

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 667 644**

51 Int. Cl.:

C12N 15/67 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **30.06.2016** E 16177122 (5)

97 Fecha y número de publicación de la concesión europea: **07.02.2018** EP 3112469

54 Título: **UTR que aumentan la eficacia de la traducción de las moléculas de ARN**

30 Prioridad:

30.06.2015 EP 15174683

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

14.05.2018

73 Titular/es:

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

72 Inventor/es:

**RUDOLPH, CARSTEN;
ANEJA, MANISH KUMAR;
FERIZI, MEHRIJE y
GEIGER, JOHANNES**

74 Agente/Representante:

ARIAS SANZ, Juan

ES 2 667 644 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

UTR que aumentan la eficacia de la traducción de las moléculas de ARN

5 La presente invención se refiere a una molécula de ARN que comprende (a) una región de codificación que codifica un polipéptido; y (b) cadena arriba de dicha región de codificación, una o más UTR que comprenden la secuencia mostrada en SEQ ID NO: 1 o una secuencia que muestra 1 a 4 sustituciones en comparación con SEQ ID NO: 1 y que da lugar a una molécula de ARN que tiene la misma o mayor eficacia de traducción que una molécula de ARN que comprende una UTR que comprende la SEQ ID NO: 1; y/o (c) cadena abajo de dicha región de codificación, una
10 o más UTR que comprenden la secuencia mostrada en SEQ ID NO: 2 o una secuencia que muestra 1 a 7 sustituciones en comparación con SEQ ID NO: 2 y que da lugar a una molécula de ARN que tiene la misma o mayor eficacia de traducción que una molécula de ARN que comprende una UTR que comprende la SEQ ID NO: 2; en la que dicho polipéptido codificado por dicha región de codificación no es un polipéptido alfa del citocromo b-245 (CYBA). Además, la presente invención se refiere a una molécula de ácido nucleico que codifica la molécula de ARN de acuerdo con la presente invención. Además, la presente invención se refiere a un vector que comprende la molécula de ácido nucleico de acuerdo con la presente invención y a una célula hospedadora que comprende el vector de acuerdo con la presente invención. Además, la presente invención se refiere a una composición farmacéutica que comprende la molécula de ARN de acuerdo con la presente invención y, opcionalmente, un vehículo farmacéuticamente aceptable. Además, la presente invención se refiere a un kit que comprende la molécula
20 de ARN de acuerdo con la presente invención. Finalmente, la presente invención se refiere al uso de una o más UTR según lo definido en (b) y/o una o más UTR según lo definido en (c) para aumentar la eficacia de traducción de una región de codificación de una molécula de ARN en un polipéptido o una proteína codificada por dicha región de codificación.

25 En los últimos años, el ARN mensajero (ARNm) se ha vuelto cada vez más relevante como nueva entidad farmacológica. A diferencia de la terapéutica génica basada en el ADN, el ARNm no necesita ser transportado al núcleo, sino que se traduce directamente en proteína en el citoplasma (1,2). Esto hace que el ARNm sea más seguro para evitar la posible mutagénesis por inserción, un riesgo poco probable, pero que existe, de los medicamentos génicos de ADN. Por consiguiente, las terapias de ARNm están surgiendo como alternativas prometedoras para terapias de reemplazo de genes y proteínas en una amplia variedad de indicaciones médicas (1-4). Sin embargo, se han de resolver la fuerte inmunogenicidad, así como la estabilidad limitada del ARNm convencional para establecer mejor su aplicabilidad clínica. Con respecto a esto, la estabilidad del ARNm y, en particular, la velocidad de traducción del ARNm es un parámetro esencial para las aplicaciones médicas previstas, porque determina, por ejemplo, la dosis y los intervalos de dosificación de los fármacos de ARNm. Varias estrategias han demostrado ser exitosas tanto para aumentar la estabilidad como para reducir la respuesta inmunogénica desencadenada por el ARNm administrado a células u organismos. Entre ellas, se encuentra la inclusión de nucleótidos químicamente modificados (5). Kormann *et al.* han demostrado que basta con un reemplazo solo del 25 % de los restos de uridina y citidina por 2-tiouridina y 5-metil-citidina para aumentar la estabilidad del ARNm y reducir la activación de la inmunidad innata desencadenada por ARNm administrado externamente *in vitro*
40 (documentos WO2012/0195936 A1; WO2007024708 A2). Además, se ha informado que las regiones no traducidas (UTR) en los ARNm desempeñan un papel fundamental en la regulación tanto de la estabilidad del ARNm como de la traducción del ARNm. Se sabe que las UTR influyen en el inicio, la elongación y la terminación de la traducción, así como en la estabilización del ARNm y la localización intracelular a través de su interacción con las proteínas de unión al ARN (6, 7). Dependiendo de los motivos específicos dentro de la UTR, se puede aumentar o reducir el recambio del ARNm (8-11). Recientemente, se han publicado datos sobre semividas de ARNm y las correspondientes secuencias de UTR (12, 43).

