseudoisocitidina, 3-metil-citidina, N4-acetilcitidina, 5-formilcitidina, 5-metilcitidina, N4-metilcitidina, 5-hidroxiacetilcitidina, 1-metil-pseudoisocitidina, pirrolo-citidina, pirrolo-pseudoisocitidina, 2-tio-citidina, 2-tio-5-metil-citidina, 4-tio-pseudoisocitidina, 4-tio-l-metil-pseudoisocitidina, 4-tio-l-metil-1-desaza-pseudoisocitidina, 1-metil-l-desaza-pseudoisocitidina, zebularina, 5-aza-zebularina, 5-metil-zebularina, 5-aza-2-tio-zebularina, 2-tio-zebularina, 2-metoxi-citidina, 2-metoxi-5-metil-citidina, 4-metoxi-pseudoisocitidina, 4-metoxi-l-metil-pseudoisocitidina, 2-aminopurina, 2, 6-diaminopurina, 7-desaza-adenina, 7-desaza-8-aza-adenina, 7-desaza-2-aminopurina, 7-desaza-8-aza-2-aminopurina, 7-desaza-2,6-diaminopurina, 7-desaza-8-aza-2,6-diaminopurina, 1-metiladenosina, N6-metiladenosina, N6-isopenteniladenosina, N6-(cis-hidroxiisopentenil)adenosina, 2-metiltio-N6-(cis-hidroxiisopentenil)adenosina, N6-glicinilcarbamoiladenosina, N6-treonilcarbamoiladenosina, 2-metiltio-N6-treonilcarbamoiladenosina, N6,N6-dimetiladenosina, 7-metiladenina, 2-metiltio-adenina, 2-metoxi-adenina, inosina, 1-metil-inosina, wyosina, wybutosina, 7-desaza-guanosina, 7-desaza-8-aza-guanosina, 6-tio-guanosina, 6-tio-7-desaza-guanosina, 6-tio-7-desaza-8-aza-guanosina, 7-metil-guanosina, 6-tio-7-metil-guanosina, 7-metilfosina, 6-metoxi-guanosina, 1-metilguanosina, N2-metilguanosina, N2,N2-dimetilguanosina, 8-oxo-guanosina, 7-metil-8-oxo-guanosina, 1-metil-6-tio-guanosina, N2-metil-6-tio-guanosina y N2,N2-dimetil-6-tio-guanosina.

10 Tal como se da a conocer en el presente documento, una molécula de ARN/polirribonucleótido modificado tal como se da a conocer en el presente documento no incluye pseudouridina. Tal como se da a conocer en el presente documento, una molécula de ARN/polirribonucleótido modificado no incluye 5-metilcitidina. Tal como se da a conocer en el presente documento, una molécula de ARN/polirribonucleótido modificado no incluye 5-metiluridina.

15 Tal como se da a conocer en el presente documento, una molécula de ARN/polirribonucleótido modificado comprende análogos de U y análogos de C, en la que tales análogos de U pueden ser todos el mismo análogo o pueden ser análogos diferentes (por ejemplo, más de un tipo de análogo), y en la que tales análogos de C pueden ser todos el mismo análogo o pueden ser análogos diferentes (por ejemplo, más de un tipo de análogo).

20 Tal como se da a conocer en el presente documento, una molécula de ARN/polirribonucleótido modificado no incluye análogos de adenosina y análogos de guanosina.

35 Tal como se describe en detalle en el presente documento, cuando una molécula de ARN/polirribonucleótido comprende un polirribonucleótido modificado, pueden estar presentes análogos como una determinada proporción de los nucleótidos en el compuesto (por ejemplo, un porcentaje dado de una nucleobase dada puede ser análogo,

tal como se describe en el presente documento).

Una molécula de ARN/polirribonucleótido que comprende al menos un nucleótido modificado es una molécula de ARN/polirribonucleótido modificado. Tal como se da a conocer en el presente documento, al menos aproximadamente el 5 % de la molécula de ARN/polirribonucleótido modificado puede incluir adenosina, citidina, guanosina o uridina modificadas o que no se producen de manera natural (por ejemplo, análogos o modificadas), tales como los nucleótidos análogos descritos en el presente documento. En algunos casos, al menos aproximadamente el 5 %, el 10 %, el 15 %, el 20 %, el 25 %, el 30 %, el 40 %, el 45 %, el 50 % de la molécula de ARN/polirribonucleótido modificado puede incluir adenosina, citidina, guanosina o uridina modificadas o que no se producen de manera natural (por ejemplo, análogos o modificadas). En algunos casos, como máximo aproximadamente el 50 %, el 45 %, el 40 %, el 35 %, el 30 %, el 25 %, el 20 %, el 15 %, el 10 %, el 5 % de la molécula de ARN/polirribonucleótido modificado incluye adenosina, citidina, guanosina o uridina modificadas o que no se producen de manera natural.

La molécula de ARN tal como se da a conocer en el presente documento puede contener una combinación de nucleótidos modificados y no modificados. Preferiblemente, la molécula de ARN contiene una combinación de nucleótidos modificados y no modificados tal como se describe en el documento WO 2011/012316. Tales moléculas de ARN también se conocen y se comercializan como "ARN SNIM®". Se notifica que la molécula de ARN descrita en el documento WO 2011/012316 muestra una estabilidad aumentada e inmunogenicidad disminuida. En una realización preferida, en una molécula de ARN modificada de este tipo, del 5 al 50 % de los nucleótidos de citidina y del 5 al 50 % de los nucleótidos de uridina están modificados. Los nucleótidos que contienen adenosina y guanosina pueden no estar modificados. Los nucleótidos de adenosina y guanosina pueden no estar modificados o estar parcialmente modificados, y preferiblemente están presentes en forma no modificada. Preferiblemente del 10 al 35 % de los nucleótidos de citidina y uridina están modificados y de manera particularmente preferible el contenido de los nucleótidos de citidina modificados se encuentra en un intervalo de desde el 7,5 hasta el 25 % y el contenido de los nucleótidos de uridina modificados en un intervalo de desde el 7,5 hasta el 25 %. Se ha encontrado que, de hecho, un contenido relativamente bajo, por ejemplo tan sólo el 10 % de cada uno, de nucleótidos de citidina y uridina modificados puede lograr las propiedades deseadas. Se prefiere particularmente que los nucleótidos de citidina modificados sean residuos de 5-metilcitidina y los nucleótidos de uridina modificados sean residuos de 2-tiouridina. Lo más preferiblemente, el contenido de nucleótidos de citidina modificados y el contenido de los nucleótidos de uridina modificados es del 25 %, respectivamente.

Tal como se da a conocer en el presente documento, en una molécula de ARN/molécula de polirribonucleótido modificada de este tipo, del 5 al 50 % de las citidinas pueden ser análogos de C y del 5 al 50 % de las uridinas pueden ser análogos de U. Tal como se da a conocer en el presente documento, en una molécula de polirribonucleótido modificada de este tipo del 5 al 40 % de las citidinas son análogos de C y del 5 al 40 % de las uridinas son análogos de U. Tal como se da a conocer en el presente documento, en una molécula de ARN/molécula de polirribonucleótido modificada de este tipo del 5 al 30 % de las citidinas son análogos de C y del 5 al 30 % de las uridinas son análogos de U. Tal como se da a conocer en el presente documento, en una molécula de ARN/molécula de polirribonucleótido modificada de este tipo del 10 al 30 % de las citidinas son análogos de C y del 10 al 30 % de las uridinas son análogos de U. Tal como se da a conocer en el presente documento, en una molécula de polirribonucleótido modificada de este tipo del 5 al 20 % de las citidinas son análogos de C y del 5 al 20 % de las uridinas son análogos de U. Tal como se da a conocer en el presente documento, en una molécula de ARN/molécula de polirribonucleótido modificada de este tipo del 5 al 10 % de los nucleótidos de citidina y del 5 al 10 % de los nucleótidos de uridina están modificados. Tal como se da a conocer en el presente documento, en una molécula de ARN/molécula de polirribonucleótido modificada de este tipo el 25 % de los nucleótidos de citidina y el 25 % de los nucleótidos de uridina están modificados. Tal como se da a conocer en el presente documento, los nucleótidos que contienen adenosina y guanosina pueden no estar modificados. Tal como se da a conocer en el presente documento, los nucleótidos de adenosina y guanosina pueden no estar modificados o estar parcialmente modificados, y preferiblemente están presentes en una forma no modificada. Tal como se indicó anteriormente, tal como se da a conocer en el presente documento, análogos de U se refiere a un único tipo de análogo de U. Tal como se da a conocer en el presente documento, análogos de U se refiere a dos o más tipos de análogos de U. Análogos de C se refiere a un único tipo de análogo de C. Tal como se da a conocer en el presente documento, análogos de C se refiere a dos o más tipos de análogos de C.

Tal como se da a conocer en el presente documento, el porcentaje de citidinas en una molécula de ARN/polirribonucleótido que son análogos de citidina no es el mismo que el porcentaje de uridinas en la molécula de ARN/polirribonucleótido que son análogos de uridina. Tal como se da a conocer en el presente documento, el porcentaje de análogos de citidina es inferior al porcentaje de análogos de uridina. Tal como se indicó anteriormente, esto puede ser en presencia o ausencia de análogos de adenosina y guanosina pero es en ausencia de análogos de adenosina y análogos de guanosina. Tal como se da a conocer en el presente documento, los polirribonucleótidos de la divulgación comprenden menos del 15 %, menos del 10 %, menos del 5 % o menos del 2 % de análogos de adenosina, análogos de guanosina o ambos.

Una molécula de ARN/polirribonucleótido tal como se da a conocer en el presente documento puede comprender análogos de citidina y análogos de uridina, y del 5 al 20 % de las citidinas son análogos de citidina y del 25 al 45 % de las uridinas son análogos de uridina. Dicho de otro modo, la molécula de ARN/polirribonucleótido comprende

citidinas modificadas y no modificadas y uridinas modificadas y no modificadas, y del 5 al 20 % de las citidinas comprenden análogos de citidina mientras que del 25 al 45 % de las uridinas comprenden análogos de uridina. Tal como se da a conocer en el presente documento, la molécula de ARN/polirribonucleótido comprende del 5 al 10 % de análogos de citidina y del 30 al 40 % de análogos de uridina, tal como el 7-9 % análogos de citidina, tal como aproximadamente el 7, el 7,5 o el 8 % y, tal como el 32-38 % de análogos de uridina, tal como aproximadamente el 33, el 34, el 35, el 36 %.

Tal como se da a conocer en el presente documento, puede usarse cualquiera de los análogos de uridina y análogos de citidina descritos en el presente documento, excluyendo opcionalmente pseudouridina. Tal como se da a conocer en el presente documento, el análogo de citidina comprende o consiste en (por ejemplo, en el caso de consistir en, es el único tipo de análogo usado) 5-yodocitidina y el análogo de uridina comprende o consiste en (por ejemplo, en el caso de consistir en, es el único tipo de análogo usado) 5-yodouridina.

Tal como se da a conocer en el presente documento en cualquiera de lo anterior, el porcentaje de análogos de un nucleótido dado se refiere al porcentaje de entrada (por ejemplo, el porcentaje de análogos en una reacción de partida, tal como una reacción de transcripción *in vitro* de partida). Tal como se da a conocer en cualquiera de lo anterior, el porcentaje de análogos de un nucleótido dado se refiere a la salida (por ejemplo, el porcentaje en un compuesto sintetizado o transcrito).

Las moléculas de ARN/moléculas de polirribonucleótido tal como se da a conocer en el presente documento pueden producirse de manera recombinante en sistemas *in vivo* mediante métodos conocidos por un experto en la técnica que se describen en más detalle a continuación.

Alternativamente, las moléculas de polirribonucleótido modificadas tal como se da a conocer en el presente documento pueden producirse en un sistema *in vitro* usando, por ejemplo, un sistema de transcripción *in vitro* que se describe en más detalle a continuación. Un sistema de transcripción *in vitro* que puede producir moléculas de ARN/polirribonucleótidos requiere una mezcla de entrada de trifosfatos de nucleósido modificados y no modificados para producir moléculas de ARN/polirribonucleótidos modificados con las propiedades deseadas de la presente invención. Tal como se da a conocer en el presente documento, del 5 al 50 % de las citidinas son análogos de citidina en una mezcla de entrada de este tipo y del 5 al 50 % de las uridinas son análogos de uridina en una mezcla de entrada de este tipo. Tal como se da a conocer en el presente documento, del 5 al 40 % de las citidinas son análogos de citidina en una mezcla de entrada de este tipo y del 5 al 40 % de las uridinas son análogos de uridina en una mezcla de entrada de este tipo. Tal como se da a conocer en el presente documento, del 5 al 30 % de las citidinas son análogos de citidina en una mezcla de este tipo y del 5 al 30 % de las uridinas son análogos de uridina en una mezcla de entrada de este tipo. Tal como se da a conocer en el presente documento, del 5 al 30 % de las citidinas son análogos de citidina en tal mezcla y del 10 al 30 % de las uridinas son análogos de uridina en tal mezcla. Tal como se da a conocer en el presente documento, del 5 al 20 % de las citidinas son análogos de citidina en una mezcla de entrada de este tipo y del 5 al 20 % de las uridinas son análogos de uridina en una mezcla de entrada de este tipo. Tal como se da a conocer en el presente documento, del 5 al 10 % de las citidinas son análogos de citidina en una mezcla de entrada de este tipo y del 5 al 10 % de las uridinas son análogos de uridina en una mezcla de entrada de este tipo. Tal como se da a conocer en el presente documento, el 25 % de las citidinas son análogos de citidina en una mezcla de entrada de este tipo y el 25 % de las uridinas son análogos de uridina en una mezcla de entrada de este tipo. Tal como se da a conocer en el presente documento, la mezcla de entrada no comprende análogos de adenosina y/o guanósina. Tal como se da a conocer en el presente documento, opcionalmente, la mezcla de entrada comprende uno o más análogos de adenosina y/o guanósina (o ninguno de los dos o ambos).

Tal como se da a conocer en el presente documento, el porcentaje de citidinas en una mezcla de entrada que son análogos de citidina no es el mismo que el porcentaje de uridinas en una mezcla de entrada que son análogos de uridina. Tal como se da a conocer en el presente documento, el porcentaje de análogos de citidina en una mezcla de entrada es inferior al porcentaje de análogos de uridina en una mezcla de entrada. Tal como se indicó anteriormente, esto puede ser en presencia o ausencia de análogos de adenosina y guanósina en la mezcla de entrada pero, en determinadas realizaciones, es en ausencia de análogos de adenosina y análogos de guanósina en la mezcla de entrada.

Tal como se da a conocer en el presente documento, una mezcla de entrada de nucleótidos para un sistema de transcripción *in vitro* que produce una molécula de ARN/polirribonucleótido de la presente invención comprende análogos de citidina y análogos de uridina, y del 5 al 20 % de las citidinas de la mezcla de entrada son análogos de citidina y del 25 al 45 % de las uridinas de la mezcla de entrada son análogos de uridina. Dicho de otro modo, la mezcla de entrada comprende citidinas modificadas y no modificadas y uridinas modificadas y no modificadas, y del 5 al 20 % de las citidinas de la mezcla de entrada comprenden análogos de citidina mientras que del 25 al 45 % de las uridinas de la mezcla de entrada comprenden análogos de uridina. En otras realizaciones, la mezcla de entrada comprende del 5 al 10 % de análogos de citidina y del 30 al 40 % de análogos de uridina, tal como el 7-9 % de análogos de citidina, tal como el 7, el 7,5 o el 8 % y, tal como el 32-38 % de análogos de uridina, tal como el 33, el 34, el 35, el 36 %.

Tal como se da a conocer en el presente documento, puede usarse cualquiera de los análogos de uridina y análogos

de citidina descritos en el presente documento, excluyendo opcionalmente pseudouridina.

Tal como se da a conocer en el presente documento, el análogo de citidina comprende o consiste en (por ejemplo, es el único tipo de análogo de C usado) 5-yodocitidina y el análogo de uridina comprende o consiste en (por ejemplo, es el único tipo de análogo de U usado) 5-yodouridina.

- 5 En las tablas anteriores se describen análogos a modo de ejemplo. Debe entenderse que para polirribonucleótidos modificados que codifican para el polipéptido deseado (módulo (a)), los análogos y el nivel de modificación, a menos que se indique lo contrario, se consideran a lo largo de todo el polirribonucleótido que codifica para el polipéptido deseado (módulo (a)), incluyendo las regiones no traducidas en 5' y 3' (por ejemplo, el nivel de modificación se basa en razones de entrada de análogos en una reacción de transcripción *in vitro* de tal manera que pueden incorporarse análogos en las posiciones que se transcriben).

Además, las moléculas de ARN/moléculas de polirribonucleótido modificadas pueden sintetizarse químicamente, por ejemplo, mediante síntesis química convencional en un sintetizador de secuencias de nucleótidos automatizado usando un soporte de fase sólida y técnicas convencionales o mediante síntesis química de las secuencias de ADN respectivas y posterior transcripción *in vitro* o *in vivo* de las mismas.

- 15 En biología molecular y genética, en el sentido de 5' y en el sentido de 3' se refieren ambos a una posición relativa en una molécula de ARN. En el contexto de la presente invención, en el sentido de 5' es hacia el extremo 5' de la molécula de ARN y en el sentido de 3' es hacia el extremo 3' de la molécula.

Por consiguiente, la UTR definida en el punto (b), anteriormente, puede estar ubicada directamente en el sentido de 5' de la región codificante según el punto (a), de manera más específica, directamente en el sentido de 5' del codón de iniciación de la región codificante. Por tanto, "directamente en el sentido de 5'" en este contexto significa que no hay ningún nucleótido adicional entre la UTR definida en el punto (b) y la secuencia codificante que se inicia con un codón de iniciación. Por tanto, la región codificante que se inicia con un codón de iniciación está inmediatamente adyacente a dicha secuencia de UTR.

La molécula de ARN puede estar presente en forma de secuencias de ARN fusionadas de los módulos (a) y (b) (definidos en los puntos (a) y (b), de manera respectiva, anteriormente) es decir, una molécula de ARN (de fusión) que se forma mediante la expresión de un gen híbrido realizado combinando al menos dos secuencias de nucleótidos que codifican para dichos módulos. Normalmente, tal como se explicará en más detalle a continuación, esto puede lograrse clonando un ADNc en un vector de expresión lo cual permite la transcripción para dar la molécula de ARN. Por consiguiente, la molécula de ADN que codifica para la molécula de ARN puede ser una secuencia de ADN fusionada, es decir, una molécula química que se forma uniendo dos o más polinucleótidos mediante el grupo fosfato de un nucleótido unido al carbono en 3' en otro nucleótido, formando un enlace fosfodiéster entre los extremos respectivos de un módulo y el extremo de otra molécula. De esta manera, las moléculas de ADN anteriores que codifican para dichos al menos dos módulos se unen entre sí en forma de una molécula de ADN. Después se transcribe una molécula de ADN recombinante de este tipo para dar su secuencia de ácido nucleico de ARN correspondiente.

R<sub>2</sub> se selecciona del grupo que consiste en:

- (i) GGGAGA (SEQ ID NO: 7);  
 (ii) GGGAGA (SEQ ID NO: 8);  
 (iii) GAAG (SEQ ID NO: 9); y  
 40 (iv) GGGA (SEQ ID NO: 10).

Tal como se da a conocer en el presente documento, la molécula de ARN que comprende la secuencia R<sub>2</sub>-CNGCCACC (SEQ ID NO: 2) puede ser una molécula de ARN, en la que el nucleótido N en la posición 2 de SEQ ID NO: 2 es un nucleótido seleccionado del grupo que consiste en U, G o C y en la que el nucleótido N no es una A.

Dicho nucleótido N en la posición 2 de SEQ ID NO: 2 es U.

45 Tal como se da a conocer en el presente documento, la molécula de ARN de la presente invención es una molécula de ARN en la que el nucleótido que sigue directamente en el sentido de 3' del codón de iniciación no es el nucleótido G. Tal como se da a conocer en el presente documento, la molécula de ARN es una molécula de ARN en la que el nucleótido que sigue directamente en el sentido de 3' del codón de iniciación es un nucleótido seleccionado del grupo que consiste en A, U y C.

50 Tal como se da a conocer en el presente documento, la molécula de ARN es una molécula de ARN que comprende un módulo (b1) tal como se definió anteriormente, en la que dicho módulo (b1) es una secuencia en la que la C en la posición 6 de SEQ ID NO: 1 está sustituida por una A y la C en la posición 7 de SEQ ID NO: 1 está sustituida por una G; y/o la A en la posición 5 de SEQ ID NO: 1 está sustituida por una G y en la que el nucleótido que sigue directamente en el sentido de 3' del codón de iniciación es un nucleótido seleccionado del grupo que consiste en A,

U y C.

5 Tal como se da a conocer en el presente documento, la molécula de ARN es una molécula de ARN que comprende un módulo (b2) tal como se definió anteriormente, en la que dicho módulo (b2) es una secuencia en la que la C en la posición 7 de SEQ ID NO: 2 está sustituida por una A y la C en la posición 8 de SEQ ID NO: 2 está sustituida por una G; y/o la A en la posición 6 de SEQ ID NO: 2 está sustituida por una G y en la que el nucleótido que sigue directamente en el sentido de 3' del codón de iniciación es un nucleótido seleccionado del grupo que consiste en A, U y C.

10 Tal como se mencionó anteriormente, la secuencia consenso de Kozak (gcc)gccRccAUGG puede ser, entre otras cosas, variable con respecto al nucleótido en la posición -3 (es decir, 3 nucleótidos en el sentido de 5' desde el codón de iniciación AUG) representado por una "R" siempre que esta posición sea una purina (es decir, adenina o guanina). En las UTR anteriormente descritas, se define que el nucleótido correspondiente a esta posición es una "A". Sin embargo, la presente divulgación también se refiere a moléculas de ARN que comprenden una UTR correspondiente que tiene una "G" en esta posición.

15 Por consiguiente, la molécula de ARN tal como se da a conocer en el presente documento puede contener una UTR según el punto (b1), anteriormente, en la que en dicha secuencia de UTR de (b1), la A en la posición 5 de SEQ ID NO: 1 está sustituida por una G; o en la que en dicha secuencia de UTR de (b2) la A en la posición 6 de SEQ ID NO: 2 está sustituida por una G.

20 Tal como se mencionó anteriormente, la molécula de ARN de la presente invención también puede albergar una cola de poli-A. Tal como se usa en el presente documento, una cola de poli-A se refiere a una secuencia de nucleótidos de adenina ubicada en el extremo 3' del ARN. Una cola de poli-A se añade habitualmente al extremo 3' del ARN mediante un proceso denominado poliadenilación. Por tanto, la presente invención se refiere a cualquiera del ARN anteriormente descrito, en el que la molécula de ARN comprende una cola de poli-A en el extremo 3'.

25 La longitud de la cola de poli-A no está particularmente limitada. Sin embargo, en realizaciones preferidas, la molécula de ARN de la presente invención comprende una cola de poli-A en el extremo 3' en la que la cola de poli-A tiene una longitud de al menos 50, 60, 70, 80, 90, 100 ó 110 nucleótidos. En una realización más preferida, la molécula de ARN de la presente invención comprende una cola de poli-A en el extremo 3' en la que la cola de poli-A tiene una longitud de al menos 120 nucleótidos. En otras realizaciones preferidas, la molécula de ARN de la presente invención comprende una cola de poli-A en el extremo 3' en la que la cola de poli-A tiene una longitud de al menos 150, 200, 250, 300, 350, 400, 500, 600, 700, 800, 900 ó 1000 nucleótidos.

30 En el caso en el que la molécula de ARN de la presente invención se produce mediante un método de transcripción *in vitro* tal como se describe a continuación en el presente documento la cola de poli-A está ubicada en el extremo 3' del ARN adyacente a la UTR en el extremo 3' de la molécula de ARN mientras que el plásmido que alberga la molécula de ARN de la presente invención se linealiza antes de la transcripción *in vitro* en el sentido de 3' de la cola de poli-A con el fin de garantizar que la molécula de ARN transcrita *in vitro* contiene dicha cola de poli-A.