Por consiguiente, aunque en la técnica anterior ya se han descrito medios y métodos para aumentar la estabilidad del ARNm reduciendo la respuesta inmunogénica desencadenada por ARNm administrado a células u organismos y aumentando la eficacia de la traducción, todavía existe la necesidad de mejoras, en particular, en lo que respecta a medios adicionales o alternativos para aumentar la eficacia de traducción, ya que la eficacia de traducción es un parámetro esencial para aplicaciones médicas previstas, porque determina, por ejemplo, la dosis y los intervalos de dosificación de los fármacos de ARNm y, en última instancia, determina la biodisponibilidad del producto final, es decir, del péptido o de la proteína codificados.

55 La presente solicitud aborda esta necesidad proporcionando las realizaciones como se definen en las reivindicaciones.

60 En particular, la presente solicitud encontró sorprendentemente que una determinada UTR confiere una mayor eficacia de traducción cuando se fusiona con un ARNm (foráneo) dado. La UTR se deriva de un ARNm del gen del *polipéptido alfa del citocromo b-245* (CYBA) humano. El gen CYBA comprende UTR 5' y 3' específicas. En general, se sabe que los motivos UTR 5' tales como los marcos de lectura abierta cadena arriba (uORF) o los sitios internos de entrada ribosómica (IRES) participan en la regulación génica, en particular, en el inicio de la traducción (13). Las UTR 3' pueden comprender incluso más funciones reguladoras que las UTR 5', obstaculizando algunas de ellas incluso la traducción del ARNm (14). El descubrimiento de la presente invención es tanto más sorprendente cuanto que, en la técnica anterior, no se han descrito motivos reguladores para la unidad UTR 5' del CYBA. Aunque se sabe

que la UTR 3' del CYBA contiene dos motivos reguladores, el hallazgo de la presente invención de que las UTR del CYBA confieren una mayor eficacia de la traducción cuando se fusionan con un ARNm dado es sorprendente, ya que estos dos motivos se describen en el contexto de la estabilidad del ARNm, pero no en el aumento de la eficacia de la traducción. Más concretamente, se sabe que la UTR 3' del CYBA alberga una señal de poliadenilación (PAS) que se sabe que interactúa con la proteína de unión al elemento de poliadenilación citoplasmática (CPEB), así como con el factor de señalización de la escisión y la poliadenilación (CPSF) (11). Se sabe que CPEB es responsable de la prolongación de la cola de poli-A en el citoplasma, mientras que CPSF prepara el pre-ARNm a través de la escisión en un sitio específico para la próxima adición de poli-A (11, 14). Un segundo motivo regulador contenido en la UTR 3' del CYBA es el elemento de estabilidad de la UTR 3' de la insulina (INS_SCE) (15). Se ha demostrado que la secuencia de INS_SCE se une a la proteína de unión al tracto de polipirimidina (PTB) en condiciones reductoras, lo que aumenta la semivida del ARNm de la insulina (15). Por lo tanto, ambos motivos reguladores de la UTR 3' de CYBA están vinculados predominantemente a la estabilidad del ARNm.

Las secuencias de ADN que muestran la secuencia de nucleótidos de las UTR 5' y 3' del gen CYBA humano presentes en la cadena de codificación del gen CYBA humano se muestran en la siguiente **Tabla 1**.