35 Tal como se mencionó anteriormente, la molécula de ARN puede estar presente en forma de secuencias de ARN fusionadas de los módulos (a) y (b), es decir, una molécula de ARN (de fusión) que se forma mediante la transcripción de un gen híbrido realizado combinando al menos dos secuencias de nucleótidos que codifican para dichos módulos. Normalmente, esto se logra clonando un ADNc en un vector de expresión lo cual permite la transcripción de toda la molécula de ARN. Se conoce una variedad de métodos para realizar constructos de fusión, incluyendo síntesis de ácido nucleico, hibridación y/o amplificación para producir una molécula de ácido nucleico bicatenaria sintética "que codifica para" la molécula de ARN de la presente invención. Una molécula de ácido nucleico bicatenaria de este tipo (es decir, molécula de ADN) alberga en una cadena (es decir, en la cadena codificante o sentido) la secuencia de ADN correspondiente a la molécula de ARN de la presente invención y, por consiguiente, "codifica para" la molécula de ARN de la presente invención. Dicho de otro modo, una molécula de ácido nucleico/ADN bicatenaria de este tipo comprende en una cadena la información genética que corresponde a la molécula de ARN transcrita de la presente invención tal como se definió anteriormente en el presente documento. El término "codificante" o "que codifica" en el contexto de la presente invención no se usa sólo en su sentido convencional, es decir, para referirse al ADN de un gen que codifica para una proteína (y, por consiguiente, la información genética que puede traducirse para dar una secuencia de aminoácidos de un polipéptido o una proteína). En vez de eso, en términos de la presente divulgación, en un constructo en el que las secuencias de ADN individuales que codifican para los módulos (a) y (b) están "fusionadas" o unidas para dar una única molécula de ADN (quimérica), el constructo también comprende un componente (es decir, módulo (b)) que no se traduce para dar una proteína. No obstante, la secuencia de ADN correspondiente al módulo (b) proporciona la información, es decir, el "código", para la estructura de las 5'UTR y, por consiguiente, el término "que codifica" en la presente invención también se refiere a la información genética para las UTR que puede expresarse, es decir, transcribirse, si, por ejemplo, está presente en una molécula de ácido nucleico bicatenaria. Por tanto, el término "que codifica" en el contexto de la presente invención, aunque habitualmente sólo se usa para referirse a la codificación/expresión de una proteína, debe entenderse de una manera que la molécula de ácido nucleico puede transcribirse para dar la molécula de ARN correspondiente que alberga partes que codifican para una proteína o un polipéptido (es decir, el módulo (a)) y partes "que codifican" para la UTR (es decir, el módulo (b)) en la que estas últimas representan el

producto final cuando se expresa dado que las UTR no se traducen para dar proteínas o polipéptidos. Un ácido nucleico bicatenario de este tipo puede insertarse en vectores de expresión mediante técnicas de biología molecular convencionales (véase, por ejemplo Sambrook *et al.*, Molecular Cloning, A laboratory manual, 2ª ed, 1989). El término "vector" tal como "vector de expresión" o "vector de clonación" en el sentido de la presente invención se entiende como una unidad circular, bicatenaria, de ADN que se replica dentro de una célula independientemente del ADN cromosómico y que se usa como vehículo para transportar material genético al interior de una célula, en la que puede replicarse y/o expresarse (es decir, transcribirse para dar ARN y traducirse para dar una secuencia de aminoácidos). Un vector que contiene ADN foráneo se denomina ADN recombinante. El propio vector es generalmente una secuencia de ADN que consiste normalmente en un inserto (es decir, módulo (b) que no se traduce para dar una proteína y módulo (a), la región codificante) y una secuencia más grande que sirve como "estructura principal" del vector. Los plásmidos en el sentido de la presente invención se encuentran lo más frecuentemente en bacterias y se usan en investigación de ADN recombinante para transferir genes entre células y, como tal, son una subpoblación de "vectores" tal como se usa en el sentido de la presente invención.

Por tanto, la presente invención también se refiere a una molécula de ácido nucleico que codifica para la molécula de ARN de la presente invención.

El ácido nucleico es, por ejemplo, un ADN, que codifica para los dos módulos principales (es decir, el módulo (a) y módulo el (b)) de la molécula de ARN de la presente invención. La molécula de ácido nucleico anterior de la presente invención es preferiblemente una molécula de ácido nucleico recombinante. La molécula de ácido nucleico de la invención puede ser sintética o semisintética.

Resulta evidente para el experto en la técnica que pueden añadirse secuencias reguladoras adicionales a la molécula de ácido nucleico de la invención que codifican para la molécula de ARN. Por ejemplo, pueden emplearse potenciadores de la transcripción y/o secuencias que permiten la expresión inducida. Un sistema inducible adecuado es, por ejemplo, expresión génica regulada por tetraciclina tal como se describe, por ejemplo, por Gossen y Bujard, Proc. Natl. Acad. Sci. USA 89 (1992), 5547-5551) y Gossen, Trends Biotech. 12 (1994), 58-62, o un sistema de expresión génica inducible por dexametasona tal como se describe, por ejemplo por Crook, EMBO J. 8 (1989), 513-519.

La presente invención también se refiere a un vector, preferiblemente un vector de expresión, que comprende la molécula de ácido nucleico de la presente invención.

Con respecto a los vectores que comprenden una molécula de ácido nucleico que codifica para la molécula de ARN de la presente invención se aplica lo mismo, cambiando lo que sea necesario, que se ha expuesto anteriormente en el contexto de los vectores que comprenden la molécula de ADN de la presente invención tal como se definió anteriormente.

La presente invención también se refiere a una célula huésped que comprende el vector de la presente invención. Por tanto, la presente invención se refiere a un huésped transfectado o transformado con el vector de la invención o a un huésped no humano que porta el vector de la presente invención, es decir a una célula huésped o a un huésped que está genéticamente modificado con una molécula de ácido nucleico según la invención o con un vector que comprende una molécula de ácido nucleico de este tipo.

Con respecto a la célula huésped que comprende el vector que comprende una molécula de ácido nucleico que codifica para la molécula de ARN de la presente invención se aplica lo mismo, cambiando lo que sea necesario, que se ha expuesto anteriormente en el contexto de las células huésped que comprenden los vectores que comprenden la molécula de ADN de la presente invención tal como se definió anteriormente.

También se dan a conocer métodos de producción de la molécula de ARN de la presente invención cultivando una célula huésped que alberga un vector de expresión que codifica para los módulos individuales de la presente invención o la molécula de ARN completa de la invención en medio de cultivo, y recuperando la molécula de ARN a partir de la célula huésped o el medio de cultivo. La presente divulgación también puede referirse a un método para producir una molécula de ARN de la presente invención que comprende cultivar la célula huésped de la presente invención y opcionalmente recuperar la molécula de ARN a partir del cultivo. El experto en la técnica conoce métodos de recuperación y/o posterior purificación de la molécula de ARN de la presente invención.

También se dan a conocer métodos de producción en una reacción *in vitro* de la molécula de ARN de la presente invención mediante métodos conocidos por el experto en la técnica. Más específicamente, la molécula de ARN de la presente invención puede producirse *in vitro* usando un sistema de transcripción *in vitro*. Habitualmente se conocen sistemas de transcripción *in vitro* y requieren habitualmente un molde de ADN lineal purificado que contiene una secuencia de ADN "que codifica para" el módulo (b) y el módulo (a) tal como se expuso anteriormente, en la que dicha secuencia de ADN está bajo el control de un promotor apropiado. Además, un sistema de transcripción *in vitro* también requiere habitualmente trifosfatos de ribonucleótidos, un sistema de tampón que incluye DTT e iones de magnesio, y una ARN polimerasa apropiada que proporciona la actividad enzimática para la transcripción *in vitro* de la secuencia de ADN para dar la molécula de ARN de la presente invención.

El experto en la técnica conoce bien métodos que se usan habitualmente para producir moléculas de ARN usando

transcripción *in vitro* y se describen, por ejemplo, en Methods Mol. Biol. 703 (2011):29-41.

5 Tal como se mencionó anteriormente, en el caso en el que la molécula de ARN de la presente invención se produce mediante un método de transcripción *in vitro* tal como se describe a continuación en el presente documento, la cola de poli-A anterior puede ser parte de la molécula de ARN de la presente invención (y no está necesariamente ubicada originalmente en el vector de clonación) y está ubicada en el extremo 3' del ARN, por ejemplo adyacente a la UTR en el extremo 3' de la molécula de ARN. En el caso en el que la molécula de ARN de la presente invención se produce mediante un método de transcripción *in vitro* el plásmido que alberga la molécula de ARN de la presente invención se linealiza antes de la transcripción *in vitro* en el sentido de 3' de la cola de poli-A con el fin de garantizar que la molécula de ARN transcrita *in vitro* contiene dicha cola de poli-A.

10 Alternativamente, la molécula de ARN también puede sintetizarse químicamente, por ejemplo, mediante síntesis química convencional en un sintetizador de secuencias de nucleótidos automatizado usando un soporte de fase sólida y técnicas convencionales.

15 También se dan a conocer métodos de producción en una reacción *in vitro* de la molécula de ARN de la presente invención mediante métodos conocidos por el experto en la técnica y tal como se expuso anteriormente y recuperación de la molécula de ARN a partir de la reacción. El experto en la técnica conoce métodos de recuperación y/o posterior purificación de la molécula de ARN de la presente invención.

La molécula de ARN de la presente invención puede usarse fácilmente en sistemas de traducción *in vitro* conocidos en la técnica para la expresión eficiente de cualquier polipéptido o proteína deseado codificado por la región codificante del módulo (a).

20 En la técnica se conocen sistemas de traducción *in vitro* y pueden usarse directamente con la molécula de ARN de la presente invención. Alternativamente, estos sistemas de traducción *in vitro* pueden combinarse con los sistemas de transcripción *in vitro* anteriores. Se conocen y están disponibles sistemas libres de células correspondientes para la transcripción *in vitro* y/o traducción *in vitro*. Estos sistemas libres de células para la síntesis de proteínas (también denominados síntesis de proteínas *in vitro* o abreviados como CFPS), permiten la expresión/producción de un polipéptido o una proteína usando maquinaria biológica sin el uso de células vivas. En estos sistemas, el entorno de síntesis de proteínas *in vitro* no está restringido por una pared celular o condiciones de homeostasis necesarias para mantener la viabilidad celular y permite el acceso directo y control del entorno de traducción que es ventajoso para varias aplicaciones incluyendo optimización de producción de proteínas, optimización de complejos de proteínas, para estudiar la síntesis de proteínas, incorporar aminoácidos no naturales, selecciones de alto rendimiento y biología sintética. Los componentes habituales de una reacción libre de células incluyen un extracto celular, una fuente de energía, un suministro de aminoácidos, cofactores tales como magnesio, y el ADN o ARN que codifica para el polipéptido o proteína deseado. Un extracto celular puede obtenerse sometiendo a lisis la célula de interés y separando por centrifugación las paredes celulares, genoma de ADN y otros residuos. Los restos son la maquinaria celular necesaria incluyendo ribosomas, aminoacil-ARNt sintetasas, factores de iniciación de la traducción y elongación, nucleasas, etc. En un sistema libre de células para la síntesis de polipéptidos o proteínas a partir de ADN (es decir, en un sistema que incluye una etapa de transcripción *in vitro* y traducción *in vitro*), habitualmente se usan dos tipos de ADN, es decir, o bien plásmidos o bien moldes de expresión lineales (LET). En un sistema libre de células para la síntesis de polipéptidos o proteínas a partir de ARN (es decir, en un sistema que sólo incluye una etapa de traducción *in vitro*) puede usarse directamente un ARN. Estas reacciones libres de células *in vitro* requieren una fuente de energía que se proporciona habitualmente mediante una mezcla independiente que contiene la fuente de energía necesaria, junto con un suministro de aminoácidos que se añaden al extracto para la reacción. Fuentes de energía habituales son piruvato de fosfoenol, fosfato de acetilo y fosfato de creatina. Extractos celulares habituales que se usan habitualmente se realizan a partir de *Escherichia coli* (ECE), reticulocitos de conejo (RRL), germen de trigo (WGE) y células de insectos (ICE). Todos estos extractos están comercialmente disponibles.

45 Por consiguiente, la presente invención también se refiere al uso de una molécula de ARN de la presente invención para la traducción *in vitro* de un polipéptido o una proteína deseado codificado por una región codificante contenida en dicha molécula de ARN.

50 Con respecto a las realizaciones preferidas de un uso de este tipo de una molécula de ARN de la presente invención, se aplica lo mismo, cambiando lo que sea necesario, que se ha expuesto anteriormente en el contexto de la molécula de ARN tal como se definió anteriormente.

55 Las moléculas de ARN tal como se definieron anteriormente son particularmente útiles en entornos médicos y en el tratamiento de una determinada enfermedad y, en particular, en terapias basadas en ARN. Por tanto, la presente invención también se refiere a una composición farmacéutica que comprende la molécula de ARN de la presente invención, la molécula de ácido nucleico de la presente invención, el vector de la presente invención o la célula huésped de la presente invención y opcionalmente un portador farmacéuticamente aceptable.

El término "tratamiento" y similares se usan en el presente documento para significar de manera general obtener un efecto farmacológico y/o fisiológico deseado. Por consiguiente, el tratamiento de la presente invención puede referirse al tratamiento de estados (agudos) de una determinada enfermedad pero también puede referirse al

tratamiento profiláctico en cuanto a prevenir completa o parcialmente una enfermedad o síntoma de la misma. Preferiblemente, debe entenderse que el término "tratamiento" es terapéutico en cuanto a curar parcial o completamente una enfermedad y/o efecto adverso y/o síntomas atribuidos a la enfermedad. Con respecto a esto, "agudo" significa que el sujeto muestra síntomas de la enfermedad. Dicho de otro modo, el sujeto que va a tratarse necesita realmente un tratamiento y el término "tratamiento agudo" en el contexto de la presente invención se refiere a las medidas tomadas para tratar realmente la enfermedad después de la aparición de la enfermedad o el brote de la enfermedad. El tratamiento también puede ser un tratamiento profiláctico o preventivo, es decir, medidas tomadas para la prevención de la enfermedad, por ejemplo, con el fin de prevenir la infección y/o la aparición de la enfermedad.

La composición farmacéutica de la presente invención puede administrarse mediante una amplia gama de clases de formas de administración conocidas por el experto. La administración puede ser sistemática, local, oral, mediante aerosoles incluyendo, pero sin limitarse a, comprimidos, inyección con aguja, el uso de inhaladores, cremas, espumas, geles, lociones y pomadas.

Tal como se menciona, la presente invención se refiere a una composición farmacéutica, que comprende una cantidad eficaz de la molécula de ARN (o la molécula de ácido nucleico, el vector o la célula huésped) de la presente invención según lo anterior y al menos un excipiente o portador farmacéuticamente aceptable.

Un excipiente o portador es una sustancia inactiva formulada junto con el principio activo, es decir, la molécula de ARN (o la molécula de ácido nucleico, el vector o la célula huésped) de la presente invención con el fin de dar volumen a formulaciones que contienen potentes principios activos. Los excipientes se denominan con frecuencia "agentes de volumen", "cargas" o "diluyentes". Dar volumen permite la dispersión conveniente y precisa de un principio activo cuando se produce una forma de dosificación. También pueden servir para diversos fines de potenciación terapéutica, tales como facilitar la solubilidad o absorción de fármaco, u otras consideraciones farmacocinéticas. Los excipientes también pueden ser útiles en el procedimiento de fabricación, para ayudar en la manipulación del principio activo en cuestión tal como facilitando la fluidez de polvo o propiedades antiadherentes, además de ayudar a la estabilidad *in vitro* tal como prevención de la desnaturalización a lo largo de la semivida esperada. La selección de excipientes apropiados también depende de la vía de administración y la forma de dosificación, así como del principio activo y otros factores.

Por tanto, la composición farmacéutica que comprende una cantidad eficaz de la molécula de ARN (o la molécula de ácido nucleico, el vector o la célula huésped) de la presente invención puede estar en forma sólida, líquida o gaseosa y puede estar, entre otras, en una forma de (un) polvo(s), (un) comprimido(s), (una) disolución/disoluciones o (un) aerosol(es). Se prefiere que dicha composición farmacéutica comprenda opcionalmente un portador y/o diluyente farmacéuticamente aceptable.

En la técnica se conocen bien ejemplos de portadores, excipientes y/o diluyentes farmacéuticos adecuados e incluyen soluciones salinas tamponadas con fosfato, agua, emulsiones, tales como emulsiones de aceite/agua, diversos tipos de agentes humectantes, disoluciones estériles, etc. Pueden formularse composiciones que comprenden tales portadores mediante métodos convencional bien conocidos. Estas composiciones farmacéuticas pueden administrarse al sujeto a una dosis adecuada, es decir, en "una cantidad eficaz" que puede determinarse fácilmente por el experto mediante métodos conocidos en la técnica. La pauta posológica se determinará por el médico encargado y los factores clínicos. Tal como se conoce bien en las técnicas médicas, las dosificaciones para cualquier paciente dependen de muchos factores, incluyendo el tamaño del paciente o sujeto, área de superficie corporal, edad, el compuesto particular que va a administrarse, sexo, tiempo y vía de administración, salud general y otros fármacos que estén administrándose de manera simultánea.

Por tanto, preferiblemente, la molécula de ARN (o la molécula de ácido nucleico, el vector o la célula huésped) de la presente invención se incluye en una cantidad eficaz. El término "cantidad eficaz" se refiere a una cantidad suficiente para inducir una respuesta terapéutica detectable en el sujeto al que va a administrarse la composición farmacéutica. Según lo anterior, el contenido de la molécula de ARN (o la molécula de ácido nucleico, el vector o la célula huésped) de la presente invención en la composición farmacéutica no está limitado siempre que sea útil para el tratamiento tal como se describió anteriormente, pero preferiblemente contiene el 0,0000001-10 % en peso por composición total. Además, la molécula de ARN (o la molécula de ácido nucleico, el vector o la célula huésped) descrita en el presente documento se emplea preferiblemente en un portador. Generalmente, se usa una cantidad apropiada de una sal farmacéuticamente aceptable en el portador para hacer que la composición sea isotónica. Los ejemplos del portador incluyen, pero no se limitan, solución salina, solución de Ringer y disolución de dextrosa. Preferiblemente, los excipientes, portadores o estabilizadores aceptables no son tóxicos a las dosificaciones y concentraciones empleadas, incluyendo tampones tales como citrato, fosfato y otros ácidos orgánicos; contraiones formadores de sales, por ejemplo sodio y potasio; polipéptidos de bajo peso molecular (> 10 residuos de aminoácido); proteínas, por ejemplo albúmina sérica o gelatina; polímeros hidrófilos, por ejemplo polivinilpirrolidona; aminoácidos tales como histidina, glutamina, lisina, asparagina, arginina o glicina; hidratos de carbono incluyendo glucosa, manosa o dextrinas; monosacáridos; disacáridos; otros azúcares, por ejemplo sacarosa, manitol, trehalosa o sorbitol; agentes quelantes, por ejemplo EDTA; tensioactivos no iónicos, por ejemplo Tween, Pluronic o polietilenglicol; antioxidantes incluyendo metionina, ácido ascórbico y tocoferol; y/o conservantes, por ejemplo cloruro de octadecildimetilbencil-amonio; cloruro de hexametonio; cloruro de benzalconio, cloruro de bencetonio;



fenol, alcohol butílico o bencílico; alquil-parabenos, por ejemplo metil o propil-parabeno; catecol; resorcinol; ciclohexanol; 3-pentanol; y m-cresol). Se describen portadores adecuados y sus formulaciones en más detalle en Remington's Pharmaceutical Sciences, 17<sup>a</sup> ed., 1985, Mack Publishing Co.

5 Puede monitorizarse la evolución terapéutica mediante evaluación periódica. La molécula de ARN (o la molécula de ácido nucleico, el vector o la célula huésped) de la presente invención o la composición farmacéutica de la invención puede estar en disoluciones, suspensiones y emulsiones estériles, acuosas o no acuosas, así como cremas y supositorios. Ejemplos de disolventes no acuosos son propilenglicol, polietilenglicol, aceites vegetales tales como aceite de oliva, y ésteres orgánicos tales como oleato de etilo. Los portadores acuosos incluyen agua, disoluciones, emulsiones o suspensiones alcohólicas/acuosas, incluyendo solución salina y medios tamponados. También pueden estar presentes conservantes y otros aditivos tales como, por ejemplo, agentes antimicrobianos, antioxidantes, agentes quelantes y gases inertes y similares. Además, la composición farmacéutica de la invención puede comprender agentes adicionales dependiendo del uso previsto de la composición farmacéutica. Dichos agentes pueden ser, por ejemplo, monolaurato de polioxietileno-sorbitano, disponible en el mercado con el nombre comercial Tween, propilenglicol, EDTA, citrato, sacarosa así como otros agentes que sean adecuados para el uso previsto de la composición farmacéutica que conoce bien el experto en la técnica.

Según esta invención, el término "composición farmacéutica" se refiere a una composición para su administración a un paciente, preferiblemente un paciente humano.

La composición farmacéutica de la presente invención puede ser para su uso en terapias basadas en ARN. Tal como se mencionó anteriormente, la molécula de ARN de la presente invención que comprende una "región codificante que codifica para un polipéptido" puede usarse en terapias basadas en ARN en las que la "región codificante que codifica para un polipéptido" codifica para un polipéptido o una proteína terapéutica o farmacéuticamente activo que tiene un efecto terapéutico o preventivo. Por tanto, tal como se da a conocer en el presente documento, la composición farmacéutica de la presente invención puede ser para su uso en terapias basadas en ARN en el tratamiento o la prevención de una enfermedad tal como se mencionó en la tabla 1 anterior. Por consiguiente, las terapias basadas en ARN según la presente invención pueden ser para su uso en el tratamiento o la prevención de una enfermedad tal como se mencionó en la tabla 1 anterior.

Por tanto, la composición farmacéutica puede ser para su uso en terapias basadas en ARN en casos en los que los defectos génicos descritos en la tabla 1 anterior conducen a una enfermedad que puede tratarse entonces o prevenirse mediante una terapia de sustitución de transcrito/terapia de sustitución de enzima con la molécula de ARN de la presente invención, en la que la molécula de ARN comprende una "región codificante para un polipéptido" que codifica para una versión intacta de la proteína o un fragmento funcional de la misma que compensa el gen defectuoso dado a conocer.

La composición farmacéutica de la presente invención puede ser para su uso en terapias basadas en ARN en el tratamiento o la prevención de enfermedades lisosómicas tales como enfermedad de Gaucher, enfermedad de Fabry, MPS I, MPS II (síndrome de Hunter), MPS VI y enfermedades de almacenamiento de glucógeno tales como por ejemplo enfermedad de almacenamiento de glucógeno tipo I (enfermedad de von Giercke), tipo II (enfermedad de Pompe), tipo III (enfermedad de Cori), tipo IV (enfermedad de Andersen), tipo V (enfermedad de McArdle), tipo VI (enfermedad de Hers), tipo VII (enfermedad de Tauri), tipo VII, tipo IX, tipo X, tipo XI (síndrome de Fanconi-Bickel), tipo XI o tipo 0. De manera beneficiosa, las terapias de sustitución de transcrito/terapias de sustitución de enzima no afectan al defecto génico subyacente, sino que aumentan la concentración de la enzima para la que el paciente presenta deficiencia. Como ejemplo, en la enfermedad de Pompe, la terapia de sustitución de transcrito/terapia de sustitución de enzima sustituye la enzima lisosómica deficiente alfa-glucosidasa ácida (GAA).

La composición farmacéutica de la presente invención puede ser para su uso en terapias basadas en ARN según la presente invención en la que la "región codificante que codifica para un polipéptido" codifica para un polipéptido, proteína o péptido terapéutica o farmacéuticamente activo que tiene un efecto terapéutico o preventivo, en la que dicho polipéptido, proteína o péptido se selecciona del grupo codificado por los genes tal como se expone en la tabla 1.

Las terapias basadas en ARN según la presente invención pueden ser para su uso en el tratamiento de cáncer, una enfermedad cardiovascular, una infección viral, una disfunción inmunitaria, una enfermedad autoinmunitaria, un trastorno neurológico, unos trastornos metabólicos heredados o un trastorno genético o cualquier enfermedad en la que una proteína o fragmento de proteína producido en una célula puede tener un efecto beneficioso para el paciente. Los ejemplos de cáncer incluyen cáncer de cabeza y cuello, cáncer de mama, cáncer renal, cáncer de vejiga, cáncer de pulmón, cáncer de próstata, cáncer de huesos, cáncer de cerebro, cáncer de cuello uterino, cáncer anal, cáncer de colon, cáncer colorrectal, cáncer de apéndice, cáncer ocular, cáncer gástrico, leucemia, linfoma, cáncer de hígado, cáncer de piel, cáncer de ovarios, cáncer de pene, cáncer de páncreas, cáncer de testículos, cáncer de tiroides, cáncer de vagina, cáncer de la vulva, cáncer endometrial, cáncer cardíaco y sarcoma.

Los ejemplos de enfermedades cardiovasculares incluyen aterosclerosis, cardiopatía coronaria, cardiopatía pulmonar y cardiomiopatía.

Los ejemplos de disfunciones inmunitarias y enfermedades autoinmunitarias incluyen, pero no se limitan a, enfermedades reumáticas, esclerosis múltiple y asma.

Los ejemplos de infecciones virales incluyen, pero no se limitan a, infecciones con virus de inmunodeficiencia humano, virus del herpes simple, papilomavirus humano así como virus de hepatitis B y C.

- 5 Los ejemplos de trastornos neurológicos incluyen, pero no se limitan a, enfermedad de Parkinson, esclerosis múltiple y demencia.

Los ejemplos de trastornos metabólicos heredados incluyen, pero no se limitan a, enfermedad de Gaucher y fenilcetonuria.

- 10 También se da a conocer un método de una terapia basada en ARN. Por tanto, se da a conocer un método para el tratamiento de una enfermedad tal como cáncer, una enfermedad cardiovascular, una infección viral, una disfunción inmunitaria, una enfermedad autoinmunitaria, un trastorno neurológico, unos trastornos metabólicos heredados o un trastorno genético mediante una terapia basada en ARN. Con respecto a las realizaciones preferidas del método para tratamiento se aplica lo mismo, cambiando lo que sea necesario, que se ha expuesto anteriormente en el contexto de la molécula de ARN o la composición farmacéutica para su uso en terapia basada en ARN tal como se definió anteriormente.
- 15

Tal como se da a conocer en el presente documento, el sujeto puede ser un mamífero tal como un perro, gato, cerdo, vaca, oveja, caballo, roedor, por ejemplo, rata, ratón y cobaya, o un primate, por ejemplo, gorila, chimpancé y ser humano. En una realización muy preferida, el sujeto es un ser humano.

- 20 Tal como se mencionó anteriormente, las moléculas de ARN tal como se definieron anteriormente son particularmente útiles en entornos médicos y en el tratamiento de una determinada enfermedad y, en particular, en terapias basadas en ARN. Por tanto, la presente invención también se refiere a una composición farmacéutica que comprende la molécula de ARN, la molécula de ácido nucleico, el vector o la célula huésped de la presente invención y opcionalmente un portador farmacéuticamente aceptable.

- 25 Sin embargo, en terapias de ARN, con frecuencia es deseable silenciar el efecto de la molécula de ARN en algún punto.

- Esto puede realizarse, por ejemplo, usando un mecanismo de iARN (interferencia de ARN) usando la cadena de ácido nucleico que es complementaria a la secuencia de UTR de la presente invención. De hecho, el pequeño tamaño de las UTR mínimas de la presente invención hace que este enfoque sea viable dado que estas UTR no forman estructuras secundarias o terciarias y no existen en células normales. Por consiguiente, la cadena complementaria de una secuencia de UTR de este tipo puede usarse de manera beneficiosa en entornos médicos después del tratamiento de las enfermedades anteriores o después de las terapias basadas en ARN anteriores que usan la composición farmacéutica de la presente invención, silenciando de ese modo las moléculas de ARN terapéuticas de la presente invención.
- 30

- Por tanto, también se da a conocer un enfoque de iARN para su uso en la preparación de una composición farmacéutica para silenciar el efecto de las moléculas de ARN terapéuticas de la presente invención.
- 35

- El término "interferencia de ARN" o "ARN de inhibición" (iARN/ARNi) describe el uso de ARN bicatenario para seleccionar como diana ARNm específicos para su degradación, silenciando de ese modo su expresión. Pueden seleccionarse moléculas de ARN de inhibición preferidas del grupo que consiste en ARN bicatenario (ARNbc), iARN, ARNic, ARNhc y ARNpt. Se sintetiza *in vitro* ARNbc que coincide con una secuencia génica y se introduce en una célula. También puede introducirse el ARNbc en una célula en forma de un vector que expresa una secuencia génica diana en orientación sentido y antisentido, por ejemplo en forma de un ARNm de horquilla. Las secuencias sentido y antisentido también pueden expresarse a partir de vectores independientes, mediante lo cual las moléculas antisentido y sentido individuales forman ARN bicatenario tras su expresión. En la técnica se sabe que en algunas ocasiones la expresión de una secuencia en orientación sentido o incluso de una secuencia de promotor es suficiente para dar lugar a ARNbc y posteriormente a ARNic debido a mecanismos de amplificación internos en una célula. Por consiguiente, todos los medios y métodos que dan como resultado una disminución de la actividad del polipéptido o la proteína codificado por la región codificante van a usarse según la presente invención. Por ejemplo, pueden usarse constructos sentido, constructos antisentido, constructos de horquilla, moléculas sentido y antisentido y combinaciones de los mismos para generar/introducir estos ARNic. El ARNbc se alimenta a un proceso natural, pero sólo parcialmente entendido, que incluye la nucleasa altamente conservada dicer que escinde moléculas precursoras de ARNbc para dar ARN de interferencia cortos (ARNic). La generación y preparación de ARNic así como el método para inhibir la expresión de un gen diana se describen, entre otros, en el documento WO 02/055693, Wei (2000) Dev. Biol. 15:239-255; La Count (2000) Biochem. Paras. 111:67-76; Baker (2000) Curr. Biol. 10:1071-1074; Svoboda (2000) Development 127:4147-4156 o Marie (2000) Curr. Biol. 10:289-292. Entonces, estos ARNic construyen la parte específica de secuencia de un complejo de silenciamiento inducido por ARN (RISC), una nucleasa de múltiples complejos que destruye ARN mensajeros homólogos al agente desencadenante de silenciamiento). Elbashir (2001) EMBO J. 20:6877-6888 mostró que pueden usarse dúplex de ARN de 21 nucleótidos en cultivo celular para interferir con la expresión génica en células de mamíferos. Ya se conoce que la
- 40
- 45
- 50
- 55

iARN está mediada de manera muy eficiente por ARNic en células de mamíferos pero la generación de líneas celulares estables o animales transgénicos no humanos era limitada. Sin embargo, pueden emplearse nuevas generaciones de vectores con el fin de expresar de manera estable, por ejemplo, ARN de horquilla cortos (ARNhc). La expresión estable de ARNic en células de mamíferos se muestra, entre otros, en Brummelkamp (2002) Science 296:550-553. Paul (2002) Nat. Biotechnol. 20:505-508 también documentó la expresión eficaz de ARN de interferencia pequeño en células humanas. La interferencia de ARN mediante expresión de ARN de interferencia cortos y ARN de horquilla en células de mamíferos también se mostró por Yu (2002) PNAS 99:6047-6052. El enfoque de ARNhc para silenciamiento génico se conoce bien en la técnica y puede comprender el uso de ARNpt (pequeños temporales); véase, entre otros, Paddison (2002) Genes Dev. 16:948-958. Estos enfoques pueden basarse en vectores, por ejemplo el vector pSUPER, o pueden emplearse vectores de ARN polIII tal como se ilustra, entre otros, en Yu (2002), anteriormente citado; Miyagishi (2002), anteriormente citado o Brummelkamp (2002), anteriormente citado. Se prevé que las secuencias reguladoras de la presente invención se usen de una manera similar a los sistemas basados en vectores pSUPER o ARN polIII.

En la técnica se conocen métodos para deducir y construir ARNic y se describen en Elbashir (2002) Methods 26:199-213, en los sitios web de Internet de proveedores comerciales de ARNic, por ejemplo Qiagen GmbH (<https://www1.qiagen.com/GeneGlobe/Default.aspx>); Dharmacon ([www.dharmacon.com](http://www.dharmacon.com)); Xeragon Inc. (<http://www.dharmacon.com/Default.aspx>) y Ambion ([www.ambion.com](http://www.ambion.com)), o en el sitio web del grupo de investigación de Tom Tuschl (<http://www.rockefeller.edu/labheads/tuschl/sima.html>). Además, están disponibles programas en línea para deducir ARNic a partir de una secuencia de ARNm dada (por ejemplo [http://www.ambion.com/techlib/misc/siRNA\\_finder.html](http://www.ambion.com/techlib/misc/siRNA_finder.html) o <http://katahdin.cshl.org:9331/RNAi/html/rnai.html>). Los residuos de uridina en la proyección en 3' de 2 nt pueden sustituirse por 2'-desoxitimidina sin pérdida de actividad, lo cual reduce significativamente los costes de síntesis de ARN y también puede potenciar la resistencia de dúplex de ARNic cuando se aplican a células de mamíferos (Elbashir (2001) anteriormente citado). Los ARNic también pueden sintetizarse de manera enzimática usando ARN polimerasas de T7 u otras (Donze (2002) Nucleic Acids Res 30:e46). También pueden producirse dúplex de ARN corto que median en la interferencia de ARN eficaz (ARNice) mediante hidrólisis con ARNasa III de *Escherichia coli* (Yang (2002) PNAS 99:9942-9947). Además, se han desarrollado vectores de expresión para expresar ARNic bicatenarios conectados mediante bucles de ARN de horquilla pequeños en células eucariotas (por ejemplo (Brummelkamp (2002) Science 296:550-553)). Todos estos constructos pueden desarrollarse con la ayuda de los programas mencionados anteriormente. Además, pueden usarse herramientas de predicción de secuencias comercialmente disponibles, incorporadas en programas de análisis de secuencias o comercializadas por separado, por ejemplo la herramienta de diseño de ARNic ofrecida por [www.oligoEngine.com](http://www.oligoEngine.com) (Seattle, WA), para la predicción de secuencias de ARNic.

Por consiguiente, pueden usarse ARN de interferencia específicos según la presente divulgación como antagonistas/silenciadores de la expresión y/o función del polipéptido o la proteína codificado por la región codificante de la molécula de ARN de la presente invención. Estos ARNic están formados por una cadena complementaria/antisentido y una sentido, mediante lo cual la cadena antisentido/sentido comprende preferiblemente al menos 10, más preferiblemente al menos 12, más preferiblemente al menos 14, más preferiblemente al menos 16, más preferiblemente al menos 18, más preferiblemente al menos 19, 20, 21 ó 22 nucleótidos.

La cadena antisentido/sentido puede comprender preferiblemente 25 o más nucleótidos.

Tal como se mencionó anteriormente, en la técnica se conocen bien métodos para preparar ARNic que van a usarse según la presente divulgación. Basándose en las enseñanzas proporcionadas en el presente documento, un experto en la técnica está fácilmente en posición no sólo de preparar tales ARNic sino también de evaluar si un ARNic puede antagonizar/inhibir/silenciar el polipéptido o la proteína codificado por la región codificante de la molécula de ARN tal como se da a conocer en el presente documento. En el presente documento se prevé que los ARNic anteriormente descritos conducen a una degradación de la molécula de ARN que alberga una región codificante que codifica para un polipéptido o una proteína y un módulo de UTR, y por tanto a una disminución del nivel de polipéptido/proteína del polipéptido o la proteína codificado por la región codificante de la molécula de ARN.

Por consiguiente, se da a conocer una molécula de ARN que es complementaria a una UTR de la presente invención tal como se describió anteriormente en el presente documento. Dicha molécula de ARN complementaria a la UTR puede comprender la secuencia CAUGGUGGCGUCUCCC (SEQ ID NO: 11) o una secuencia que muestra de 1 a 4 sustituciones en comparación con SEQ ID NO: 11.

Dicha molécula de ARN complementaria a la UTR puede comprender la secuencia CAUGGUGGCGUCUCCC (SEQ ID NO: 12), en la que el nucleótido N en la posición 10 de SEQ ID NO: 12 es un nucleótido seleccionado del grupo que consiste en U, G, C o A, o una secuencia que muestra de 1 a 4 sustituciones en comparación con SEQ ID NO: 12 y que puede antagonizar/inhibir/silenciar el polipéptido o la proteína codificado por la región codificante de la molécula de ARN.

Dicha molécula de ARN complementaria a la UTR puede comprender la secuencia CAUCUUGGCGUCUCCC (SEQ ID NO: 13), o una secuencia que muestra de 1 a 4 sustituciones en comparación con SEQ ID NO: 13.

- 5 Dicha molécula de ARN complementaria a la UTR puede comprender la secuencia CAUCUUGGCGUCUCCC (SEQ ID NO: 14), en la que el nucleótido N en la posición 10 de SEQ ID NO: 14 es un nucleótido seleccionado del grupo que consiste en U, G, C o A aunque se prefiere más A, o una secuencia que muestra de 1 a 4 sustituciones en comparación con SEQ ID NO: 14 y que puede antagonizar/inhibir/silenciar el polipéptido o la proteína codificado por la región codificante de la molécula de ARN.
- Dicha molécula de ARN complementaria a la UTR puede comprender la secuencia CAUGGCGGCGUCUCCC (SEQ ID NO: 15), o una secuencia que muestra de 1 a 4 sustituciones en comparación con SEQ ID NO: 15.
- 10 Dicha molécula de ARN complementaria a la UTR puede comprender la secuencia CAUGGCGGCGUCUCCC (SEQ ID NO: 16), en la que el nucleótido N en la posición 10 de SEQ ID NO: 16 es un nucleótido seleccionado del grupo que consiste en U, G, C o A, o una secuencia que muestra de 1 a 4 sustituciones en comparación con SEQ ID NO: 16 y que puede antagonizar/inhibir/silenciar el polipéptido o la proteína codificado por la región codificante de la molécula de ARN.
- Dicha molécula de ARN complementaria a la UTR puede comprender la secuencia CAUCUCGGCGUCUCCC (SEQ ID NO: 17), o una secuencia que muestra de 1 a 4 sustituciones en comparación con SEQ ID NO: 17.
- 15 Dicha molécula de ARN complementaria a la UTR puede comprender la secuencia CAUCUCGGCGUCUCCC (SEQ ID NO: 18), en la que el nucleótido N en la posición 10 de SEQ ID NO: 18 es un nucleótido seleccionado del grupo que consiste en U, G, C o A aunque se prefiere más A, o una secuencia que muestra de 1 a 4 sustituciones en comparación con SEQ ID NO: 18 y que puede antagonizar/inhibir/silenciar el polipéptido o la proteína codificado por la región codificante de la molécula de ARN.
- 20 Dicha molécula de ARN complementaria a la UTR puede comprender la secuencia CAUGGUGGCGUCCC (SEQ ID NO: 19), o una secuencia que muestra de 1 a 4 sustituciones en comparación con SEQ ID NO: 19.
- Dicha molécula de ARN complementaria a la UTR puede comprender la secuencia CAUGGUGGCGUCCC (SEQ ID NO: 20), en la que el nucleótido N en la posición 10 de SEQ ID NO: 20 es un nucleótido seleccionado del grupo que consiste en U, G, C o A, o una secuencia que muestra de 1 a 4 sustituciones en comparación con SEQ ID NO: 20 y que puede antagonizar/inhibir/silenciar el polipéptido o la proteína codificado por la región codificante de la molécula de ARN.
- 25 Dicha molécula de ARN complementaria a la UTR puede comprender la secuencia CAUCUUGGCGUCCC (SEQ ID NO: 21), o una secuencia que muestra de 1 a 4 sustituciones en comparación con SEQ ID NO: 21.
- Dicha molécula de ARN complementaria a la UTR puede comprender la secuencia CAUCUUGGCGUCCC (SEQ ID NO: 22), en la que el nucleótido N en la posición 10 de SEQ ID NO: 22 es un nucleótido seleccionado del grupo que consiste en U, G, C o A aunque se prefiere más A, o una secuencia que muestra de 1 a 4 sustituciones en comparación con SEQ ID NO: 22 y que puede antagonizar/inhibir/silenciar el polipéptido o la proteína codificado por la región codificante de la molécula de ARN.
- 30 Dicha molécula de ARN complementaria a la UTR puede comprender la secuencia CAUGGCGGCGUCCC (SEQ ID NO: 23), o una secuencia que muestra de 1 a 4 sustituciones en comparación con SEQ ID NO: 23.
- 35 Dicha molécula de ARN complementaria a la UTR puede comprender la secuencia CAUGGCGGCGUCCC (SEQ ID NO: 24), en la que el nucleótido N en la posición 10 de SEQ ID NO: 24 es un nucleótido seleccionado del grupo que consiste en U, G, C o A, o una secuencia que muestra de 1 a 4 sustituciones en comparación con SEQ ID NO: 24 y que puede antagonizar/inhibir/silenciar el polipéptido o la proteína codificado por la región codificante de la molécula de ARN.
- 40 Dicha molécula de ARN complementaria a la UTR puede comprender la secuencia CAUCUCGGCGUCCC (SEQ ID NO: 25), o una secuencia que muestra de 1 a 4 sustituciones en comparación con SEQ ID NO: 25.
- Dicha molécula de ARN complementaria a la UTR puede comprender la secuencia CAUCUCGGCGUCCC (SEQ ID NO: 26), en la que el nucleótido N en la posición 10 de SEQ ID NO: 26 es un nucleótido seleccionado del grupo que consiste en U, G, C o A aunque se prefiere más A, o una secuencia que muestra de 1 a 4 sustituciones en comparación con SEQ ID NO: 26 y que puede antagonizar/inhibir/silenciar el polipéptido o la proteína codificado por la región codificante de la molécula de ARN.
- 45 Dicha molécula de ARN complementaria a la UTR puede comprender la secuencia CAUGGUGGCGCUUC (SEQ ID NO: 27), o una secuencia que muestra de 1 a 4 sustituciones en comparación con SEQ ID NO: 27.
- 50 Dicha molécula de ARN complementaria a la UTR puede comprender la secuencia CAUGGUGGCGCUUC (SEQ ID NO: 28), en la que el nucleótido N en la posición 10 de SEQ ID NO: 28 es un nucleótido seleccionado del grupo que consiste en U, G, C o A, o una secuencia que muestra de 1 a 4 sustituciones en comparación con SEQ ID NO: 28 y que puede antagonizar/inhibir/silenciar el polipéptido o la proteína codificado por la región codificante de la molécula de ARN.

Dicha molécula de ARN complementaria a la UTR puede comprender la secuencia CAUCUUGGCGCUUC (SEQ ID NO: 29), o una secuencia que muestra de 1 a 4 sustituciones en comparación con SEQ ID NO: 29.

5 Dicha molécula de ARN complementaria a la UTR puede comprender la secuencia CAUCUUGGCGNGCUUC (SEQ ID NO: 30), en la que el nucleótido N en la posición 10 de SEQ ID NO: 30 es un nucleótido seleccionado del grupo que consiste en U, G, C o A aunque se prefiere más A, o una secuencia que muestra de 1 a 4 sustituciones en comparación con SEQ ID NO: 30 y que puede antagonizar/inhibir/silenciar el polipéptido o la proteína codificado por la región codificante de la molécula de ARN.

Dicha molécula de ARN complementaria a la UTR puede comprender la secuencia CAUGGCGGCGCUUC (SEQ ID NO: 31), o una secuencia que muestra de 1 a 4 sustituciones en comparación con SEQ ID NO: 31.

10 Dicha molécula de ARN complementaria a la UTR puede comprender la secuencia CAUGGCGGCGNGCUUC (SEQ ID NO: 32), en la que el nucleótido N en la posición 10 de SEQ ID NO: 32 es un nucleótido seleccionado del grupo que consiste en U, G, C o A, o una secuencia que muestra de 1 a 4 sustituciones en comparación con SEQ ID NO: 32 y que puede antagonizar/inhibir/silenciar el polipéptido o la proteína codificado por la región codificante de la molécula de ARN.

15 Dicha molécula de ARN complementaria a la UTR puede comprender la secuencia CAUCUCGGCGCUUC (SEQ ID NO: 33), o una secuencia que muestra de 1 a 4 sustituciones en comparación con SEQ ID NO: 33.

20 Dicha molécula de ARN complementaria a la UTR puede comprender la secuencia CAUCUCGGCGNGCUUC (SEQ ID NO: 34), en la que el nucleótido N en la posición 10 de SEQ ID NO: 34 es un nucleótido seleccionado del grupo que consiste en U, G, C o A aunque se prefiere más A, o una secuencia que muestra de 1 a 4 sustituciones en comparación con SEQ ID NO: 34 y que puede antagonizar/inhibir/silenciar el polipéptido o la proteína codificado por la región codificante de la molécula de ARN.

25 También se da a conocer una molécula de ARN seleccionada del grupo que consiste en SEQ ID NO: 11 a 34 que alberga (un) nucleótido(s) adicional(es) en el extremo 5' que se extiende más allá del triplete complementario al codón de iniciación y que es complementaria a las secuencias del polipéptido o la proteína deseado codificado por la región codificante de la molécula de ARN tal como se da a conocer en el presente documento. Preferiblemente, las secuencias complementarias que comprenden las secuencias anteriores complementarias a la secuencias de UTR tal como se da a conocer en el presente documento (es decir, una molécula de ARN seleccionada del grupo que consiste en SEQ ID NO: 11 a 34) comprenden preferiblemente al menos 15, más preferiblemente al menos 16, más preferiblemente al menos 17, más preferiblemente al menos 18, más preferiblemente al menos 19, más preferiblemente al menos 20, 21, 22, 23 ó 24 nucleótidos. Tal como se da a conocer en el presente documento, estas secuencias pueden comprender 25, 30, 35, 40 o más nucleótidos. Puede desearse aumentar la longitud en el extremo 5' con el fin de aumentar la especificidad de la secuencia complementaria previniendo de ese modo efectos secundarios no deseados.

35 La presente divulgación no sólo se refiere a cualquiera de las moléculas de ARN anteriores sino también a una molécula de ARN seleccionada del grupo que consiste en SEQ ID NO: 11 a 34 que comprende hasta el 5 %, el 10 %, el 20 % o el 30 % de coincidencias erróneas con las moléculas de ARN descritas anteriormente. Además, las moléculas de ARN pueden modificarse químicamente tal como se describió anteriormente en el presente documento.

40 La presente invención también se refiere a un kit que comprende una molécula de ADN de la presente invención, una molécula de ARN de la presente invención, una molécula de ácido nucleico de la presente invención, un vector de la presente invención o una célula huésped de la presente invención. Con respecto a las realizaciones preferidas, se aplica lo mismo, cambiando lo que sea necesario, que se ha expuesto anteriormente en el contexto de la molécula de ADN, molécula de ARN, molécula de ácido nucleico, vector o la célula huésped según la presente invención. Ventajosamente, el kit de la presente invención comprende además, opcionalmente, (un) tampón/tampones, disoluciones de almacenamiento y/o reactivos o materiales restantes requeridos para la realización de los usos y métodos anteriores y siguientes. Además, partes del kit de la invención pueden acondicionarse de manera individual en viales o botellas o en combinación en recipientes o unidades de múltiples recipientes. El kit de la presente invención puede usarse ventajosamente, entre otras cosas, para llevar a cabo los métodos de la invención o para la preparación de la molécula de ARN de la invención y puede emplearse en una variedad de aplicaciones a las que se hace referencia en el presente documento, por ejemplo, en los usos tal como se expusieron anteriormente y a continuación. Otro componente que puede incluirse en el kit son instrucciones de uso para una persona que usa un kit. La fabricación de los kits sigue preferiblemente procedimientos convencionales que conoce el experto en la técnica.

55 La presente invención también se refiere al uso de una UTR tal como se describió anteriormente en el presente documento para traducir una región codificante de una molécula de ARN para dar un polipéptido o una proteína codificado por dicha región codificante.

En una realización más preferida, la presente invención también se refiere al uso de una UTR tal como se describió anteriormente en el presente documento para aumentar la eficiencia de traducción de una región codificante de una

molécula de ARN para dar un polipéptido o una proteína codificado por dicha codificación. Con respecto a las realizaciones preferidas del uso se aplica lo mismo, cambiando lo que sea necesario, que se ha expuesto anteriormente en el contexto de la molécula de ARN de la presente invención.

En el contexto de la presente divulgación, también se da a conocer en el mismo los siguientes puntos 1 a 20:

- 5 1. Una molécula de ADN, que puede transcribirse para dar un ARNm, que comprende una cadena con los siguientes elementos:
- (a) una región codificante, que incluye un codón de iniciación en su extremo 5', que codifica para un polipéptido; y
- (b) directamente en el sentido de 5' de dicha secuencia codificante, una secuencia seleccionada del grupo que consiste en:
- 10 (b1) R<sub>1</sub>-CGCCACC (SEQ ID NO: 1);
- o una secuencia en la que en dicha secuencia la C en la posición 6 de SEQ ID NO: 1 está sustituida por una A y la C en la posición 7 de SEQ ID NO: 1 está sustituida por una G; y/o la A en la posición 5 de SEQ ID NO: 1 está sustituida por una G; y
- 15 (b2) R<sub>1</sub>-CNGCCACC (SEQ ID NO: 2), en la que el nucleótido N en la posición 2 de SEQ ID NO: 2 es un nucleótido seleccionado del grupo que consiste en T, G, C o A;
- o una secuencia en la que en dicha secuencia la C en la posición 7 de SEQ ID NO: 2 está sustituida por una A y la C en la posición 8 de SEQ ID NO: 2 está sustituida por una G; y/o la A en la posición 6 de SEQ ID NO: 2 está sustituida por una G,
- en la que R<sub>1</sub> es un promotor que se reconoce por una ARN polimerasa dependiente de ADN;
- 20 o que comprende la cadena complementaria.
2. La molécula de ADN según el punto 1, en la que el promotor que se reconoce por una ARN polimerasa dependiente de ADN se selecciona del grupo que consiste en:
- (i) TAATACGACTCACTATAGGGAGA (SEQ ID NO: 3) o una secuencia que muestra de 1 a 6 sustituciones en comparación con SEQ ID NO: 3 y que se reconoce por una ARN polimerasa dependiente de ADN de T7;
- 25 (ii) AATTAACCCCTACTAAAGGGAGA (SEQ ID NO: 4) o una secuencia que muestra de 1 a 6 sustituciones en comparación con SEQ ID NO: 4 y que se reconoce por una ARN polimerasa dependiente de ADN de T3;
- (iii) ATTTAGGTGACACTATAGAAG (SEQ ID NO: 5) o una secuencia que muestra de 1 a 6 sustituciones en comparación con SEQ ID NO: 5 y que se reconoce por una ARN polimerasa dependiente de ADN de SP6; y
- 30 (iv) AATTAGGGCACACTATAGGGA (SEQ ID NO: 6) o una secuencia que muestra de 1 a 6 sustituciones en comparación con SEQ ID NO: 6 y que se reconoce por una ARN polimerasa dependiente de ADN de K11.
3. La molécula de ADN según el punto 1 ó 2, en la que el nucleótido N en la posición 2 de SEQ ID NO: 2 es un nucleótido seleccionado del grupo que consiste en T, G o C y en la que el nucleótido N no es una A.
4. La molécula de ADN según el punto 3, en la que dicho nucleótido N en la posición 2 de SEQ ID NO: 2 es T.
5. Un vector que comprende la molécula de ADN según el punto 4.
- 35 6. Una célula huésped que comprende el vector según el punto 5.
7. Una composición que comprende:
- la molécula de ADN según uno cualquiera de los puntos 1 a 4, el vector según el punto 5 o la célula huésped según el punto 6.
8. Una molécula de ARN que comprende
- 40 (a) una región codificante, que incluye un codón de iniciación en su extremo 5', que codifica para un polipéptido; y
- (b) directamente en el sentido de 5' de dicha secuencia codificante, una UTR seleccionada del grupo que consiste en:
- (b1) una UTR de la secuencia
- R<sub>2</sub>-CGCCACC (SEQ ID NO: 1),

o una secuencia en la que en dicha secuencia de UTR la C en la posición 6 de SEQ ID NO: 1 está sustituida por una A y la C en la posición 7 de SEQ ID NO: 1 está sustituida por una G; y/o la A en la posición 5 de SEQ ID NO: 1 está sustituida por una G; y

(b2) una UTR de la secuencia

5 R<sub>2</sub>-CNGCCACC (SEQ ID NO: 2), en la que el nucleótido N en la posición 2 de SEQ ID NO: 2 es un nucleótido seleccionado del grupo que consiste en U, G, C o A, o una secuencia en la que en dicha secuencia de UTR la C en la posición 7 de SEQ ID NO: 2 está sustituida por una A y la C en la posición 8 de SEQ ID NO: 2 está sustituida por una G; y/o la A en la posición 6 de SEQ ID NO: 2 está sustituida por una G,

10 en la que R<sub>2</sub> es una secuencia de ARN correspondiente a la parte de una región de promotor que empieza con el nucleótido en el que una ARN polimerasa dependiente de ADN inicia la síntesis de ARN.