Tabla 1: Código genético de las UTR del gen CYBA humano

Región no traducida	Secuencia de ADN (de 5' a 3')
5'	GGCGGGGTTTCGGCCGGGAGCGCAGGGGGCGGCAGTGC GCGCCT
	AGCAGTGTCCCAGCCGGGTTCTGTGTCGCC (SEQ ID NO:5)
3'	<u>CCTCGCCCCGGACCTGCCCTCCCGCCAGGTGCACCCACCTGCA</u> <u>ATAAATGCAGCGAAGCCGGGAGCGCGT</u> (SEQ ID NO:6)

La **Tabla 1** muestra el código genético exacto de las UTR del gen CYBA humano. Las secuencias de ADN se muestran desde el extremo 5' hasta el extremo 3'. La señal de poliadenilación (PAS) de la UTR 3' se muestra en negrita, y el elemento de estabilidad de la UTR 3' de la insulina (INS_SCE) está subrayado. La UTR 5' consiste en 71 pares de bases, mientras que la UTR 3' contiene 70 pares de bases. Ambas UTR son más cortas que las UTR humanas medias, que consisten en alrededor de 200 nucleótidos en el caso de la UTR 5' y aproximadamente 1.000 nucleótidos en el caso de la UTR 3'.

En la **Tabla 1** anterior, las secuencias de ADN que muestran las UTR 5' y 3' del gen CYBA humano se muestran como SEQ ID NO: 5 y SEQ ID NO: 6, respectivamente. En vista del hecho de que la presente invención se refiere predominantemente a una molécula de ARN, se hace referencia a continuación a las secuencias de ARN correspondientes. Derivada de la secuencia de ADN anterior, la SEQ ID NO: 5 corresponde a la siguiente secuencia de UTR en el ARN:

5'-CGCGCCUAGCAGUGUCCAGCCGGUUCGUGUCGCC-3' (SEQ ID NO: 1).

Esta secuencia UTR 5' precede inmediatamente al codón de inicio del gen CYBA humano. Derivada de la secuencia de ADN anterior, SEQ ID NO: 6 corresponde a la siguiente secuencia de UTR en el ARN:

5'-CCUCGCCCCGGACCUGCCUCCCGCCAGGUGCACCC
ACCUGCAAUAAAUGCAGCGAAGCCGGGA-3' (SEQ ID NO:2)).

Otra característica importante que influye en la eficacia de la traducción del ARNm es la cola de poli-A, que se encuentra en el extremo 3'. Se ha demostrado que una prolongación de la cola de poli-A hasta 120 nucleótidos tiene efectos beneficiosos sobre la expresión de proteínas, presumiblemente debido al efecto protector de las colas poli-A más largas contra la degradación del ARNm (16). A diferencia de las colas poli-A largas, se afirma que los ARNm con colas poli-A que tienen menos de 50 nucleótidos no se traducen del todo (11, 17). Por lo tanto, en la terapia de ARNm, las construcciones de ARNm recombinante se proveen ventajosamente con una cola de poli-A de 120 nucleótidos o más. La degradación de la mayoría de las transcripciones de ARNm en células eucariotas comienza con una desadenilación exonucleolítica de 3' a 5', que genera la eliminación de la mayor parte de la cola de poli-A. Posteriormente, se sabe que entran en juego dos rutas principales que son responsables de la degradación del resto

del cuerpo del ARNm. Por un lado, el extremo 5' es desprotegido del capuchón por el complejo Dcp1/Dcp2, a lo que le sigue una degradación exonucleolítica 5'-3' catalizada por Xrn1p. Por otro lado, el exosoma permite una degradación exorribonucleolítica de 3'-5' con la retención del capuchón 5' (18). Además, se sabe que la interacción del capuchón 5' con la cola de poli-A 3' da lugar a formas circulares del ARNm. Se supone que la forma circular del ARNm aumenta la tasa de inicio de los ribosomas tras la traducción del primer codón de terminación y también protege al ARNm contra la degradación (19).