9. La molécula de ARN según el punto 8, en la que R<sub>2</sub> se selecciona del grupo que consiste en:

(i) GGGAGA (SEQ ID NO: 7);

(ii) GGGAGA (SEQ ID NO: 8);

(iii) GAAG (SEQ ID NO: 9); y

15 (iv) GGGA (SEQ ID NO: 10).

10. La molécula de ARN según el punto 8 ó 9, en la que el nucleótido N en la posición 2 de SEQ ID NO: 2 es un nucleótido seleccionado del grupo que consiste en U, G o C y en la que el nucleótido N no es una A.

11. La molécula de ARN según el punto 10, en la que dicho nucleótido N en la posición 2 de SEQ ID NO: 2 es U.

20 12. La molécula de ARN según uno cualquiera de los puntos 8 a 11, en la que la molécula de ARN comprende una cola de poli-A en el extremo 3'.

13. La molécula de ARN según uno cualquiera de los puntos 8 a 12, en la que la cola de poli-A tiene una longitud de al menos 120 nucleótidos.

14. Una molécula de ácido nucleico que codifica para la molécula de ARN según uno cualquiera de los puntos 8 a 13.

25 15. Un vector que comprende la molécula de ácido nucleico según el punto 14.

16. Una célula huésped que comprende el vector según el punto 15.

17. Una composición farmacéutica que comprende la molécula de ARN según uno cualquiera de los puntos 8 a 13, la molécula de ácido nucleico según el punto 14, el vector según el punto 15 o la célula huésped según el punto 16 y opcionalmente un portador farmacéuticamente aceptable.

30 18. La composición farmacéutica según el punto 17 para su uso en terapias basadas en ARN.

19. Un kit que comprende la molécula de ADN según uno cualquiera de los puntos 1 a 4, la molécula de ARN según uno cualquiera de los puntos 8 a 13, la molécula de ácido nucleico según el punto 14, el vector según el punto 5 ó 15 o la célula huésped según el punto 6 ó 16.

35 20. Uso de una UTR según el punto 8(b) para traducir una región codificante de una molécula de ARN para dar un polipéptido o una proteína codificado por dicha región codificante.

Figura 1:

Muestra las secuencias que albergan una secuencia de "UTR mínima" junto con el nombre de los constructos de indicadores de luciferasa respectivos usados en la presente invención. Las secuencias albergan partes del promotor de T7 y del elemento de Kozak seguido por un codón de iniciación ATG. Las primeras 10 bases incluyendo la secuencia TATA y las 6 bases posteriores (GGGAGA) son secuencias derivadas de promotor de T7 mientras que las bases restantes en el sentido de 5' del codón de iniciación ATG pertenecen al elemento de Kozak (GCCACC). "Sp30" es una secuencia aleatoria de 30 nucleótidos. La secuencia subrayada en la secuencia n.º 9 es la secuencia de 5'UTR de alfa-globina humana ("hAg") que tiene una longitud de 30 nucleótidos. Las secuencias 1 a 9 tal como se muestran en la figura 1 corresponden a SEQ ID NO: 37 a 45, respectivamente.

45 Figuras 2A y B:

Muestra que la "C" adicional en la "UTR mínima" es esencial (secuencia n.º 1 y n.º 2 en la figura 1). Se sembraron la línea celular epitelial alveolar humana (A549) y la línea celular de carcinoma hepatocelular humana (HepG2) a la

densidad de 20.000 células / pocillo y 40.000 células / pocillo, respectivamente, en una placa de 96 pocillos. 24 horas tras la siembra, se transfectaron células con diferentes constructos de ARN SNIM que codifican para luciferasa (secuencias 1 y 2 en la figura 1) usando Lipofectamine2000. Se midió la expresión de luciferasa a las 24 horas tras la transfección. Los valores representan la media  $\pm$  DE de 3 repeticiones y se representaron gráficamente frente a la dosis de transfección y se analizaron los datos mediante GraphPad Prism. Tanto en células A549 como HepG2, la delección de C dio como resultado una expresión inferior. Por tanto, se incluyó esta C adicional en el diseño de todos los constructos adicionales.

Figura 3:

Muestra el efecto de nucleótidos individuales tal como se indica y demuestra el efecto de la distancia entre la "C" adicional y el elemento de Kozak en células transfectadas A549. Las células se transfectaron y el ensayo de luciferasa se realizó tal como se describe en materiales y métodos. Dado que dosis superiores estaban fuera del intervalo lineal, en este caso sólo se presenta la respuesta a la dosis hasta 62,5 ng / pocillo. Se usó 5'UTR de alfa-globina humana como control positivo. Se realizaron experimentos de transfección con moléculas de ARN SNIM que albergaban las secuencias 3-8 de la figura 1, respectivamente. Se sembró la línea celular epitelial alveolar humana (A549) a la densidad de 20.000 células / pocillo en una placa de 96 pocillos. 24 horas tras la siembra, se transfectaron células con diferentes constructos de ARN SNIM que codifican para luciferasa (secuencias n.º 3-8 de la figura 1) usando Lipofectamine2000. Se midió la expresión de luciferasa a las 24 horas tras la transfección. Se representaron gráficamente los valores frente a la dosis de transfección y se analizaron los datos mediante GraphPad Prism. Los valores representan la media  $\pm$  DE de 3 repeticiones. En línea celular epitelial alveolar (A549), la inserción de una "A" adicional entre C y el elemento de Kozak (secuencia n.º 3 en la figura 1) dio como resultado una expresión significativamente inferior (figura 3). La inserción de una única "T" entre C y el elemento de Kozak (secuencia n.º 4 en la figura 1) dio como resultado niveles de expresión comparables al logrado con 5'UTR de alfa-globina humana que se usó como control positivo.

Figura 4:

Muestra el efecto de nucleótidos individuales tal como se indica y demuestra el efecto de la distancia entre la "C" adicional y el elemento de Kozak en células transfectadas HepG2. Las células se transfectaron y el ensayo de luciferasa se realizó tal como se describe en materiales y métodos. Dado que dosis superiores estaban fuera del intervalo lineal, en este caso sólo se presenta la respuesta a la dosis hasta 62,5 ng / pocillo. Se realizaron experimentos de transfección con las secuencias n.º 3-8 de la figura 1. Se sembró la línea celular de carcinoma hepatocelular (HepG2) a la densidad de 40.000 células / pocillo en una placa de 96 pocillos. 24 horas tras la siembra, se transfectaron células con diferentes constructos de ARN SNIM que codifican para luciferasa (secuencias n.º 3-8 de la figura 1) usando Lipofectamine2000. Se midió la expresión de luciferasa a las 24 horas tras la transfección. Se representaron gráficamente los valores frente a la dosis de transfección y se analizaron los datos mediante GraphPad Prism. Los valores representan la media  $\pm$  DE de 3 repeticiones. En ambas líneas celulares (células A549 (figura 3) y HepG2 (figura 4)), la inserción de una "A" adicional entre C y el elemento de Kozak (secuencia n.º 3 de la figura 1) dio como resultado una expresión significativamente inferior (figuras 3 y 4). En ambos tipos de células, la inserción de una única "T" entre C y el elemento de Kozak (secuencia n.º 4 de la figura 1) dio como resultado niveles de expresión comparables al logrado con 5'UTR de alfa-globina humana que se usó como control positivo. En células HepG2, la secuencia n.º 1 (figura 1) también fue igual de eficaz.

Figura 5:

Muestra el efecto del elemento de TISU sobre la expresión de luciferasa en células A549. Se realizó una respuesta a la dosis y ajuste de curva detalladas para constructos que codifican para luciferasa seleccionados. Basándose en datos anteriores de las figuras 2-4, se puso el elemento de TISU en la combinación de la secuencia 4 (figura 1) que contenía los dos atributos deseables: (C entre el promotor de T7 y el elemento de Kozak y T adicional entre C y el elemento de Kozak para lograr la secuencia n.º 9 de la figura 1).

Se sembraron la línea celular epitelial alveolar humana (A549) (figuras 5 A y B) y la línea celular de carcinoma hepatocelular humana (HepG2) (figuras 5 C y D) a la densidad de 20.000 células / pocillo y 40.000 células / pocillo, respectivamente, en una placa de 96 pocillos. 24 horas tras la siembra, se transfectaron células con diferentes constructos de ARN SNIM que codifican para luciferasa usando Lipofectamine2000. Se midió la expresión de luciferasa a las 24 y 48 horas tras la transfección (figura 5E). Se representaron gráficamente los valores frente a la dosis de transfección y se analizaron los datos mediante GraphPad Prism. Se indica la transfección de células A549 (A, B) y HepG2 (C, D) con diferentes ARNm que codifican para luciferasa. Se midió la actividad de luciferasa a las 24 (A, C) y 48 (B, D) horas tras la transfección. Los valores representan la media  $\pm$  DE de 3 repeticiones. En ambas líneas celulares y en ambos puntos de tiempo medidos, se obtuvo una expresión significativamente superior con el constructo de luciferasa que contenía el elemento de TISU (figuras 5A-D).

Figura 6:

Muestra el efecto del elemento de TISU sobre la expresión de luciferasa en células A549 (figura 6A) y en células HepG2 (figura 6B). Se sembraron la línea celular epitelial alveolar humana (A549) y la línea celular de carcinoma



hepatocelular humana (HepG2) a la densidad de 20.000 células / pocillo y 40.000 células / pocillo, respectivamente, en una placa de 96 pocillos. 24 horas tras la siembra, se transfectaron células con diferentes constructos de ARN SNIM que codifican para luciferasa usando Lipofectamine2000 (el eje de las X muestra la cantidad en ng de ARN SNIM por pocillo de una placa de 96 pocillos). Se midió la expresión de luciferasa a las 24 horas tras la transfección. Se representaron gráficamente los valores frente a la dosis de transfección y se analizaron los datos mediante GraphPad Prism. Se indica la transfección de células A549 (A) y HepG2 (B) con diferentes ARNm que codifican para luciferasa.

Figura 7:

Muestra los resultados de los experimentos *in vivo* en ratones con diferentes constructos de ARNm que codifican para luciferasa. Los constructos de luciferasa tal como se indican en la figura 7 (para el elemento de secuencia de UTR respectivo véase la figura 1) se sometieron a prueba *in vivo* en ratones Balb/c (hembra, 6-8 semanas). Para este conjunto de experimentos, también se sometió a prueba un elemento de UTR adicional que se ha mostrado que potencia la expresión transgénica (número de publicación internacional WO 2012/170930 A1) para determinar su eficiencia. El constructo de luciferasa que contiene este elemento de UTR se ha denominado Luc2-SUSA. Se complejaron 20 µg del ARN SNIM respectivo con LF-44 y se inyectaron por vía intravenosa en ratones Balb/c. Se realizó la obtención de imágenes *in vivo* a las 6 horas tras la inyección empleando un sistema de obtención de imágenes IVIS y se representaron gráficamente valores cuantificados como fotones/s/cm<sup>2</sup>/sr. En la figura 7A se muestran resultados de obtención de imágenes de animales completos y los resultados de obtención de imágenes del órgano completo se muestran en las figuras 7B (hígado), 7C (pulmón), 7D (bazo), respectivamente.

Se congelaron órganos extirpados a partir de los animales en nitrógeno líquido y se homogeneizaron. Se sometieron células a lisis en tampón de lisis Tris-HCl y se midió la actividad de luciferasa. Los resultados se muestran en las figuras 7E (hígado), 7F (pulmón), 7G (bazo), respectivamente.

La inserción del elemento de TISU dio como resultado una expresión superior en comparación con las 5' y 3'UTR anteriormente publicadas (número de publicación internacional WO 2012/170930 A1). La adición de una única T entre C y Kozak (secuencia n.º 4 de la figura 1) conduce a niveles de expresión comparables observados con UTR de alfa-globina humana (secuencia n.º 8 de la figura 1). La adición de un elemento de TISU en la secuencia n.º 4 (figura 1) aumentó adicionalmente la expresión (secuencia n.º 9 de la figura 1). Se encontró sorprendentemente que el efecto de UTR de alfa-globina humana no se encontró que fuera específico de secuencia. Una secuencia 30 de nucleótidos aleatoria soportó un nivel de expresión similar a la 5'UTR de alfa-globina humana.

Basándose en los resultados *in vitro* en líneas celulares y experimentos *in vivo* en ratones, se proponen las secuencias n.º 1, 4, 7 y 9 (figura 1) como candidatos prometedores para secuencias que albergan "UTR mínimas" para terapia de transcrito. Estas secuencias de UTR mínimas no tienen ningún efecto negativo sobre el rendimiento de ARN durante la transcripción *in vitro* y el ARNm resultante se traduce de manera mucho más eficiente en comparación con los ARNm que contienen UTR del estado de la técnica.

Figura 8:

Muestra los valores de recuento de glóbulos blancos (WBC) (figura 8A), glóbulos rojos (RBC) (figura 8B), plaquetas (figura 8C), hemoglobina (figura 8D) y hematocrito (figura 8E) a partir de ratones con diferentes constructos de ARNm que codifican para luciferasa. El experimento se realizó esencialmente tal como se describió en la figura 7 y los parámetros sanguíneos se analizaron empleando un analizador de hematología automatizado Sysmex KX-21N™ (IL, EE.UU.).

Figura 9:

Muestra experimentos de expresión con elemento de TISU que contiene ARNm que codifica para EPO humana en comparación con el de ARNm que codifica para EPO humana que contiene 5' y 3'UTR de (número de publicación internacional WO 2012/170930 A1: figuras 1 y 2) (UTR de SUSA) que se sabe que soporta una expresión de EPO muy alta. Se sembraron la línea celular epitelial alveolar humana (A549) y la línea celular de carcinoma hepatocelular humana (HepG2) a la densidad de 20.000 células / pocillo y 40.000 células / pocillo, respectivamente, en una placa de 96 pocillos. 24 horas tras la siembra, se transfectaron células con 250 ng de diferentes constructos de ARN SNIM que codifican para EPO usando Lipofectamine2000. Se cuantificaron las cantidades de EPO a las 24 horas tras la transfección mediante ELISA (kit de ELISA Quantikine IVD de eritropoyetina humana de R&D Systems (MN, EE.UU.)) y se analizaron los datos mediante GraphPad Prism. Los valores representan la media ± DE de 3 repeticiones.

Figura 10:

Muestra experimentos de expresión con OTC humana. Para OTC humana, se comparó la expresión a partir del elemento de TISU que contiene ARNm que codifica para hOTC con la de ARNm que codifica para hOTC que contiene 5' UTR de alfa-globina humana que se sabe que produce la expresión más alta en comparación con todas las demás combinaciones conocidas hasta ahora.

Se sembró la línea celular de carcinoma hepatocelular humana (HepG2) en placas de 96 pocillos y, 24 horas tras la siembra, se transfectaron células con diferentes constructos de ARN SNIM que codifican para hOTC usando Lipofectamine2000. 24 h tras la transfección, se sometieron las células a lisis y se cuantificaron las cantidades de OTC usando inmunotransferencia de tipo Western.

- 5 Tanto hAg como elemento de TISU que contenía ARN SNIM que codifican para hOTC dieron como resultado un nivel similar de expresión de hOTC (figura 10A). Se usó vinculina como mantenimiento y se cuantificaron las intensidades de bandas y se usaron como patrón de cuantificación interno (figura 10B).

Figura 11:

- 10 Estructuras secundarias predichas de un espaciador de 30 nucleótido de longitud aleatorio presente en la secuencia 7 (izquierda) y 5'UTR de alfa-globina humana presente en la secuencia 8 (derecha).

Otros aspectos y ventajas de la invención se describirán en los siguientes ejemplos, que se facilitan con fines de ilustración y no a modo de limitación.

## Ejemplos

### I. Materiales y métodos

#### 15 Vectores de plásmido

- Se sintetizaron las secuencias de 5'UTR respectivas junto con una secuencia de luciferasa de codones optimizados mediante GeneScriptG (NJ, EE.UU.) y se clonaron en pUC57-Kan (GeneScript). En el caso de EPO (eritropoyetina humana de codones optimizados) y OTC (ornitina transcarbamilasa humana de codones optimizados), se substituyó el gen de luciferasa de secuencia codificante por la secuencia codificante del gen de EPO (SEQ ID NO: 35) y de OTC (SEQ ID NO: 36), respectivamente. Las secuencias de UTR usada en los constructos junto con el nombre del constructo indicador de luciferasa respectivo se muestran en la figura 1.

#### Producción de ARNm

- 25 Para generar ARNm transcrito *in vitro* (ARNm IVT), se linealizaron plásmidos mediante digestión con BstBI y se purificaron mediante extracción con cloroformo y precipitación en etanol. Se usaron plásmidos lineales purificados como molde para la transcripción *in vitro* usando sistema de producción de ARN a gran escala RiboMax de T7 (Promega, Alemania). Se añadió análogo de caperuza anti-inverso (ARCA) a la mezcla de reacción para generar ARNm con caperuza en 5' y se poliadeniló el ARNm (Thermo Scientific) para generar la cola de poli-A en 3'.

- 30 Adicionalmente, para la producción de ARNm SNIM, se añadieron nucleótidos químicamente modificados, concretamente metil-CTP y tio-UTP (Jena Bioscience, Alemania), hasta una concentración final de ATP:CTP:UTP:metil-CTP:tio-UTP:GTP de 7,57 mM:5,68 mM:5,68 mM:1,89 mM:1,89 mM:1,21 mM. Se incubó la mezcla de IVT completa a 37°C durante 2 horas seguido por una digestión de ADN con ADNasa I durante 20 minutos a 37°C. Se precipitó ARN con acetato de amonio (concentración final de 2,5 M) y se lavó con EtOH al 70 %. La etapa de lavado se realizó dos veces. Finalmente, volvió a suspenderse el sedimento de ARN en agua libre de ARNasa. Se verificaron todos los ARNm en geles de agarosa al 1 %. Los ARN transcritos están químicamente modificados ya que aproximadamente el 25 % de los residuos de uridina son 2-tiouridina (s2U) y aproximadamente el 25 % de los residuos de citidina son 5-metilcitidina (m5C). Las secuencias de las UTR se facilitan en la figura 1.

#### Transfección *in vitro*

- 40 Se sembraron la línea celular epitelial alveolar humana (A549) y la línea celular de carcinoma hepatocelular humana (HepG2) a la densidad de 20.000 células / pocillo y 40.000 células / pocillo, respectivamente, en una placa de 96 pocillos. 24 horas tras la siembra, se transfectaron células con diferentes constructos de ARN SNIM que codifican para luciferasa usando el reactivo de transfección comercial LipofectamineTM2000 a una razón de 2,5 µl de LipofectamineTM2000 por 1 µg de ARNm (el eje de las X en las figuras 2-6 muestra la cantidad en ng de ARN SNIM por pocillo de una placa de 96 pocillos). Se preparó la formación de complejo de la siguiente manera: se diluyeron por separado LipofectamineTM2000 y ARNm en medio de transfección OptiMEM para alcanzar un volumen total de 45 µl, cada uno. Se incubaron estas mezclas a temperatura ambiente durante 5 minutos. Después se mezcló la disolución de LipofectamineTM2000 con la disolución de ARNm, seguido por otros 20 minutos de incubación a temperatura ambiente. Se incubaron las células en un volumen de transfección total de 90 µl a 37°C (nivel de CO2 del 5 %) durante una hora. Después de eso se retiró el medio de transfección y se lavaron las células con PBS. Posteriormente, volvieron a incubarse las células con medio L-15 de Leibovitz que contenía FBS al 10 %.

#### 50 Cultivo celular

Se hizo crecer una línea celular de adenocarcinoma alveolar humana (A549, ATCC CCL-185) en medio F12K de Ham complementado con FBS al 10 %. Se cultivó una línea celular de carcinoma hepatocelular humana (HepG2, ATCC HB-8065) en medio DMEM, complementado con suero bovino fetal al 10 %. Se hicieron crecer todas las

líneas celulares en una atmósfera humidificada a un nivel de CO<sub>2</sub> del 5 %.

#### Medición de bioluminiscencia

5 La luciferasa de luciérnaga (FFL) es una proteína indicadora común que no está presente de manera endógena en mamíferos y puede detectarse fácilmente mediante obtención de imágenes luminiscentes. La luciferasa cataliza la reacción de luciferina y oxígeno que da como resultado la emisión de bioluminiscencia.

10 Se sembraron la línea celular epitelial alveolar humana (A549) y la línea celular de carcinoma hepatocelular humana (HepG2) a la densidad de 20.000 células / pocillo y 40.000 células / pocillo, respectivamente, en una placa de 96 pocillos. 24 horas tras la siembra, se transfectoron células con diferentes constructos de ARN SNIM que codifican para luciferasa usando Lipofectamine2000 (el eje de las X muestra la cantidad en ng de ARN SNIM por pocillo de una placa de 96 pocillos). Se midió la bioluminiscencia a las 24 horas tras la transfección. Se representaron gráficamente los valores frente a la dosis de transfección y se analizaron los datos mediante GraphPad Prism.

Para cuantificar la expresión de luciferasa en lisado tisular homogeneizado, se extirparon órganos a partir de los animales, se congelaron en nitrógeno líquido, se homogeneizaron y se sometieron las células a lisis en tampón de lisis (Tris-HCl 25 mM, pH 7,5 con Triton-X100 al 0,1 %).

#### 15 Animales

20 Se obtuvieron ratones BALB/c hembra de seis a ocho semanas de edad a partir de Janvier, Route Des Chênes SecsBP5, F-53940 Le Genest St. Isle, Francia, y se mantuvieron en condiciones libres de patógenos específicos. Se aclimataron los ratones al entorno de la instalación para animales durante al menos siete días antes de los experimentos. Todos los procedimientos con animales se aprobaron y controlaron por el comité ético local y se llevaron a cabo según las directrices de la legislación alemana de protección de la vida animal.

#### Formulaciones de lipídeos

25 Se formularon lipídeos con ARNm de la siguiente manera: se disolvieron C12-(2-3-2), DOPE, Chol y DSPE-PEG2k (razón en peso de 3,6:0,18:0,76:1) en etanol y se inyectaron rápidamente en una disolución tamponada con citrato (ácido cítrico 10 mM, NaCl 150 mM, pH=4,5) que comprendía ARNm químicamente modificado que codificaba para luciferasa de luciérnaga a una razón en peso de lípido/ARNm de 10,5 para producir una concentración en etanol final del 20 % y se dializó frente a agua. Los complejos de lipídeo/ARNm resultantes dieron como resultado nanopartículas cargadas positivamente ( $92,6 \pm 0,7$  nm;  $21,0 \pm 0,2$  mV) y se inyectaron por vía intravenosa en la vena de la cola de ratones sujetos. En un segundo experimento, se ajustaron los complejos de lipídeo/ARNm a PBS antes de la inyección intravenosa, lo cual dio como resultado nanopartículas casi sin carga ( $91,5 \pm 0,6$  nm;  $-0,7 \pm 0,2$  mV).

#### Medición de la actividad de Luc en ratones usando obtención de imágenes bioluminiscentes *in vivo*

35 Veinticuatro horas tras la administración, se anestesiaron los ratones mediante inyección intraperitoneal de medetomidina (11,5 µg/kg de PC), midazolam (115 µg/kg de PC) y fentanilo (1,15 µg/kg de PC). Se aplicó sustrato de D-luciferina (3 mg/100 µl de PBS por ratón) mediante inyección intravenosa. Se midió la bioluminiscencia 10 minutos después, usando un sistema de obtención de imágenes IVIS 100 (Xenogen, Alameda, EE.UU.) y los ajustes de cámara: grupo (HS), campo de visión 10, f1 de apertura, agrupación de alta resolución y tiempo de exposición de 5 min. Se cuantificó la señal y se analizó usando el software Living Image versión 2.50 (Xenogen, Alameda, EE.UU.).

#### Análisis mediante inmunotransferencia de tipo Western de la proteína OTC

40 Se descongelaron placas congeladas y se realizó la lisis celular directa en la placa. Se sometieron proteínas a lisis usando tampón de lisis (TRIS 25 mM, Triton-X 100 al 0,1 %, Sigma-Aldrich, Alemania) complementado con inhibidor de proteasa (cOmplete, libre de EDTA, Roche Diagnostics, Alemania) y ADNasa (disolución de ADNasa I (2500 U/ml), (Thermo Fisher, EE.UU.). Después de la lisis se mezclaron las muestras con tampón de muestra NuPage® LDS y agente de reducción de muestra (Thermo Fisher, EE.UU.) y se calentaron durante 10 min a 70°C.

45 Se llevó a cabo la electroforesis en gel usando 15 µl del lisado en geles Midi de Bis-Tris al 10 % de NuPAGE con el sistema XCell4 SureLock™ Midi, Bio-Rad Criterion™ (Thermo Fisher, EE.UU.). Se transfirieron las proteínas usando el sistema de transferencia TransBlot® Turbo™ (Biorad, Alemania) durante 30 min. Después de la transferencia se bloquearon las membranas con NET-gelatina durante 30 min antes de incubar la membrana durante la noche a 4°C con el anticuerpo primario, diluido en NET-gelatina 1:2000 (anticuerpo policlonal frente a OTC (centro), AP6928c-AB Biocat, Alemania). Después de tres etapas de lavado con NET-gelatina, se añadió anticuerpo secundario conjugado con peroxidasa del rábano (HRP-IgG de cabra anti-conejo, sc-2004, Santa Cruz Biotechnology, EE.UU.), diluido 1:10.000 en NET-gelatina, durante 1 h a TA. Se lavó la membrana de nuevo tres veces con NET-gelatina hasta que se visualizaron las señales con un kit de sustrato quimioluminiscente (sustrato de HRP tipo Western Luminata Crescendo, Merck Millipore, Alemania) y se visualizaron usando el sistema ChemiDoc™ MP (Biorad, Alemania).

#### 55 Materiales

Se adquirieron FBS, medio L-15 de Leibovitz (Gibco), Lipofectamine™2000 y OptiMEM (Gibco) de Invitrogen, Alemania. Se preparó PBS estéril de manera interna. Se adquirieron F-12K de Ham, DMEM y tripsina-EDTA de c.c.pro GmbH, Alemania.

II. Resultados

5 II.a Experimentos de cultivo celular

Las figuras 2A y B muestran que la "C" adicional entre el promotor de T7 y el elemento de Kozak es esencial. La delección de esa base da como resultado la expresión reducida en ambos tipos de células comparados. Para ambos constructos (secuencias n.º 1 y 2 de la figura 1), se presentan el intervalo de dosis entero y el intervalo lineal (valores excluidos: dosis superior a 62,5 ng / pocillo excluida del análisis) por separado por conveniencia de comparación. Tanto en células A549 como HepG2, la delección de C dio como resultado una expresión inferior. Por tanto, se incluyó esta C adicional en el diseño de todos los constructos adicionales.

Basándose en los resultados obtenidos en células A549 y HepG2, se llevaron a cabo experimentos adicionales con el constructo que contenía la "C" adicional (secuencia número 1: T7Luc2).

Figura 3 y figura 4.

15 Se usó la secuencia 1 como molde y, a esta secuencia, se le incorporó o bien un único nucleótido (A, T, G o C: secuencias número 3 - 6 de la figura 1, respectivamente), o bien una secuencia aleatoria, de 30 nucleótidos de longitud y desprovista de cualquier estructura secundaria predecible (secuencia 7) o 5'UTR de alfa-globina humana (secuencia 8), entre la "C" investigada y el elemento de Kozak.

20 Se transfectaron las células y se realizó un ensayo de luciferasa tal como se describe en materiales y métodos. Dado que dosis superiores estaban fuera del intervalo lineal, en este caso sólo se presenta la respuesta a la dosis hasta 62,5 ng / pocillo. Se usó 5'UTR de alfa-globina humana como control positivo.

25 Para resumir los resultados anteriores, las figuras 1 a 4 muestran que una "C" adicional entre el promotor de T7 y el elemento de Kozak es esencial con respecto a lograr una alta expresión de proteína empleando una 5'UTR minimalista. La delección del nucleótido da como resultado una expresión reducida. La adición de una "A" adicional entre la "C" adicional y el elemento de Kozak afecta de manera negativa la expresión. Cuando se añade una base de pirimidina y, lo más preferiblemente, una "T" en esa posición, se obtienen niveles comparables a los observados con 5'UTR de hAg.

Posteriormente, se realizaron experimentos adicionales para:

- esclarecer el efecto de elemento de TISU cuando se combina con la secuencia que mejor funciona (secuencia 9), y
- 30 - determinar si el efecto de 5'UTR a partir de hAg es un efecto específico de secuencia o si es la distancia entre la caperuzas en 5' y el codón de iniciación lo que es importante.

35 La figura 5 muestra el efecto del elemento de TISU sobre la expresión de luciferasa en células A549. El "elemento de TISU" incorpora "AG" en vez de "CC" en la secuencia n.º 9 tal como se muestra en la figura 1 con respecto a la secuencia n.º 4 tal como se muestra en la figura 1. Las células A549 (figuras 5A y B) así como las células HepG2 (figuras 5C y D) mostraron una expresión significativamente superior de luciferasa con el constructo de luciferasa que contenía el elemento de TISU junto con la "C" de la secuencia n.º 1 y la "T" adicional entre esta "C" y el elemento de Kozak a las 24 (A, C) y 48 (B, D) horas tras la transfección.

40 La figura 6 muestra los resultados a partir del mismo experimento que el de la figura 5 pero con la adición de una 5'UTR que contenía una secuencia aleatoria de 30 nucleótidos, para permitir una comparación directa de la UTR de alfa-globina humana (secuencia 8 de la figura 1) con una secuencia aleatoria de la misma longitud (secuencia 7 de la figura 1). Se midió la expresión de luciferasa en células HepG2 (figura 6A) y A549 (figura 6B) 24 horas después de la transfección con los ARN SNIM tal como se indica.

45 La figura 9 muestra los resultados a partir de experimentos de expresión con el elemento de TISU que contiene ARNm que codifica para hEPO en comparación con el de ARNm que codifica para hEPO que contiene 5' y 3'UTR de (número de publicación internacional WO 2012/170930 A1: figuras 1 y 2) (UTR de SUSA) que se usó como patrón después de la transfección de células A549 y HepG2 con el ARN SNIM respectivo. Se cuantificaron las cantidades de EPO a las 24 horas tras la transfección mediante ELISA. Los valores representan la media ± DE de 3 repeticiones.

50 En células A549 humanas, la incorporación del elemento de TISU dio como resultado una expresión superior en comparación con la lograda con la incorporación de 5' y 3'UTR (figura 9A). Se observaron niveles de expresión comparables en células HepG2 (figura 9B). Esto es especialmente sorprendente ya que la incorporación de las 5' y 3'UTR de SUSA hace que los ARN sean aproximadamente 200 nucleótidos más largos en comparación con la UTR según la presente invención.

- La figura 10 muestra experimentos de expresión con OTC humana. Para comparación, se comparó el elemento de TISU que contenía ARNm que codifica para hOTC con el de ARNm que codifica para hOTC que contiene 5' UTR de alfa-globina humana que se sabe que produce la expresión más alta en comparación con todas las demás combinaciones conocidas hasta ahora. Se transfectaron células HepG2 con diferentes constructos de ARN SNIM que codifican para hOTC, se sometieron a lisis 24 horas después y se cuantificaron las cantidades de OTC mediante inmunotransferencia de tipo Western.
- Tanto hAg como elemento de TISU que contenía ARN SNIM que codifican para hOTC dieron como resultado un nivel similar de expresión de hOTC (figura 10A). Se usó vinculina como mantenimiento y se compararon las intensidades de bandas usando densitometría (figura 10B).
- 10 II.b Aplicación i.v. de constructos de Luc2 en ratones
- Los resultados se muestran en la figura 7 y la figura 8.
- Se han usado los siguientes constructos en aplicaciones i.v. en ratones:
- Luc2 (+8+A)
- Luc2 (+8+T)
- 15 Luc2 (+8+T) + TISU
- Luc2-hAg
- Luc2-Sp30
- Luc2-UTR de SUSA
- 20 Se complejaron 20 µg del ARN SNIM respectivo con LF-44 y se inyectaron i.v. en ratones Balb/c. Como control adicional, también se produjo la secuencia de Luc2 flanqueada por potenciador de CMV humano en el extremo 5' (Luc2-SUSA) y 3'UTR de hormona del crecimiento humana en el extremo 3'. Las secuencias usadas como UTR en este constructo se han tomado de la patente de Shire (documento WO 2012/170930 A1: ID de secuencia 1 / figura 1).
- 25 Se realizó la obtención de imágenes *in vivo* a las 6 horas tras la inyección empleando un sistema de obtención de imágenes IVIS y se representaron gráficamente valores cuantificados como fotones/s/cm<sup>2</sup>/sr. En la figura 7A se muestran resultados de obtención de imágenes de animales completos y los resultados de obtención de imágenes del órgano completo se muestran en las figuras 7B (hígado), 7C (pulmón), 7D (bazo), respectivamente.
- 30 Se congelaron órganos extirpados a partir de los animales en nitrógeno líquido, se homogeneizaron, se sometieron a lisis y se midió la actividad de luciferasa. Los resultados se muestran en las figuras 7E (hígado), 7F (pulmón), 7G (bazo), respectivamente.
- Se analizaron los parámetros sanguíneos de los animales empleando un analizador de hematología automatizado Sysmex KX-21N™: los valores de recuento de glóbulos blancos (WBC) (figura 8A), glóbulos rojos (RBC) (figura 8B), plaquetas (figura 8C), hemoglobina (figura 8D) y hematocrito (figura 8E) a partir de ratones con diferentes constructos de ARNm que codifican para luciferasa no muestran diferencias significativas.
- 35 Figura 11: estructuras secundarias predichas de un espaciador de 30 nucleótidos de longitud aleatorio presente en la secuencia 7 (izquierda) y 5'UTR de alfa-globina humana de la misma longitud presente en la secuencia 8 (derecha). Aunque las estructuras secundarias de ambas secuencias ni siquiera son similares, dieron como resultado niveles de expresión similares (figuras 6A y 6B) que fueron ambos igualmente bajos en comparación con T7Luc2(+8+T)-TISU.

**Lista de secuencias**

<110> ethris GmbH

<120> Secuencias de UTR mínimas novedosas

<130> Y2004 PCT S3

<150> Documento EP 16 16 3264.1

<151> 31-03-2016

<150> Documento EP 16 17 7094.6

<151> 30-06-2016

<160> 45

<170> BiSSAP 1.2

<210> 1  
 <211> 7  
 <212> ADN  
 <213> Secuencia artificial  
 <220>  
 <221> fuente  
 <222> 1..7  
 <223> /tipo\_mol = "ADN sin asignar"  
         /nota = "UTR1 mínima"  
         /organismo = "Secuencia artificial"

<400> 1

cgccacc

7

<210> 2  
 <211> 8  
 <212> ADN  
 <213> Secuencia artificial  
 <220>  
 <221> fuente  
 <222> 1..8  
 <223> /tipo\_mol = "ADN sin asignar"  
         /nota = "UTR2 mínima"  
         /organismo = "Secuencia artificial"

<220>

<221> variación

<222> 2

<223> /sustituir = "A"

        /sustituir = "T"

        /sustituir = "C"

        /sustituir = "G"

<400> 2

cngccacc

8

<210> 3  
 <211> 23  
 <212> ADN  
 <213> Secuencia artificial  
 <220>  
 <221> fuente

```

<222> 1..23
<223> /tipo_mol = "ADN sin asignar"
      /nota = "Promotor de T7"
      /organismo = "Secuencia artificial"
<400> 3
taatacgact cactataggg aga                                23
<210> 4
<211> 23
<212> ADN
<213> Secuencia artificial
<220>
<221> fuente
<222> 1..23
<223> /tipo_mol = "ADN sin asignar"
      /nota = "Promotor de T3"
      /organismo = "Secuencia artificial"
<400> 4
aattaaccct cactaaaggg aga                                23
<210> 5
<211> 21
<212> ADN
<213> Secuencia artificial
<220>
<221> fuente
<222> 1..21
<223> /tipo_mol = "ADN sin asignar"
      /nota = "Promotor de SP6"
      /organismo = "Secuencia artificial"
<400> 5
atttaggtga cactatagaa g                                  21
<210> 6
<211> 21
<212> ADN
<213> Secuencia artificial
<220>
<221> fuente
<222> 1..21

```



<223> /tipo\_mol = "ADN sin asignar"  
     /nota = "Promotor de K11"  
     /organismo = "Secuencia artificial"  
 <400> 6  
 aattagggca cactataggg a 21  
 <210> 7  
 <211> 6  
 <212> ADN  
 <213> Secuencia artificial  
 <220>  
 <221> fuente  
 <222> 1..6  
 <223> /tipo\_mol = "ADN sin asignar"  
     /nota = "Promotor de T7"  
     /organismo = "Secuencia artificial"  
 <400> 7  
 gggaga 6  
 <210> 8  
 <211> 6  
 <212> ADN  
 <213> Secuencia artificial  
 <220>  
 <221> fuente  
 <222> 1..6  
 <223> /tipo\_mol = "ADN sin asignar"  
     /nota = "Promotor de T3"  
     /organismo = "Secuencia artificial"  
 <400> 8  
 gggaga 6  
 <210> 9  
 <211> 4  
 <212> ADN  
 <213> Secuencia artificial  
 <220>  
 <221> fuente  
 <222> 1..4  
 <223> /tipo\_mol = "ADN sin asignar"

```

        /nota = "Promotor de SP6"
        /organismo = "Secuencia artificial"
<400> 9
gaag
4
<210> 10
<211> 4
<212> ADN
<213> Secuencia artificial
<220>
<221> fuente
<222> 1..4
<223> /tipo_mol = "ADN sin asignar"
        /nota = "Promotor de K11"
        /organismo = "Secuencia artificial"
<400> 10
ggga
4
<210> 11
<211> 16
<212> ARN
<213> Secuencia artificial
<220>
<221> fuente
<222> 1..16
<223> /tipo_mol = "ARN sin asignar"
        /nota = "Secuencia complementaria inversa de UTR1 mínima"
        /organismo = "Secuencia artificial"
<400> 11
caugguggcg ucuccc
16
<210> 12
<211> 17
<212> ARN
<213> Secuencia artificial
<220>
<221> fuente
<222> 1..17
<223> /tipo_mol = "ARN sin asignar"
        /nota = "Secuencia complementaria inversa de UTR2 mínima"

```

/organismo = "Secuencia artificial"  
 <220>  
 <221> variación  
 <222> 10  
 <223> /sustituir = "G"  
     /sustituir = "C"  
     /sustituir = "U"  
     /sustituir = "A"  
 <400> 12  
 caugguggcn gucuccc 17  
 <210> 13  
 <211> 16  
 <212> ARN  
 <213> Secuencia artificial  
 <220>  
 <221> fuente  
 <222> 1..16  
 <223> /tipo\_mol = "ARN sin asignar"  
     /nota = "Secuencia complementaria inversa de UTR1 mínima que incluye el elemento de TISU"  
     /organismo = "Secuencia artificial"  
 <400> 13  
 caucuuggcg ucuccc 16  
 <210> 14  
 <211> 17  
 <212> ARN  
 <213> Secuencia artificial  
 <220>  
 <221> fuente  
 <222> 1..17  
 <223> /tipo\_mol = "ARN sin asignar"  
     /nota = "Secuencia complementaria inversa de UTR2 mínima que incluye el elemento de TISU"  
     /organismo = "Secuencia artificial"  
 <220>  
 <221> variación  
 <222> 10  
 <223> /sustituir = "G"  
     /sustituir = "C"

```

        /sustituir = "A"
        /sustituir = "U"
<400> 14
caucuuggcn gucuccc                                17
<210> 15
<211> 16
<212> ARN
<213> Secuencia artificial
<220>
<221> fuente
<222> 1..16
<223> /tipo_mol = "ARN sin asignar"
        /nota = "Secuencia complementaria inversa de la UTR1 mínima alternativa"
        /organismo = "Secuencia artificial"
<400> 15
cauggcggcg ucucucc                                16
<210> 16
<211> 17
<212> ARN
<213> Secuencia artificial
<220>
<221> fuente
<222> 1..17
<223> /tipo_mol = "ARN sin asignar"
        /nota = "Secuencia complementaria inversa de la UTR2 mínima alternativa"
        /organismo = "Secuencia artificial"
<220>
<221> variación
<222> 10
<223> /sustituir = "U"
        /sustituir = "A"
        /sustituir = "G"
        /sustituir = "C"
<400> 16
cauggcggcn gucuccc                                17
<210> 17
<211> 16

```

<212> ARN  
 <213> Secuencia artificial  
 <220>  
 <221> fuente  
 <222> 1..16  
 <223> /tipo\_mol = "ARN sin asignar"  
     /nota = "Secuencia complementaria inversa de la UTR1 mínima alternativa que incluye el elemento de TISU"  
     /organismo = "Secuencia artificial"

<400> 17  
 caucucggcg ucuccc 16

<210> 18  
 <211> 17  
 <212> ARN  
 <213> Secuencia artificial  
 <220>  
 <221> fuente  
 <222> 1..17  
 <223> /tipo\_mol = "ARN sin asignar"  
     /nota = "Secuencia complementaria inversa de la UTR2 mínima alternativa que incluye el elemento de TISU"  
     /organismo = "Secuencia artificial"

<220>  
 <221> variación  
 <222> 10  
 <223> /sustituir = "U"  
     /sustituir = "G"  
     /sustituir = "C"  
     /sustituir = "A"

<400> 18  
 caucucggcn guccccc 17

<210> 19  
 <211> 14  
 <212> ARN  
 <213> Secuencia artificial  
 <220>  
 <221> fuente  
 <222> 1..14  
 <223> /tipo\_mol = "ARN sin asignar"

```

        /nota = "Secuencia complementaria inversa de UTR1 mínima"
        /organismo = "Secuencia artificial"
<400> 19
caugguggcg uccc                                     14
<210> 20
<211> 15
<212> ARN
<213> Secuencia artificial
<220>
<221> fuente
<222> 1..15
<223> /tipo_mol = "ARN sin asignar"
        /nota = "Secuencia complementaria inversa de UTR2 mínima"
        /organismo = "Secuencia artificial"
<220>
<221> variación
<222> 10
<223> /sustituir = "A"
        /sustituir = "C"
        /sustituir = "G"
        /sustituir = "U"
<400> 20
caugguggcn guccc                                     15
<210> 21
<211> 14
<212> ARN
<213> Secuencia artificial
<220>
<221> fuente
<222> 1..14
<223> /tipo_mol = "ARN sin asignar"
        /nota = "Secuencia complementaria inversa de UTR1 mínima que incluye el elemento de TISU"
        /organismo = "Secuencia artificial"
<400> 21
caucuuggcg uccc                                     14
<210> 22
<211> 15

```

<212> ARN  
 <213> Secuencia artificial  
 <220>  
 <221> fuente  
 <222> 1..15  
 <223> /tipo\_mol = "ARN sin asignar"  
       /nota = "Secuencia complementaria inversa de UTR2 mínima que incluye el elemento de TISU"  
       /organismo = "Secuencia artificial"  
 <220>  
 <221> variación  
 <222> 10  
 <223> /sustituir = "A"  
       /sustituir = "C"  
       /sustituir = "G"  
       /sustituir = "U"  
 <400> 22  
 caucuuggcn guccc 15  
 <210> 23  
 <211> 14  
 <212> ARN  
 <213> Secuencia artificial  
 <220>  
 <221> fuente  
 <222> 1..14  
 <223> /tipo\_mol = "ARN sin asignar"  
       /nota = "Secuencia complementaria inversa de la UTR1 mínima alternativa"  
       /organismo = "Secuencia artificial"  
 <400> 23  
 cauggcggcg uccc 14  
 <210> 24  
 <211> 15  
 <212> ARN  
 <213> Secuencia artificial  
 <220>  
 <221> fuente  
 <222> 1..15  
 <223> /tipo\_mol = "ARN sin asignar"

/nota = "Secuencia complementaria inversa de la UTR2 mínima alternativa"  
 /organismo = "Secuencia artificial"  
 <220>  
 <221> variación  
 <222> 10  
 <223> /sustituir = "A"  
         /sustituir = "C"  
         /sustituir = "G"  
         /sustituir = "U"  
 <400> 24  
 cauggcggcn guccc 15  
 <210> 25  
 <211> 14  
 <212> ARN  
 <213> Secuencia artificial  
 <220>  
 <221> fuente  
 <222> 1..14  
 <223> /tipo\_mol = "ARN sin asignar"  
         /nota = "Secuencia complementaria inversa de la UTR1 mínima alternativa que incluye el elemento de TISU"  
         /organismo = "Secuencia artificial"  
 <400> 25  
 caucucggcg uccc 14  
 <210> 26  
 <211> 15  
 <212> ARN  
 <213> Secuencia artificial  
 <220>  
 <221> fuente  
 <222> 1..15  
 <223> /tipo\_mol = "ARN sin asignar"  
         /nota = "Secuencia complementaria inversa de la UTR2 mínima alternativa que incluye el elemento de TISU"  
         /organismo = "Secuencia artificial"  
 <220>  
 <221> variación  
 <222> 10  
 <223> /sustituir = "A"



```

/sustituir = "C"
/sustituir = "G"
/sustituir = "U"
<400> 26
caucucggcn guccc                                     15
<210> 27
<211> 14
<212> ARN
<213> Secuencia artificial
<220>
<221> fuente
<222> 1..14
<223> /tipo_mol = "ARN sin asignar"
        /nota = "Secuencia complementaria inversa de UTR1 mínima"
        /organismo = "Secuencia artificial"
<400> 27
caugguggcg cuuc                                     14
<210> 28
<211> 15
<212> ARN
<213> Secuencia artificial
<220>
<221> fuente
<222> 1..15
<223> /tipo_mol = "ARN sin asignar"
        /nota = "Secuencia complementaria inversa de UTR2 mínima"
        /organismo = "Secuencia artificial"
<220>
<221> variación
<222> 10
<223> /sustituir = "A"
        /sustituir = "C"
        /sustituir = "G"
        /sustituir = "U"
<400> 28
caugguggcn gcuuc                                     15
<210> 29

```

<211> 14  
 <212> ARN  
 <213> Secuencia artificial  
 <220>  
 <221> fuente  
 <222> 1..14  
 <223> /tipo\_mol = "ARN sin asignar"  
     /nota = "Secuencia complementaria inversa de UTR1 mínima que incluye el elemento de TISU"  
     /organismo = "Secuencia artificial"  
 <400> 29  
 caucuuggcg cuuc 14  
 <210> 30  
 <211> 15  
 <212> ARN  
 <213> Secuencia artificial  
 <220>  
 <221> fuente  
 <222> 1..15  
 <223> /tipo\_mol = "ARN sin asignar"  
     /nota = "Secuencia complementaria inversa de UTR2 mínima que incluye el elemento de TISU"  
     /organismo = "Secuencia artificial"  
 <220>  
 <221> variación  
 <222> 10  
 <223> /sustituir = "A"  
     /sustituir = "C"  
     /sustituir = "G"  
     /sustituir = "U"  
 <400> 30  
 caucuuggcn gcuuc 15  
 <210> 31  
 <211> 14  
 <212> ARN  
 <213> Secuencia artificial  
 <220>  
 <221> fuente  
 <222> 1..14

<223> /tipo\_mol = "ARN sin asignar"  
 /nota = "Secuencia complementaria inversa de la UTR1 mínima alternativa"  
 /organismo = "Secuencia artificial"

<400> 31  
 cauggcggcg cuuc 14

<210> 32  
 <211> 15  
 <212> ARN  
 <213> Secuencia artificial  
 <220>  
 <221> fuente  
 <222> 1..15

<223> /tipo\_mol = "ARN sin asignar"  
 /nota = "Secuencia complementaria inversa de la UTR2 mínima alternativa"  
 /organismo = "Secuencia artificial"

<220>  
 <221> variación  
 <222> 10  
 <223> /sustituir = "A"  
 /sustituir = "C"  
 /sustituir = "G"  
 /sustituir = "U"

<400> 32  
 cauggcggcn gcuuc 15

<210> 33  
 <211> 14  
 <212> ARN  
 <213> Secuencia artificial  
 <220>  
 <221> fuente  
 <222> 1..14

<223> /tipo\_mol = "ARN sin asignar"  
 /nota = "Secuencia complementaria inversa de la UTR1 mínima alternativa que incluye el elemento de TISU"  
 /organismo = "Secuencia artificial"

<400> 33  
 caucucggcg cuuc 14

<210> 34

<211> 15

<212> ARN

<213> Secuencia artificial

<220>

<221> fuente

<222> 1..15

<223> /tipo\_mol = "ARN sin asignar"

/nota = "Secuencia complementaria inversa de la UTR2 mínima alternativa que incluye el elemento de TISU"

/organismo = "Secuencia artificial"

<220>

<221> variación

<222> 10

<223> /sustituir = "A"

/sustituir = "C"

/sustituir = "G"

/sustituir = "U"

<400> 34

caucucggcn gcuuc

15

<210> 35

<211> 637

<212> ADN

<213> Secuencia artificial

<220>

<221> fuente

<222> 1..637

<223> /tipo\_mol = "ADN sin asignar"

/nota = "T7-TISU-hEPO (de codones optimizados)"

/organismo = "Secuencia artificial"

<400> 35

ES 2 835 051 T3

gtgactagat cttaatacga ctactatag ggagactgcc aagatgggcg tgcacgaatg 60  
tcttcttggt ctgtggctgc tgctgagcct gctgtctctg cctctgggac tgctgtgct 120  
gggagcccct cctagactga tctgagacag ccgggtgctg gaaagatacc tgctggaagc 180  
caaagaggcc gagaacatca ccaccggctg cgcagagcac tgcagcctga acgagaatat 240  
cacctgccc gacaccaaag tgaacttcta cgcctggaag cggatggaag tgggccagca 300  
ggctgtgga gtgtggcagg gactggccct gctgagcgaa gctgtgctga gaggacaggc 360  
tctgctctg aacagcagcc agccttggga gcctctgcag ctgcacgtgg acaaggcct 420  
gtctggcctg agaagcctga ccactctgct gagagccctg ggggccaga aagaggccat 480  
ctctccacct gatgcgcct ctgcccctc tetgagaacc atcaccgccc acaccttcag 540  
aaagctgttc cgggtgtaca gcaacttct cgggggcaag ctgaagctgt acacaggcga 600  
ggcctgccc accggcgata gataattoga agtgact 637