La presente solicitud, entre otras cosas, encontró que, sorprendentemente, se puede conferir un aumento de la eficacia de la traducción de un ARNm de CYBA natural a un ARNm foráneo en virtud de flanquear su secuencia de codificación con combinaciones de UTR 5' y 3' de CYBA acortadas. En este sentido, cabe señalar que tanto la UTR 5' como la UTR 3' de la presente invención, mostradas en SEQ ID NO: 1 y SEQ ID NO: 2, respectivamente, son más cortas que las secuencias de ADN anteriores que presentan las UTR 5' y 3' del gen CYBA humano que se muestran como SEQ ID NO: 5 y SEQ ID NO: 6, respectivamente. Esto se ha realizado mediante un análisis de una sola célula de películas de lapso de tiempo de transfección del ARNm, que recientemente han demostrado ser capaz de evaluar cursos de tiempo de expresión individuales (26) mientras que se ha informado que es posible usar micropatrones regulares para ubicar células en una rejilla regular de sitios de adhesión (27). Por lo tanto, la presente solicitud ha demostrado que esta tecnología ofrece la resolución de seleccionar rápidamente y comparar diferentes combinaciones de UTR en un ARNm foráneo. Para abordar esto, se ha escogido la secuencia de codificación de la proteína de fluorescencia verde mejorada desestabilizada (d2EGFP) para acortar artificialmente el ciclo de vida de la proteína indicadora dentro de la célula (28). Las combinaciones incluían la inserción de las respectivas UTR de CYBA en los extremos 5' o 3', respectivamente, en ambos extremos 5' y 3', en el extremo 5' combinado con dos repeticiones de la UTR 3' en el extremo 3', o dos repeticiones de UTR 3' sin UTR 5'. Se compararon todas ellas con una construcción de control sin UTR. Se evaluaron los tiempos de vida del ARNm funcional y de la proteína, y la velocidad de expresión de cada una de las transcripciones comparadas. Se comparó el análisis de células individuales de la dinámica de la expresión génica después de la transfección del ARNm con los métodos basados en la población (citometría de flujo, imágenes de microscopía de fluorescencia y la medición de la bioluminiscencia de la actividad de la luciferasa). Sorprendentemente, se ha demostrado que se mejora la expresión de proteína total durante un periodo de tres días para todas las combinaciones de UTR en comparación con el control.

Este hallazgo conduce a la provisión de las realizaciones como se caracterizan en las reivindicaciones. Por lo tanto, la presente invención se refiere a una molécula de ARN que comprende:

(a) una región de codificación que codifica un polipéptido; y

(b) cadena arriba de dicha región de codificación, una o más UTR que comprenden la secuencia mostrada en SEQ ID NO: 1 o una secuencia que muestra de 1 a 4 sustituciones en comparación con SEQ ID NO: 1 y que da lugar a una molécula de ARN que tiene la misma o mayor eficacia de traducción que una molécula de ARN que comprende una UTR que comprende la SEQ ID NO: 1; y/o

(c) cadena abajo de dicha región de codificación, una o más UTR que comprenden la secuencia mostrada en SEQ ID NO: 2 o una secuencia que muestra de 1 a 7 sustituciones en comparación con SEQ ID NO: 2 y que da lugar a una molécula de ARN que tiene la misma o mayor eficacia de traducción que una molécula de ARN que comprende una UTR que comprende la SEQ ID NO: 2,

en la que dicho polipéptido codificado por dicha región de codificación no es un polipéptido alfa del citocromo b-245 (CYBA).

Una molécula de ácido ribonucleico (ARN), como se usa de acuerdo con la presente invención, se refiere a una molécula polimérica que está ensamblada como una cadena de los nucleótidos denominada G, A, U y C. Cada nucleótido del ARN contiene un azúcar de ribosa, con átomos de carbono numerados del 1' al 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 polimérico, un grupo fosfato está unido a la posición 3' de una ribosa y a la posición 5' de la siguiente. Por lo tanto, los nucleótidos en una molécula de ARN polimérico están unidos covalentemente entre sí, en la que el grupo fosfato de un nucleótido se une al átomo de carbono 3' del nucleótido subsiguiente, formando de ese modo un enlace fosfodiéster. Por consiguiente, una cadena de ARN tiene un extremo 5' y un extremo 3', así denominado por los átomos de carbono en el anillo de ribosa. Por convención, cadena arriba y cadena abajo se refieren al sentido de 5' a 3' en el que tiene lugar la transcripción del ARN. Preferentemente, la molécula de ARN es una molécula de ARN mensajero (ARNm). El ARNm es una gran familia de moléculas de ARN que transmite información genética desde el ADN al ribosoma, donde especifican la secuencia de aminoácidos de los productos proteicos de la expresión génica. Tras la transcripción del ARNm de transcripción primaria (conocido como pre-ARNm) por la ARN polimerasa, el ARNm maduro, procesado, se traduce a un polímero de aminoácidos: una proteína, como se resume en el dogma central de la biología molecular. Al igual que en el ADN, la información genética del ARNm está en la secuencia de nucleótidos, que están dispuestos en codones que consisten en tres bases cada uno. Cada codón codifica un aminoácido específico, a excepción de los codones de parada, que terminan la síntesis de proteínas.