<210> 36

<211> 1120

<212> ADN

<213> Secuencia artificial

<220>

<221> fuente

<222> 1..1120

<223> /tipo\_mol = "ADN sin asignar"

/nota = "T7-TISU-hOTC (de codones optimizados)"

/organismo = "Secuencia artificial"

<400> 36

gtgactgcta gctaatacga ctactatag ggagactgcc aagatgctgt tcaacctgcg 60  
gatcctgctg aacaacgccc ccttcggaa cggccacaac ttcattggtgc gcaacttcag 120  
atgcccagc cccctgcaga acaagggtca gctgaagggc agggacctgc tgacctgaa 180  
gaacttcacc ggogaagaga tcaagtacat gctgtggctg agcgcgacc tgaagttccg 240  
gatcaagcag aagggcgagt acctgcccct gctgcagggc aagtctctgg gcatgatctt 300  
cgagaagcgg agcaccgga cccggctgct taccgagaca ggatttgccc tgctgggccc 360  
ccaccttgc tttctgacca ccaggatat ccacctgggc gtgaacgaga gcctgaccga 420  
cacagccaga gtgctgagca gcatggccga tgccgtgctg gccagagtgt acaagcagag 480  
cgacctggac acctggcca aagaggccag catccccatc atcaacggcc tgtccgacct 540  
gtaccacccc atccagatcc tggccgacta cctgacctg caggaacct acagctcct 600  
gaagggcctg aactgagct ggatcggcga oggcaacaac atcctgact ctatcatgat 660

ES 2 835 051 T3

gagcgccgcc aagttcggca tgcattctgca ggccgccacc cccaagggct atgagcctga 720  
 tgccagcgtg accaagetgg cagagcagta cgccaaagag aacggcacca agctgctget 780  
 gaccaacgac cctctggaag ccgcccacgg cggcaatgtg ctgatcaccg atacctggat 840  
 cagcatgggc caggaagagg aaaagaagaa gcggctgcag gccttccagg gctaccaagt 900  
 gacatgaag accgccaag tggccgccag cgactggacc ttctgcaact gcctgccag 960  
 aaagcccgaag gaggtggagc acgaggtggt ctacagcccc cggtccttgg tgtttcccga 1020  
 ggccgagaac cggaagtgga ccatcatggc tgtgatgggt tctctgctga ccgactactc 1080  
 ccccagctg cagaagccca agttotgaag cgetgtgact 1120

<210> 37

<211> 20

<212> ADN

<213> Secuencia artificial

<220>

<221> fuente

<222> 1..20

<223> /tipo\_mol = "ADN sin asignar"

/nota = "T7Luc2"

/organismo = "Secuencia artificial"

<400> 37

tatagggaga cgccaccatg 20

<210> 38

<211> 19

<212> ADN

<213> Secuencia artificial

<220>

<221> fuente

<222> 1..19

<223> /tipo\_mol = "ADN sin asignar"

/nota = "T7Luc2 ( $\Delta$ C)"

/organismo = "Secuencia artificial"

<400> 38

tatagggaga gccaccatg 19

<210> 39

<211> 21

<212> ADN

<213> Secuencia artificial

<220>  
 <221> fuente  
 <222> 1..21  
 <223> /tipo\_mol = "ADN sin asignar"  
     /nota = "T7Luc2 (+8+A)"  
     /organismo = "Secuencia artificial"  
 <400> 39  
 tataggaga cagccacat g 21  
 <210> 40  
 <211> 21  
 <212> ADN  
 <213> Secuencia artificial  
 <220>  
 <221> fuente  
 <222> 1..21  
 <223> /tipo\_mol = "ADN sin asignar"  
     /nota = "T7Luc2 (+8+T)"  
     /organismo = "Secuencia artificial"  
 <400> 40  
 tataggaga ctgccacat g 21  
 <210> 41  
 <211> 21  
 <212> ADN  
 <213> Secuencia artificial  
 <220>  
 <221> fuente  
 <222> 1..21  
 <223> /tipo\_mol = "ADN sin asignar"  
     /nota = "T7Luc2 (+8+G)"  
     /organismo = "Secuencia artificial"  
 <400> 41  
 tataggaga cggccacat g 21  
 <210> 42  
 <211> 21  
 <212> ADN  
 <213> Secuencia artificial  
 <220>

<221> fuente  
 <222> 1..21  
 <223> /tipo\_mol = "ADN sin asignar"  
       /nota = "T7Luc2 (+8+C)"  
       /organismo = "Secuencia artificial"  
 <400> 42  
 tatagggaga ccgccaccat g 21  
 <210> 43  
 <211> 21  
 <212> ADN  
 <213> Secuencia artificial  
 <220>  
 <221> fuente  
 <222> 1..21  
 <223> /tipo\_mol = "ADN sin asignar"  
       /nota = "T7Luc2 (+8+Sp30)"  
       /organismo = "Secuencia artificial"  
 <220>  
 <221> variación  
 <222> 12  
 <223> /sustituir = "secuencia de una secuencia de 30 nucleótidos aleatoria"  
 <400> 43  
 tatagggaga cngccaccat g 21  
 <210> 44  
 <211> 50  
 <212> ADN  
 <213> Secuencia artificial  
 <220>  
 <221> fuente  
 <222> 1..50  
 <223> /tipo\_mol = "ADN sin asignar"  
       /nota = "T7Luc2 (+8+hAg)"  
       /organismo = "Secuencia artificial"  
 <400> 44  
 tatagggaga ctctctggt cccacagac tcagagagaa cgccacatg 50  
 <210> 45  
 <211> 21



<212> ADN

<213> Secuencia artificial

<220>

<221> fuente

<222> 1..21

<223> /tipo\_mol = "ADN sin asignar"

    /nota = "T7Luc2 (+8+T)+TISU"

    /organismo = "Secuencia artificial"

<400> 45

tatagggaga ctgccaagat g

21

## REIVINDICACIONES

1. Una molécula de ADN, que puede transcribirse para dar un ARNm, que comprende una cadena con los siguientes elementos:
- 5 (a) una región codificante, que incluye un codón de iniciación en su extremo 5', que codifica para un polipéptido; y
- (b) directamente en el sentido de 5' de dicha secuencia codificante, la secuencia:  
 $R_1$ -CNGCCACC (SEQ ID NO: 2), en la que el nucleótido N en la posición 2 de SEQ ID NO: 2 es T;  
 en la que  $R_1$  es un promotor que se reconoce por una ARN polimerasa dependiente de ADN;  
 o que comprende la cadena complementaria,
- 10 en la que el promotor que se reconoce por una ARN polimerasa dependiente de ADN se selecciona del grupo que consiste en:
- (i) TAATACGACTCACTATAGGGAGA (SEQ ID NO: 3) que se reconoce por una ARN polimerasa dependiente de ADN de T7;
- 15 (ii) AATTAACCCTCACTAAAGGGAGA (SEQ ID NO: 4) que se reconoce por una ARN polimerasa dependiente de ADN de T3;
- (iii) ATTTAGGTGACACTATAGAAG (SEQ ID NO: 5) que se reconoce por una ARN polimerasa dependiente de ADN de SP6; y
- (iv) AATTAGGGCACACTATAGGGA (SEQ ID NO: 6) que se reconoce por una ARN polimerasa dependiente de ADN de K11.
- 20 2. Un vector que comprende la molécula de ADN según la reivindicación 1.
3. Una célula huésped que comprende el vector según la reivindicación 2.
4. Una composición que comprende:  
 la molécula de ADN según la reivindicación 1, el vector según la reivindicación 2 o la célula huésped según la reivindicación 3.
- 25 5. Una molécula de ARN que comprende
- (a) una región codificante, que incluye un codón de iniciación en su extremo 5', que codifica para un polipéptido; y
- (b) directamente en el sentido de 5' de dicha secuencia codificante, una UTR de la secuencia  $R_2$ -CNGCCACC (SEQ ID NO: 2), en la que el nucleótido N en la posición 2 de SEQ ID NO: 2 es U,
- 30 en la que  $R_2$  es una secuencia de ARN correspondiente a la parte de una región de promotor que empieza con el nucleótido en el que una ARN polimerasa dependiente de ADN inicia la síntesis de ARN,  
 en la que  $R_2$  se selecciona del grupo que consiste en:
- (i) GGGAGA (SEQ ID NO: 7);
- 35 (ii) GGGAGA (SEQ ID NO: 8);
- (iii) GAAG (SEQ ID NO: 9); y
- (iv) GGGGA (SEQ ID NO: 10); y
- en la que la molécula de ARN comprende una cola de poli-A en el extremo 3';
- 40 en la que la UTR tal como se define en (b) tiene una longitud máxima de 14 nucleótidos cuando  $R_2$  es (i) o (ii); o
- en la que la UTR tal como se define en (b) tiene una longitud máxima de 12 nucleótidos cuando  $R_2$  es (iii) o (iv).
6. La molécula de ARN según la reivindicación 5, en la que la cola de poli-A tiene una longitud de al menos

120 nucleótidos.

7. Una molécula de ácido nucleico que codifica para la molécula de ARN según la reivindicación 5 o 6.
8. Un vector que comprende la molécula de ácido nucleico según la reivindicación 7.
9. Una célula huésped que comprende el vector según la reivindicación 8.
- 5 10. Una composición farmacéutica que comprende la molécula de ARN según la reivindicación 5 o 6, la molécula de ácido nucleico según la reivindicación 7, el vector según la reivindicación 8 o la célula huésped según la reivindicación 9 y opcionalmente un portador farmacéuticamente aceptable.
11. La composición farmacéutica según la reivindicación 10, para su uso en terapias basadas en ARN.
- 10 12. Un kit que comprende la molécula de ADN según la reivindicación 1, la molécula de ARN según la reivindicación 5 o 6, la molécula de ácido nucleico según la reivindicación 7, el vector según la reivindicación 2 u 8 o la célula huésped según la reivindicación 3 o 9.
13. Uso de una UTR según la reivindicación 5b, para traducir una región codificante de una molécula de ARN para dar un polipéptido o una proteína codificado por dicha región codificante.

1.) ...TATAGGGAGACGCCACCATG	T7Luc2 (SEQ ID NO:37)
2.) ...TATAGGGAGAGGCCACCATG	T7Luc2 ( $\Delta$ C) (SEQ ID NO:38)
3.) ...TATAGGGAGACAGGCCACCATG	T7Luc2 (+8+A) (SEQ ID NO:39)
4.) ...TATAGGGAGACTGCCACCATG	T7Luc2 (+8+T) (SEQ ID NO:40)
5.) ...TATAGGGAGACGGCCACCATG	T7Luc2 (+8+G) (SEQ ID NO:41)
6.) ...TATAGGGAGACCGCCACCATG	T7Luc2 (+8+C) (SEQ ID NO:42)
7.) ...TATAGGGAGACSp30GCCACCATG	T7Luc2 (+8+Sp30) (SEQ ID NO:43)
8.) ... TATAGGGAGACTCTTCTGGTCCCCAC <u>AGACTCAGAGAGAACGCCACCATG</u>	T7Luc2 (+8+hAg) (SEQ ID NO:44)
9.) ...TATAGGGAGACTGCCAAGATG	T7Luc2 (+8+T)+TISU (SEQ ID NO:45)

Sp30: una secuencia aleatoria de 30 nucleótidos

subrayado en la secuencia n.º 8) 5'UTR de alfa-globina humana (hAg) (30 nucleótidos de longitud)

## Figura 1

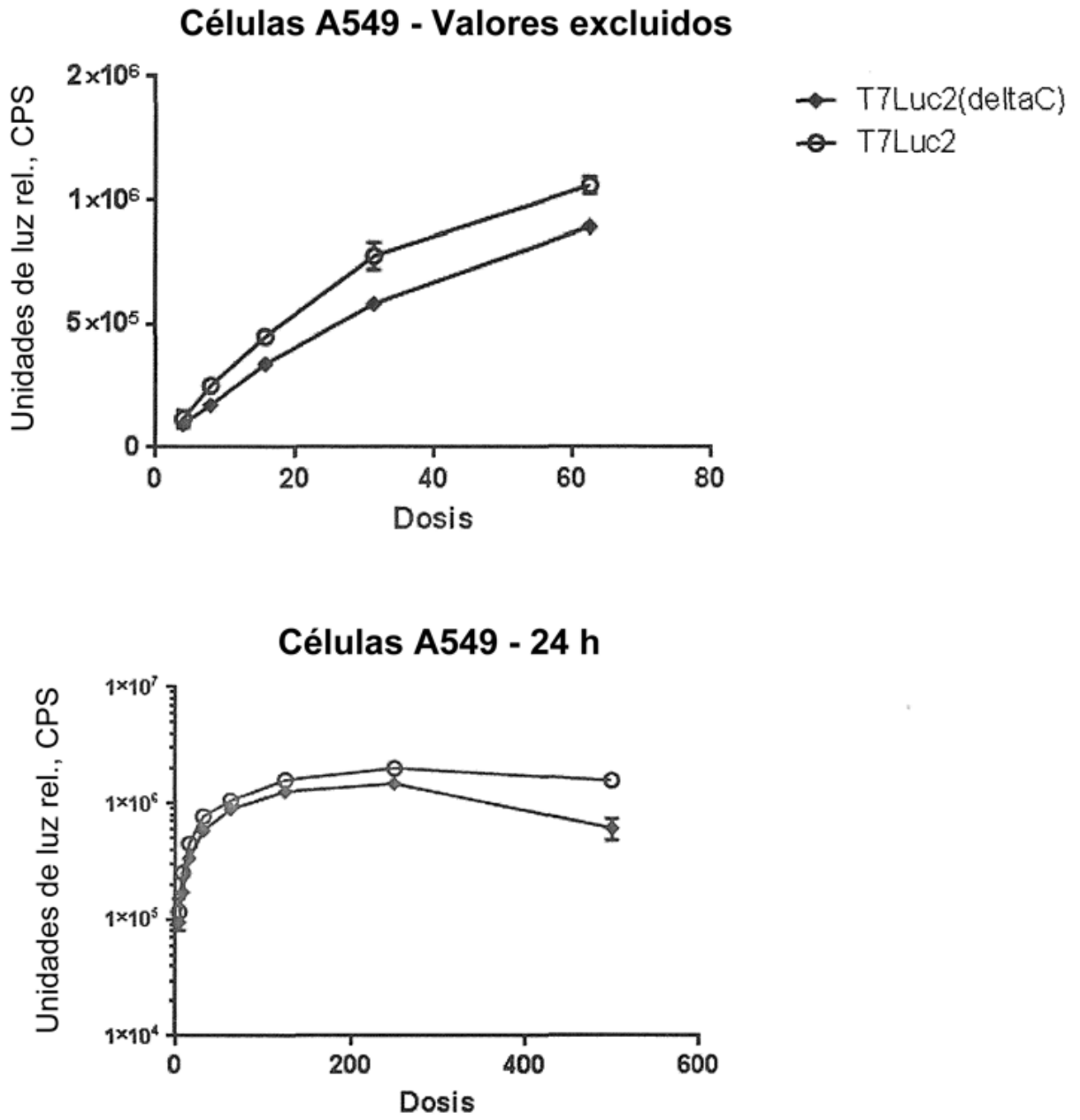


Figura 2A

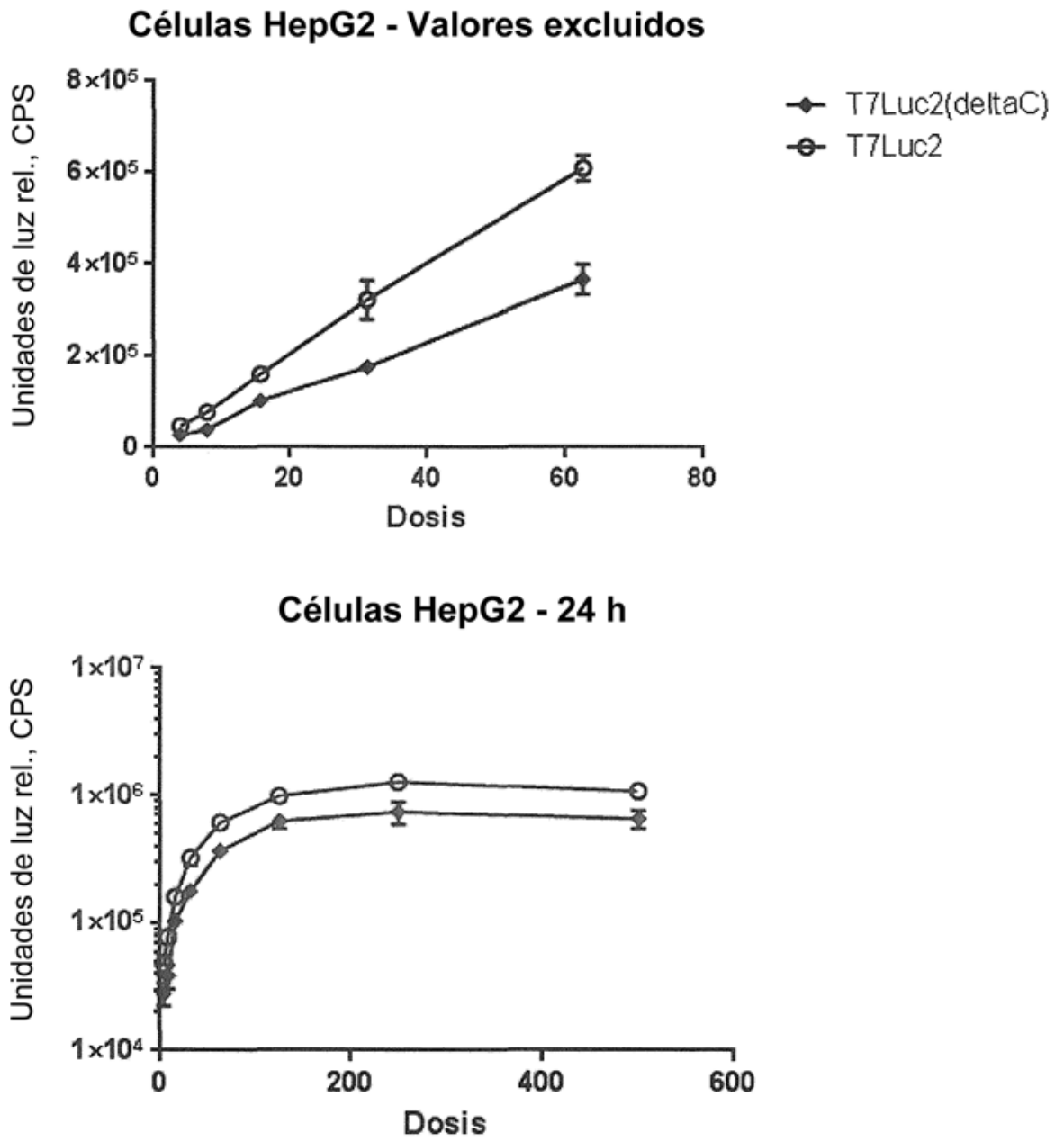


Figura 2B

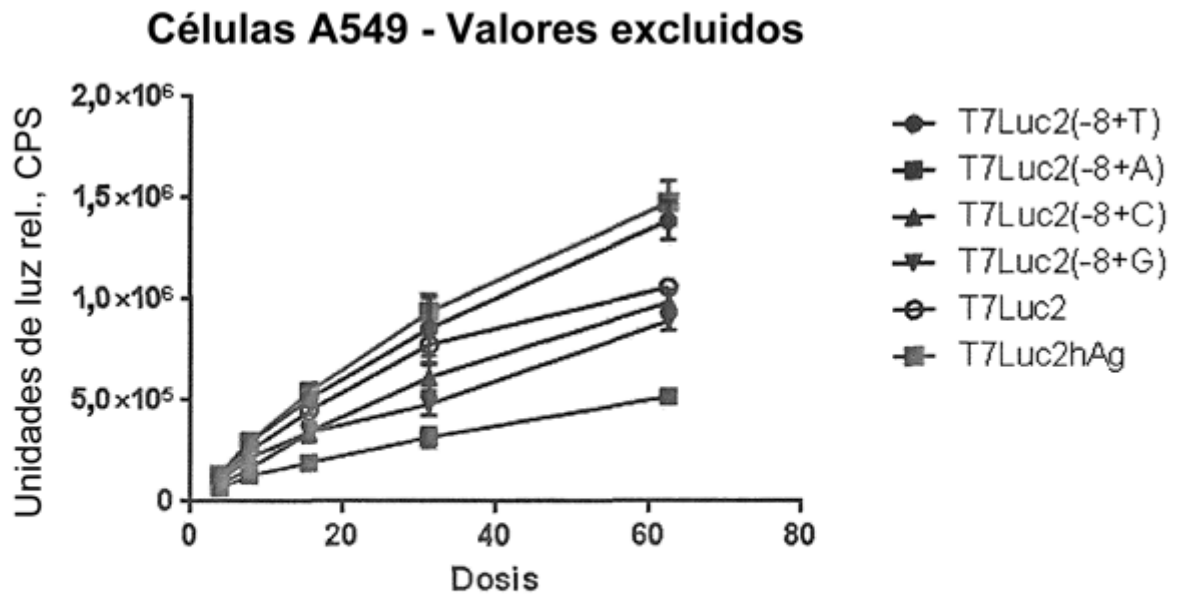


Figura 3

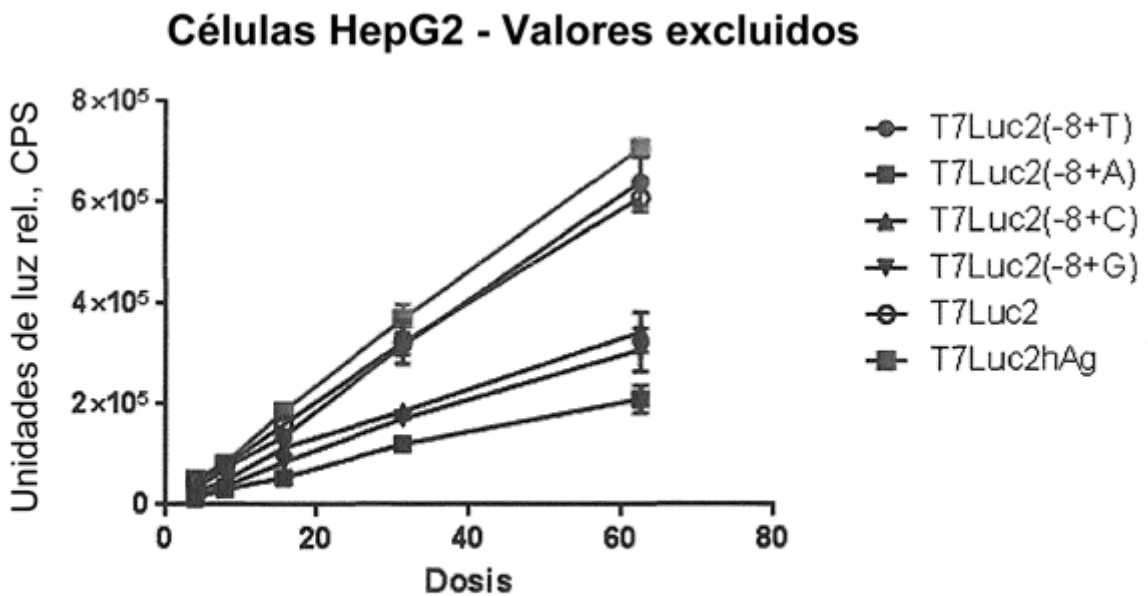


Figura 4

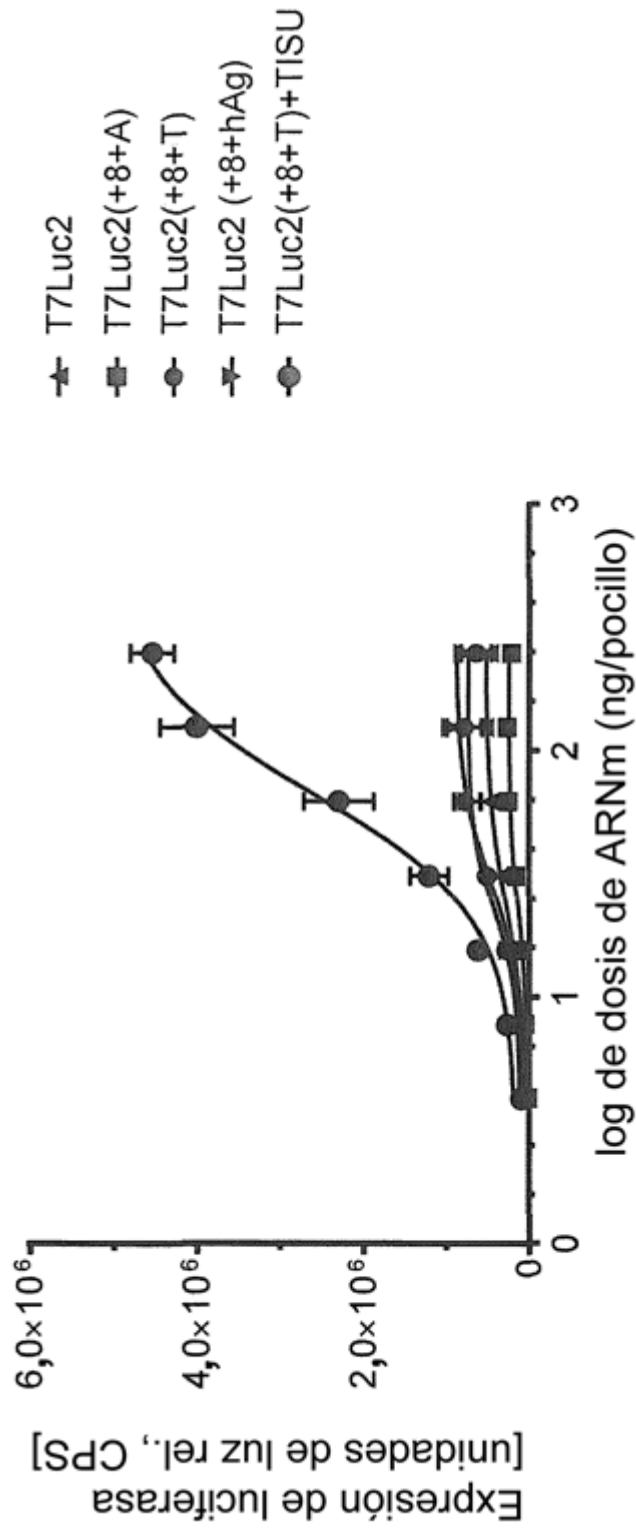


Figura 5A



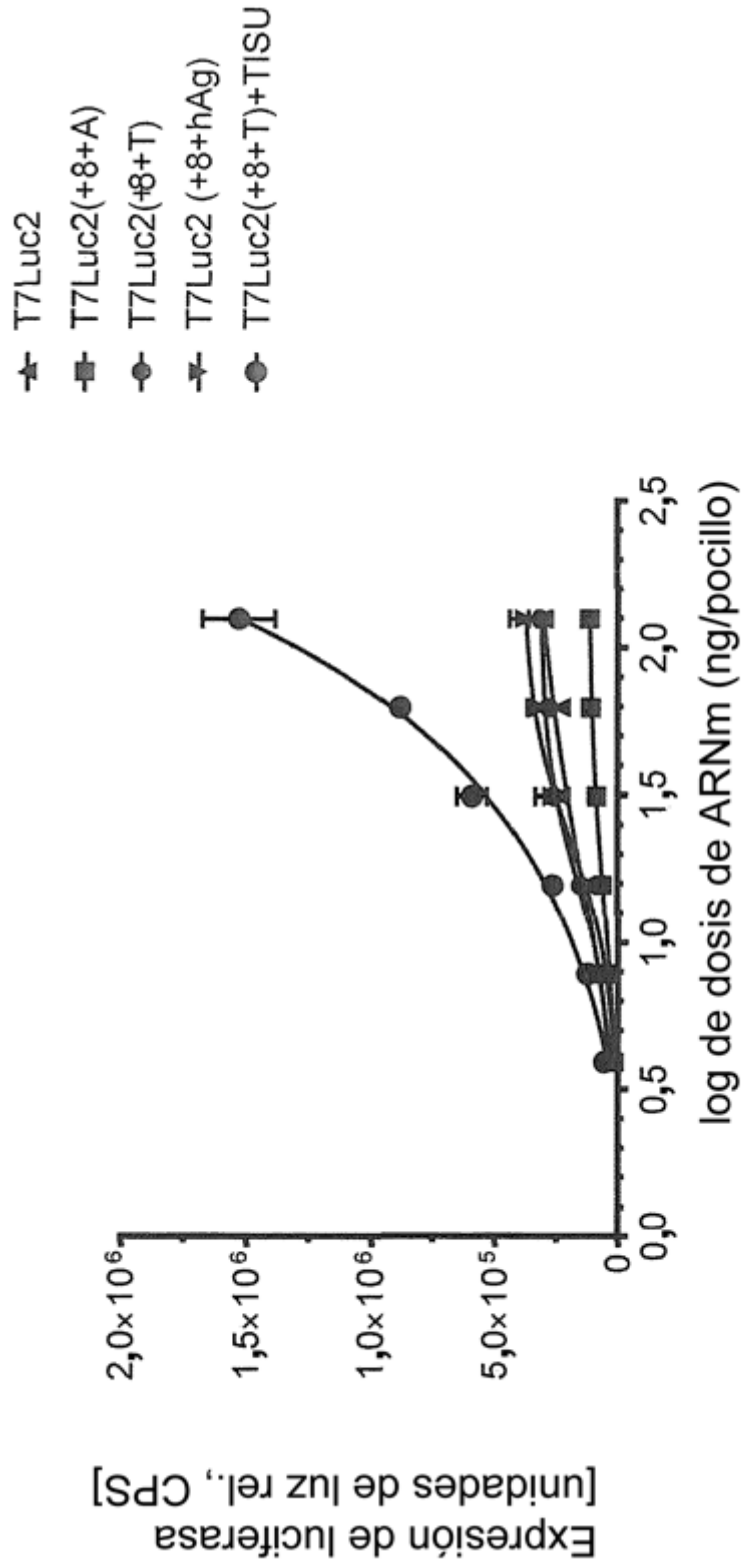


Figura 5B

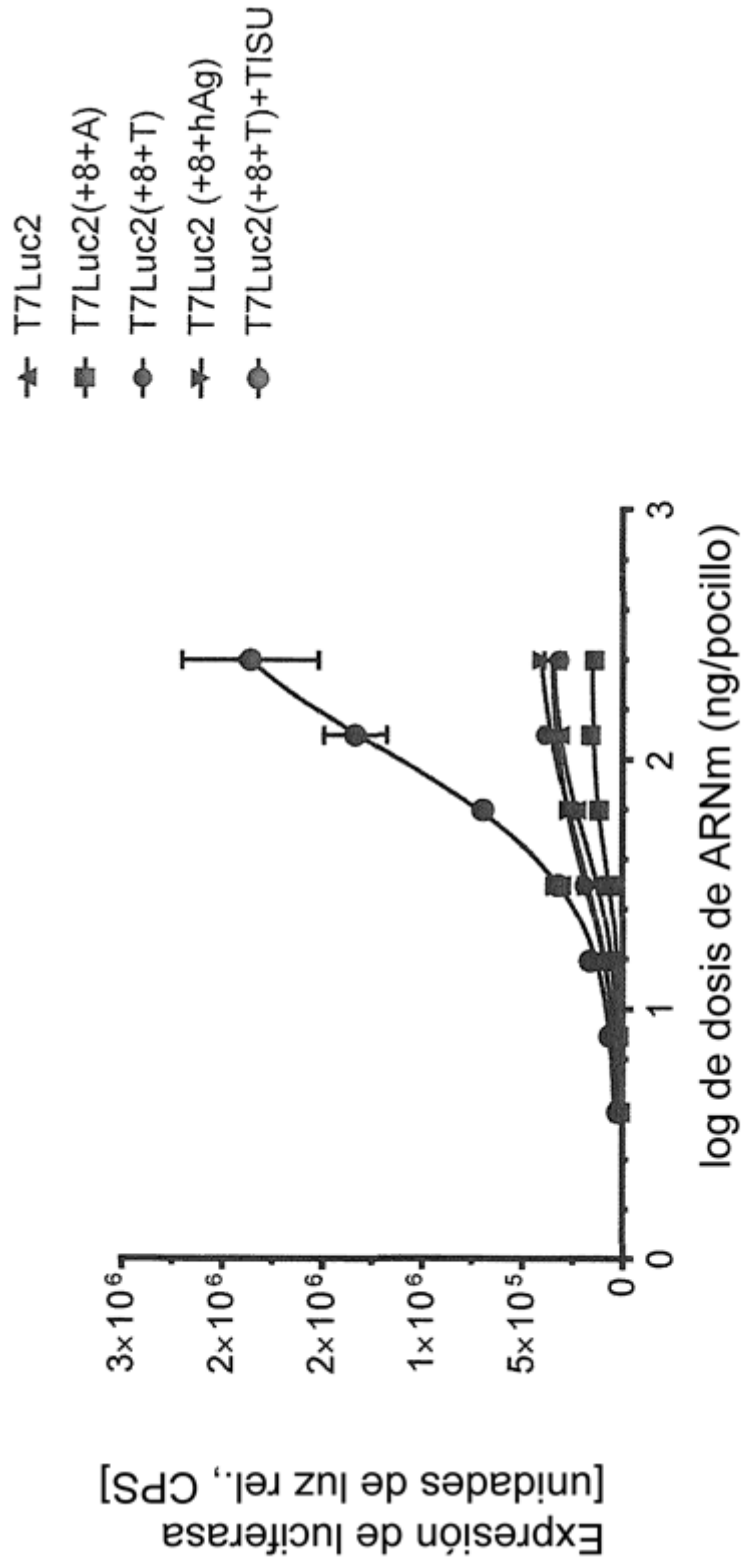


Figura 5C

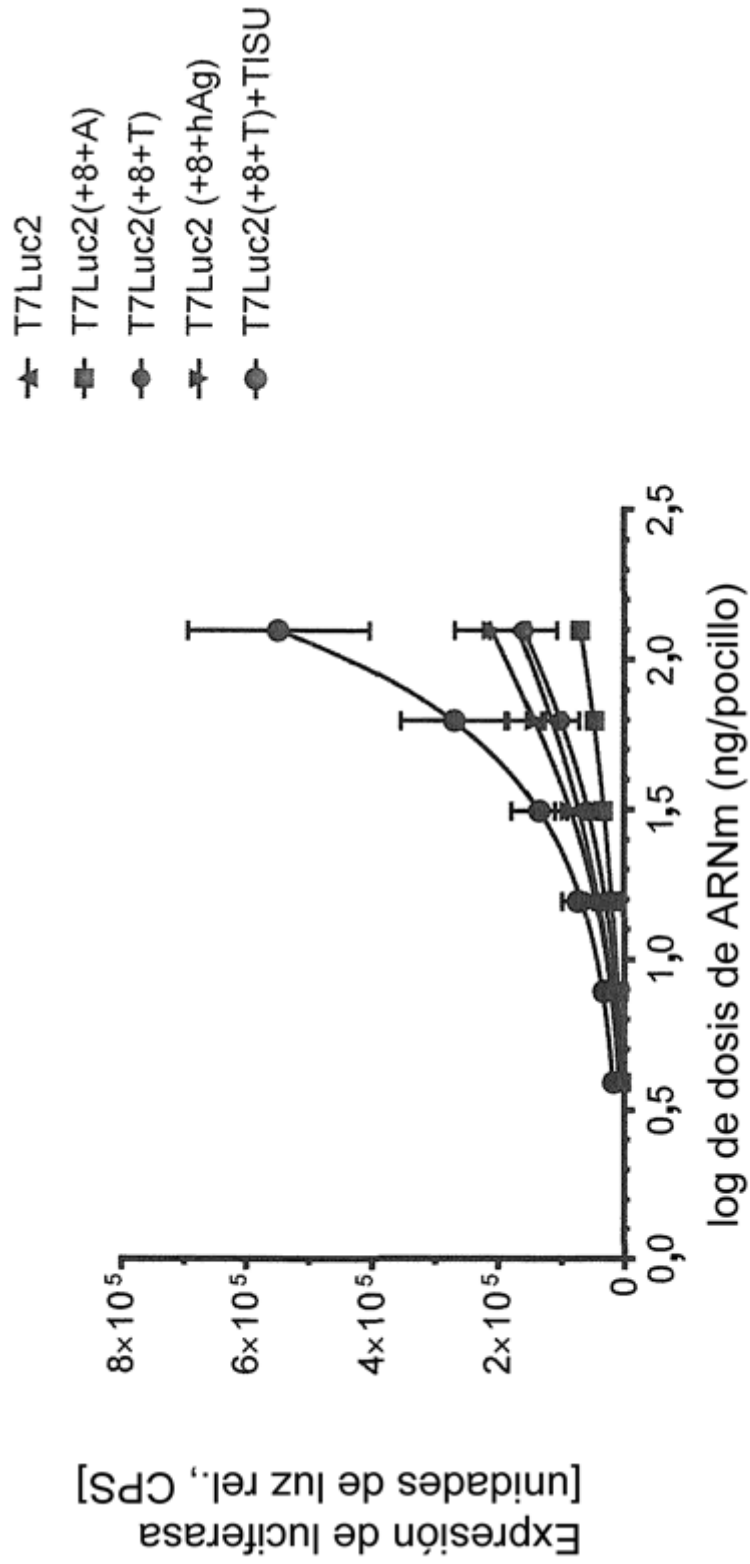


Figura 5D

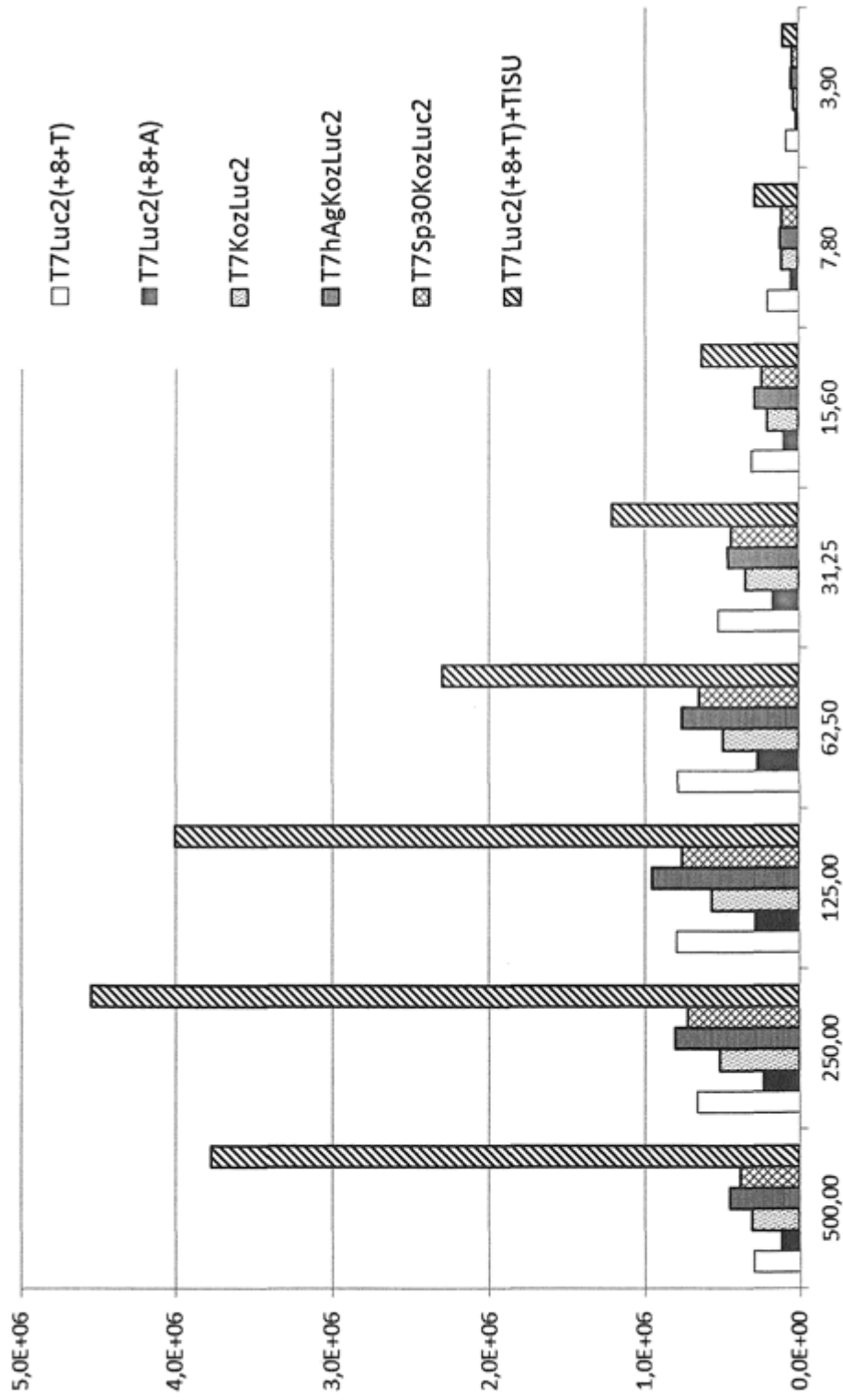


Figura 6A

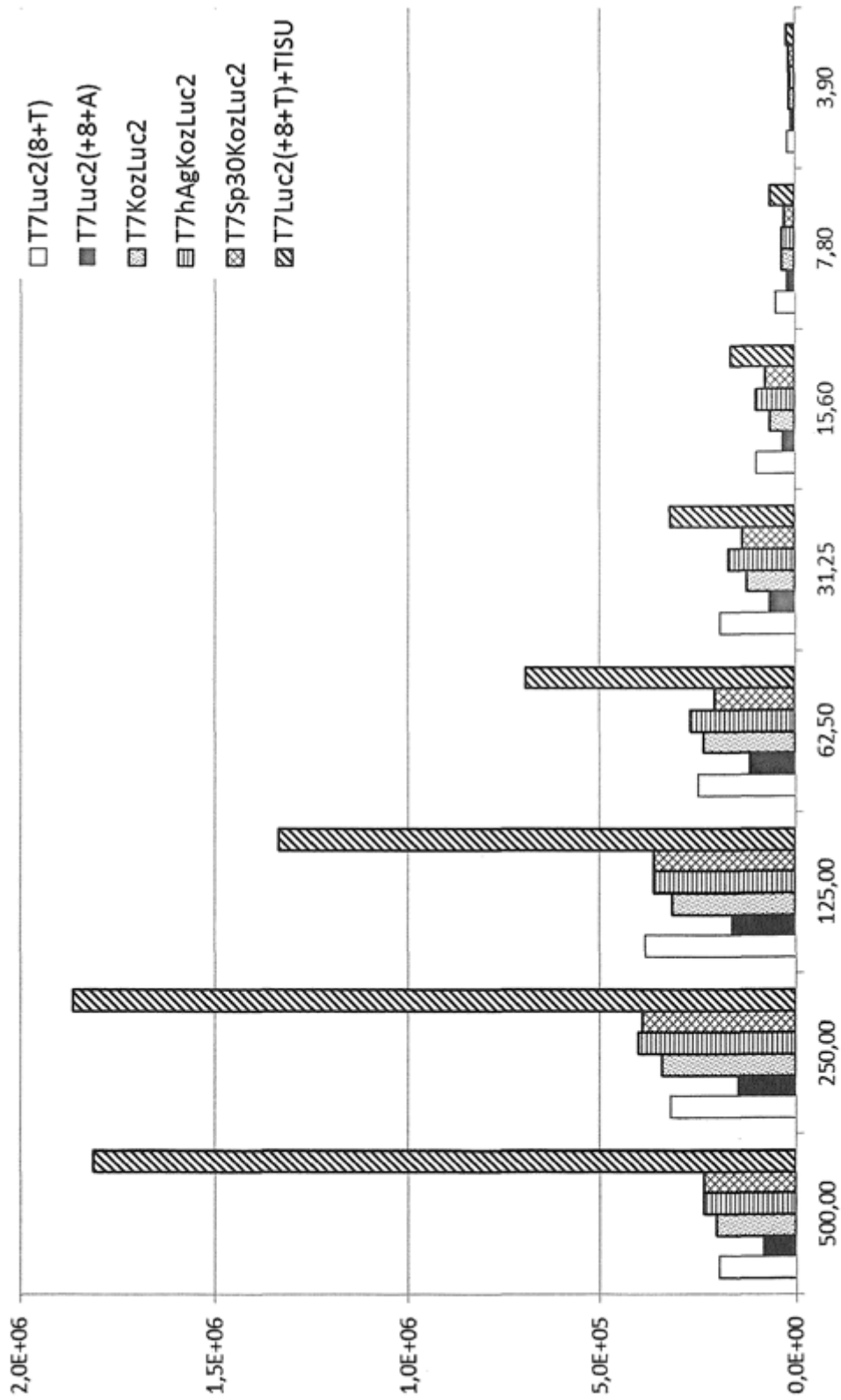


Figura 6B

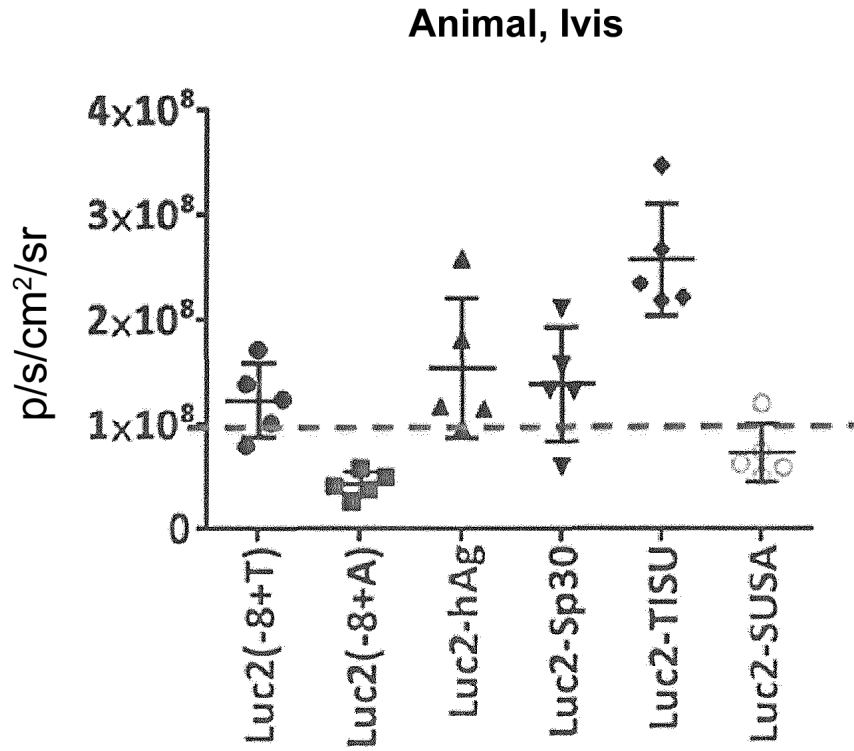


Figura 7A

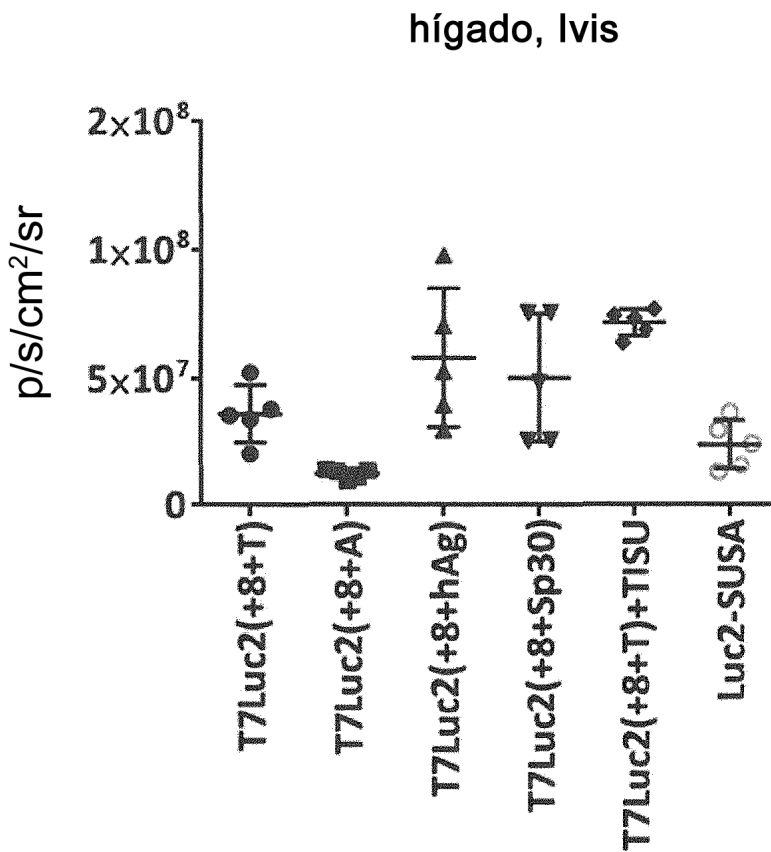


Figura 7B

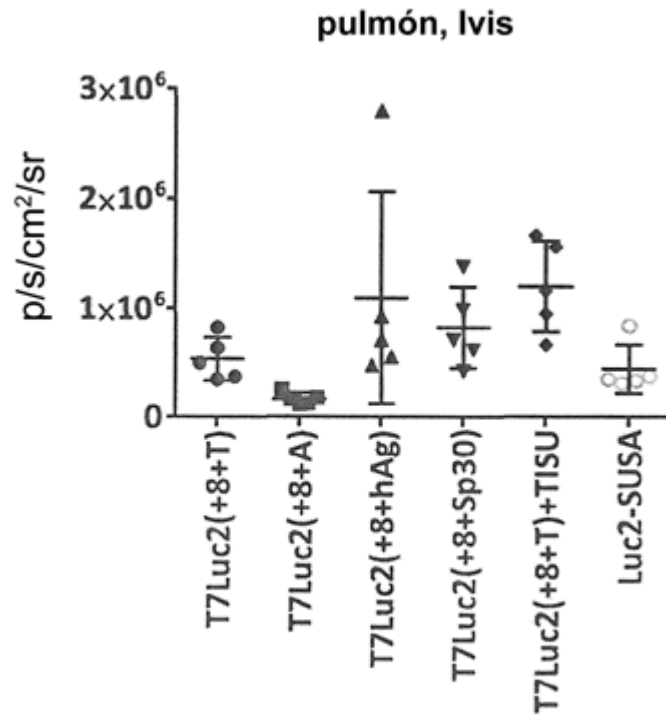


Figura 7C

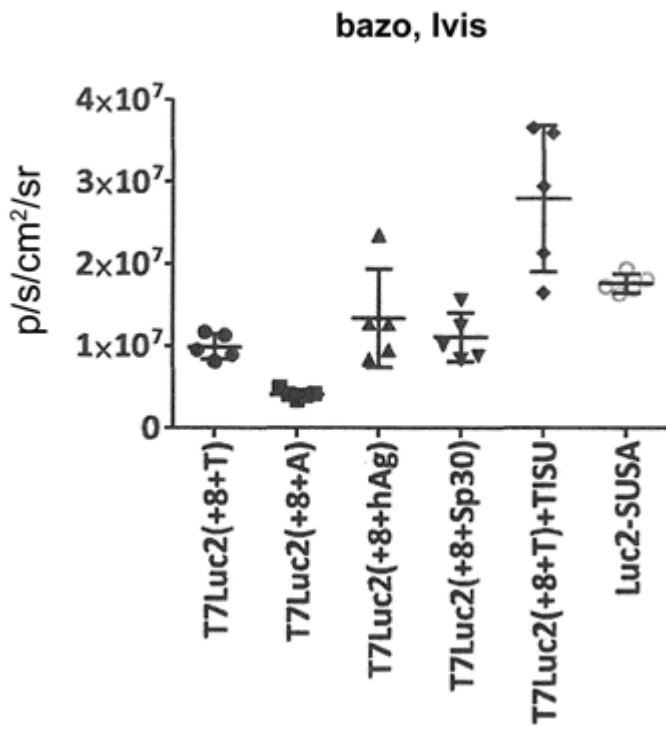
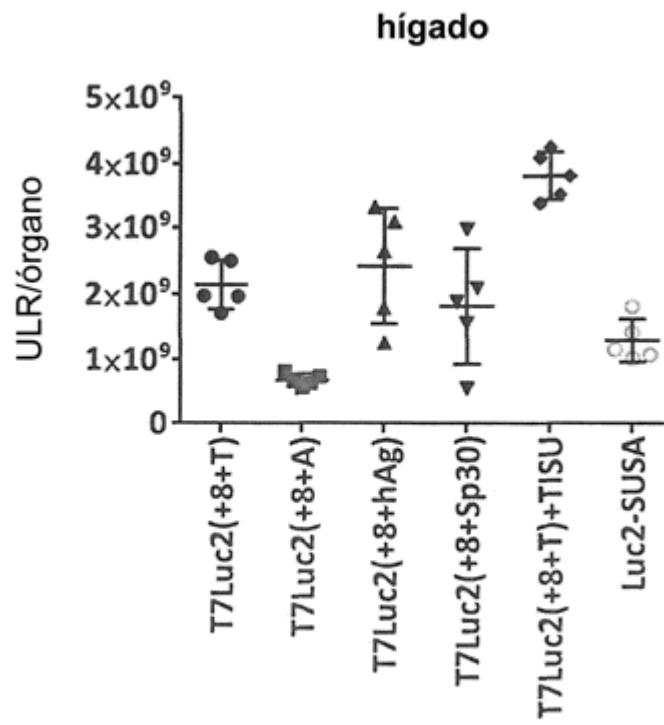