Como se describirá con más detalle a continuación, una molécula de ácido ribonucleico (ARN) de la presente invención comprende dos o incluso tres módulos principales, es decir, (a) una región de codificación que codifica un

polipéptido; (b) cadena arriba de dicha región de codificación, una o más UTR; y/o (c) cadena abajo de dicha región de codificación, una o más UTR que son diferentes de las UTR del módulo (b). Por lo tanto, la molécula de ARN de la presente invención se parece en su estructura a una molécula de ARNm "normal" que se produce en la naturaleza, que alberga una región de codificación, así como regiones no traducidas (UTR) (5' y 3') y, opcionalmente, una cola de poli-A. La expresión "región de codificación", como se usa de acuerdo con la presente invención, se refiere a una molécula de ARN polimérico que se compone de codones, que son descodificados y traducidos a proteínas por el ribosoma de acuerdo con la información proporcionada por el "código genético". Las regiones de codificación comúnmente comienzan con un codón de inicio y terminan con un codón de parada. En general, el codón de inicio es un triplete AUG y el codón de parada es UAA, UAG o UGA. Además de codificar proteínas, hay partes de las regiones de codificación que pueden servir como secuencias reguladoras en el pre-ARNm como potenciadores de corte y empalme exónico o silenciadores del corte y empalme exónico. La región de codificación de un gen que codifica un polipéptido o una proteína, como se usa de acuerdo con la presente invención, también se conoce como la secuencia de codificación o CDS (de la secuencia de ADN de codificación) y es la parte del ADN o ARN de un gen, compuesta de exones, que codifica un polipéptido o una proteína. Como se ha mencionado, la región está limitada más cerca del extremo 5' por un codón de inicio y más cerca del extremo 3' con un codón de parada. La región de codificación del ARNm está flanqueada por las cinco regiones no traducidas principales (UTR 5') y las tres regiones no traducidas principales (UTR 3'), que también forman parte de los exones. La región de codificación o CDS es esa parte de la transcripción de ARNm, es decir, de la región de codificación que codifica un polipéptido, como se usa de acuerdo con la presente invención, que es traducida por un ribosoma en un polipéptido o una proteína.

La expresión "región no traducida" o "UTR", como se usa de acuerdo con la presente invención, se refiere a secciones del ARNm cadena arriba del codón de inicio y cadena abajo del codón de parada que no están traducidas y se denominan, por lo tanto, las cinco regiones no traducidas principales (UTR 5') y tres regiones no traducidas principales (UTR 3'), respectivamente. Estas regiones se transcriben con la región de codificación y, por lo tanto, son exónicas, ya que están presentes en el ARNm maduro. Como se usa en la presente invención, la región no traducida 3' (UTR 3') se refiere a la sección de ARN mensajero (ARNm) que sigue inmediatamente al codón de terminación de la traducción. Una molécula de ARNm se transcribe a partir de la secuencia de ADN y luego se traduce en proteína. Varias regiones de la molécula de ARNm no se traducen a proteínas, incluyendo el capuchón 5', la UTR 5', la UTR 3' y la cola de poli-A. Como se usa en la presente invención, la región 5' no traducida (UTR 5') (también conocida como secuencia líder o ARN líder) es la región de un ARNm que está directamente cadena arriba del codón de inicio. La UTR 5' comienza en el sitio de inicio de la transcripción y termina un nucleótido (nt) antes del codón de inicio (normalmente, AUG) de la región de codificación. En procariotas, la longitud de la UTR 5' tiende a ser de 3 a 10 nucleótidos, mientras que, en eucariotas, tiende a ser más larga, en general, de 100 a varios miles de nucleótidos de longitud, pero a veces también aparecen UTR más cortas en eucariotas. Como se usa en la presente invención, la UTR 3' puede comprender regiones reguladoras dentro de la región no traducida 3' que se sabe que influyen en la poliadenilación y en la estabilidad del ARNm. Muchas UTR 3' también contienen elementos ricos en AU (ARE). Además, la UTR 3' contiene la secuencia AAUAAA que dirige la adición de varios cientos de restos de adenina denominados cola de poli (A) al final de la transcripción de ARNm.