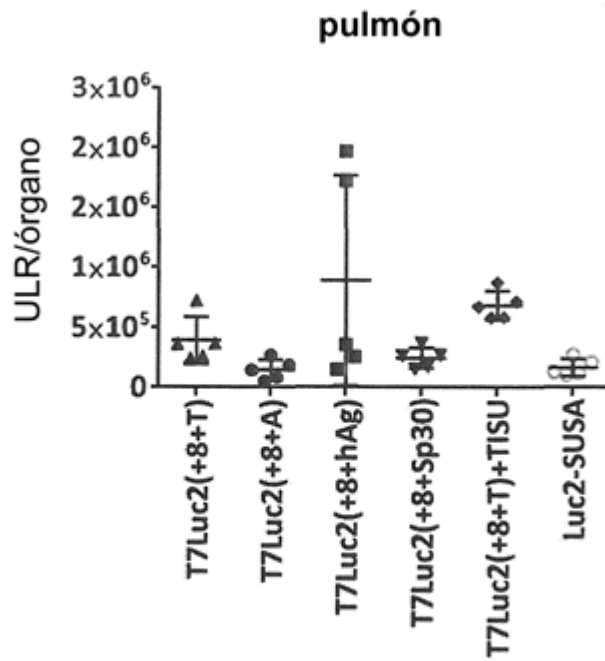


Figura 7D

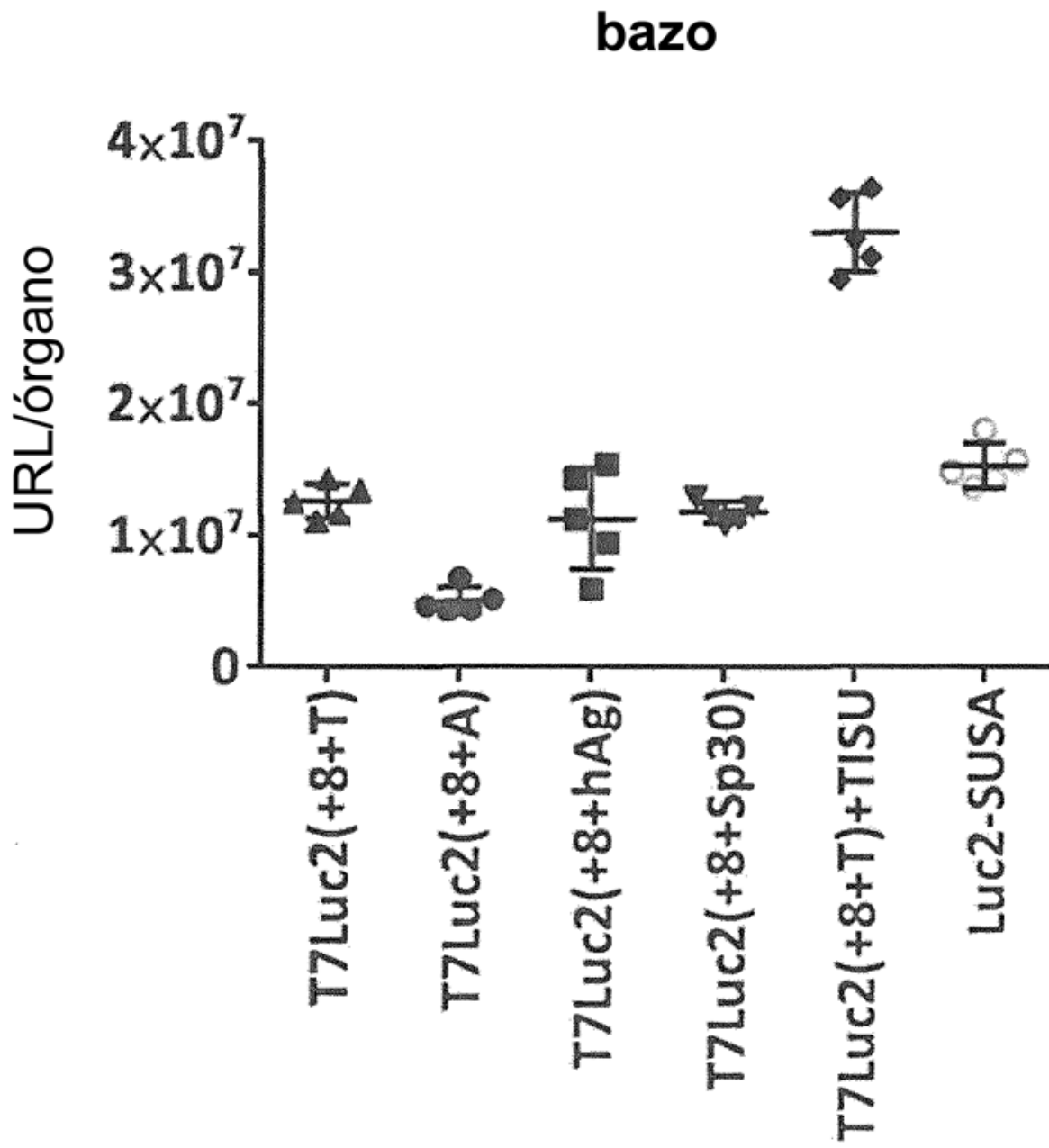


**Figura 7E**

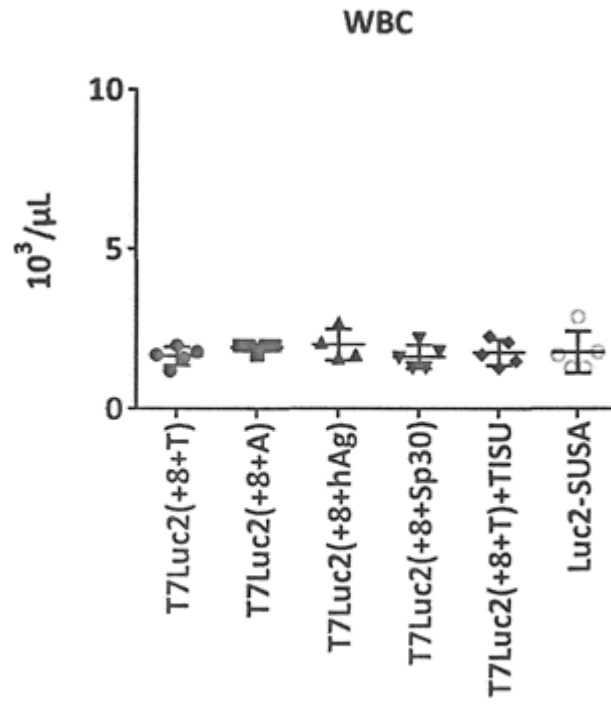


**Figura 7F**

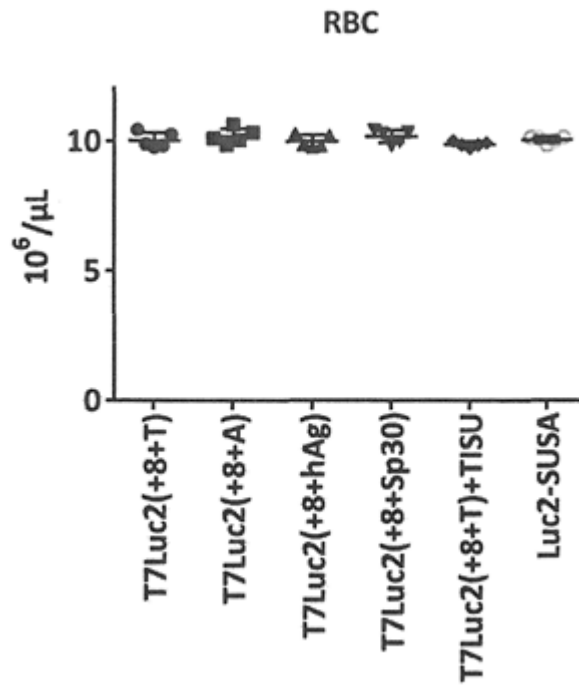




**Figura 7G**



**Figura 8A**



**Figura 8B**

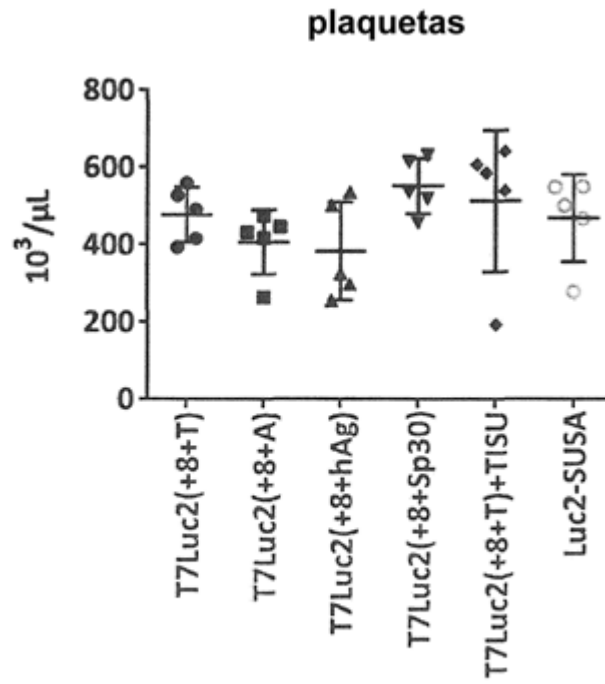


Figura 8C

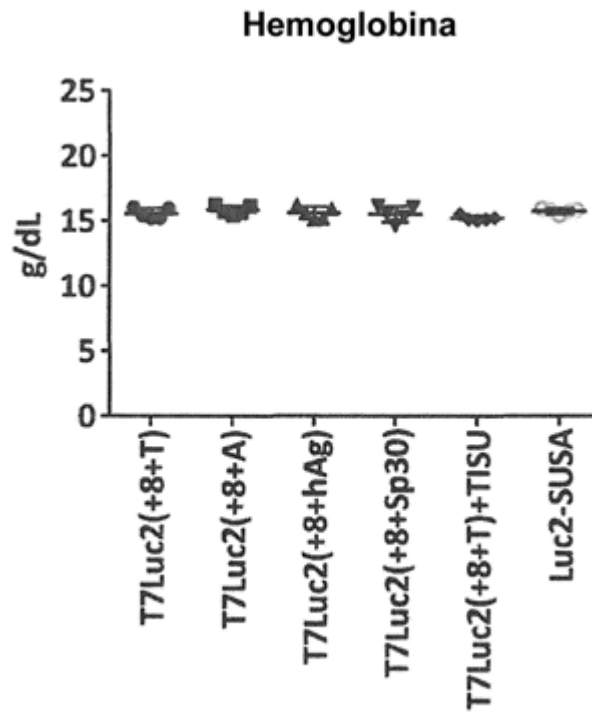


Figura 8D

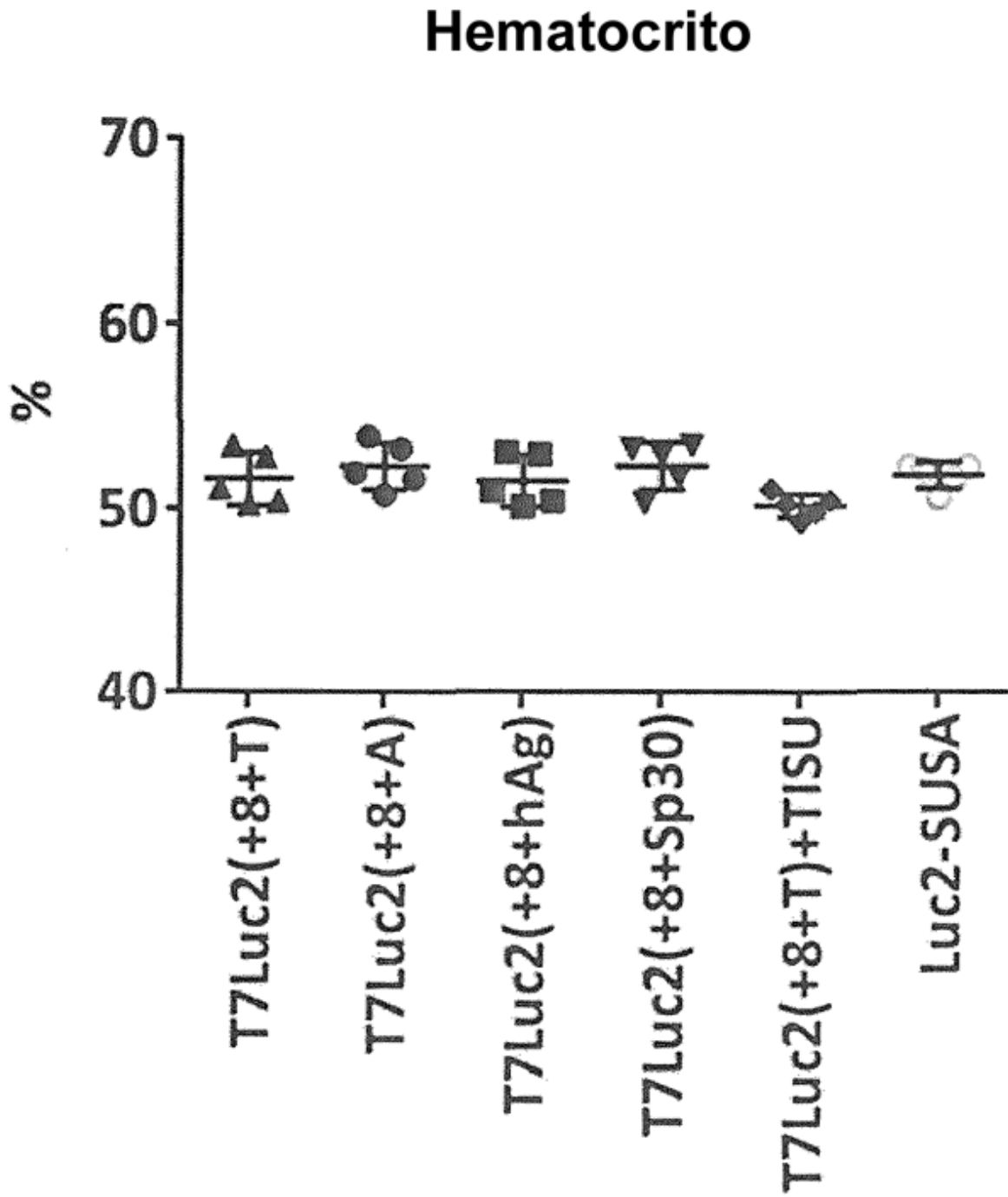


Figura 8E

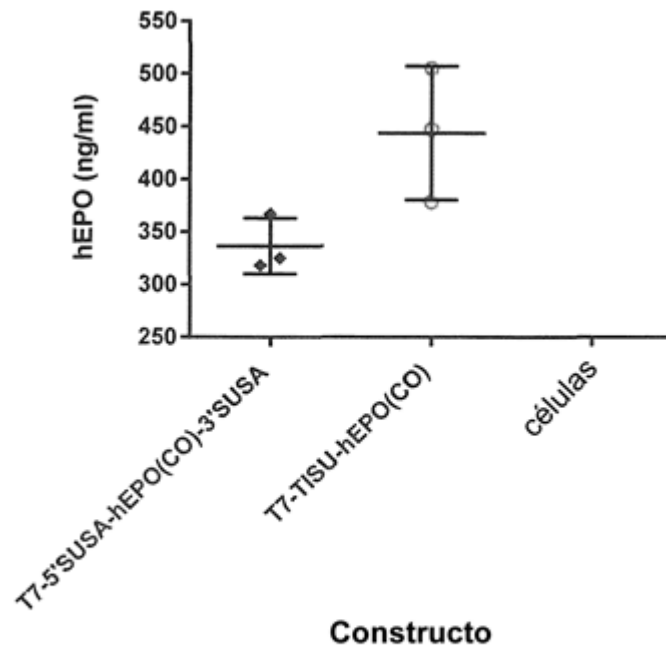


Figura 9A

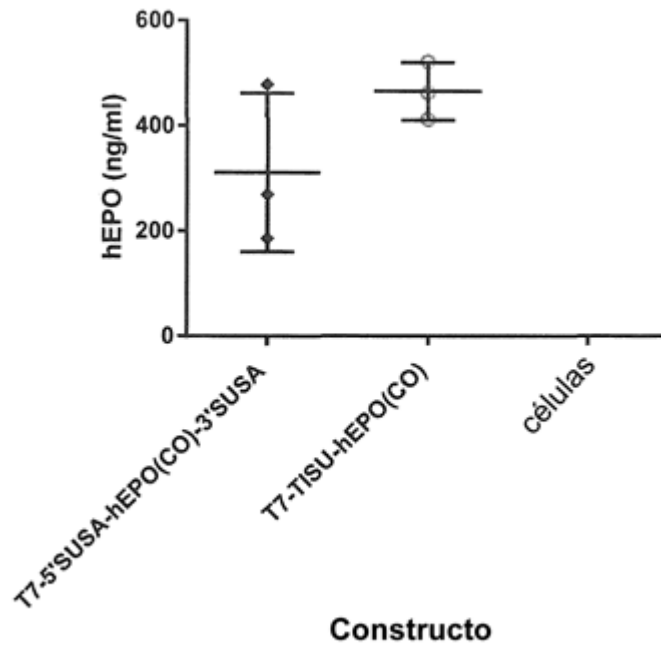
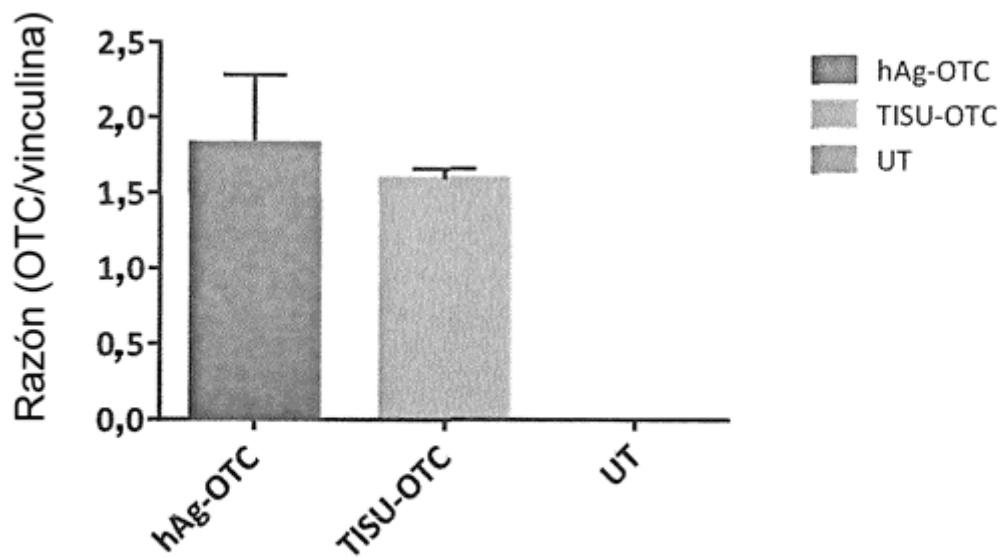


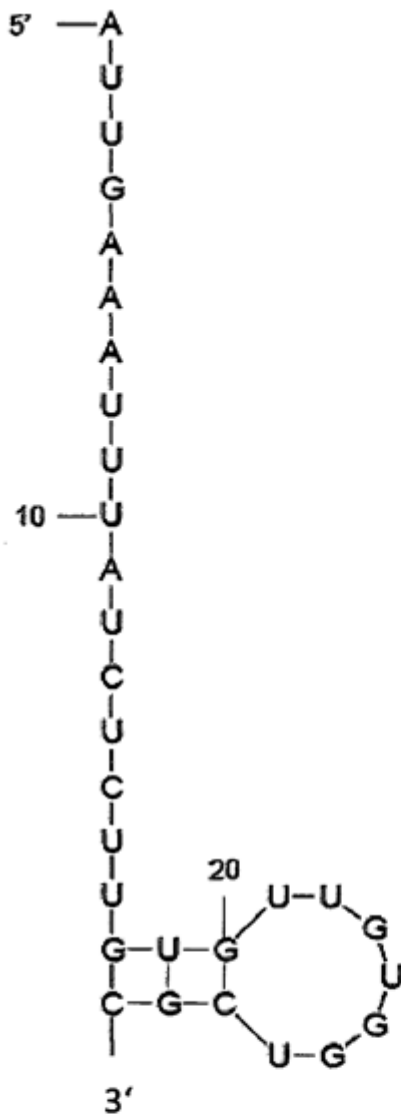
Figura 9B



**Figura 10A**

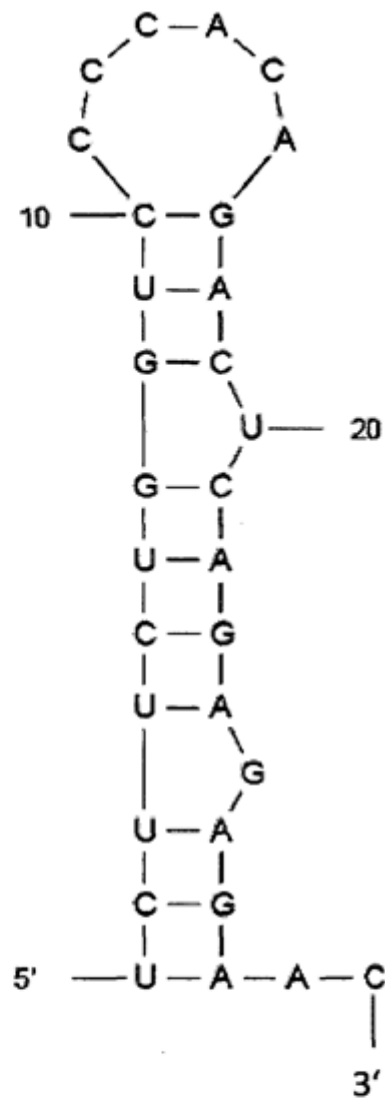


**Figura 10B**



dG=0,50 [inicialmente 0,50]

15Jun11-06-04-27-711141900e



dG=-8,20 [inicialmente -8,20]

15Jun11-06-07-44-bd7612cf14

Figura 11