Como se describirá con más detalle más adelante, una molécula de ARN, como se usa de acuerdo con la presente invención, también puede contener una cola de poli-A. Una cola de poli-A es una larga secuencia de nucleótidos de adenina (a menudo de varios cientos) añadida al extremo 3' del pre-ARNm mediante un proceso denominado poliadenilación. Esta cola potencia la exportación desde el núcleo y la traducción, y protege el ARNm de 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; en otras palabras, es un tramo de ARN que solo tiene bases de adenina. En eucariotas, la poliadenilación es parte del proceso que produce el ARN mensajero maduro (ARNm) para la traducción.

Como se ha mencionado anteriormente, la molécula de ARN de la presente invención preferentemente comprende dos o tres módulos principales, es decir, (a) una región de codificación que codifica un polipéptido; y (b) cadena arriba de dicha región de codificación, una o más UTR que comprenden la secuencia mostrada en SEQ ID NO: 1 o una secuencia que muestra de 1 a 4 sustituciones en comparación con SEQ ID NO: 1 y que da lugar a una molécula de ARN que tiene la misma o mayor eficacia de traducción que una molécula de ARN que comprende una UTR que comprende la SEQ ID NO: 1; y/o (c) cadena abajo de dicha región de codificación, una o más UTR que comprenden la secuencia mostrada en SEQ ID NO: 2 o una secuencia que muestra de 1 a 7 sustituciones en comparación con SEQ ID NO: 2 y que da lugar a una molécula de ARN que tiene la misma o mayor eficacia de traducción que una molécula de ARN que comprende una UTR que comprende la SEQ ID NO: 2.

Por lo tanto, es obligatorio que la molécula de ARN de la presente invención comprenda dos módulos principales, es decir, el módulo (a) anterior y el módulo (b) y, opcionalmente, también el módulo (c). En otra realización preferida, la molécula de ARN de la presente invención comprende tres módulos principales, es decir, el módulo (a) anterior y el módulo (b) y el módulo (c). Sin embargo, aunque el módulo (a) es obligatorio, también se prevé que la molécula de ARN también pueda carecer de uno de los módulos (b) o (c).

Un módulo de la molécula de ARN, es decir, "una región de codificación que codifica un polipéptido" (módulo (a)) no

se limita en particular, y puede ser cualquier región de codificación deseada que se exprese en una célula dada. Por lo tanto, este módulo puede ser una región de codificación que codifique un polipéptido deseado, es decir, el producto final deseado. La presente invención no está limitada con respecto a la "región de codificación que codifica un polipéptido", ya que la naturaleza de la región de codificación depende del producto deseado que se vaya a producir en la célula. Dicha región de codificación también puede ser una secuencia de nucleótidos que difiere de una secuencia natural conocida y contiene mutaciones (es decir, mutaciones puntuales, mutación de inserción, eliminaciones y combinaciones de las mismas). Además, dicha región de codificación puede ser parcialmente o en toda su extensión una secuencia con optimización de codones derivada de la secuencia natural que se usará como módulo (a). La optimización de codones es una técnica para aumentar al máximo la expresión de proteínas aumentando la eficacia de la traducción de un gen de interés. Se sabe que los genes naturales no usan los codones disponibles al azar, sino que muestran una cierta preferencia por determinados codones para el mismo aminoácido. Por lo tanto, debido a la degeneración del código genético (un aminoácido puede estar codificado 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. Sin embargo, se excluye la región de codificación (módulo (a)) del gen del polipéptido alfa del citocromo b-245 (CYBA) y, por consiguiente, la molécula de ARN de la presente invención es una molécula de ARN que comprende el módulo (a), es decir, una región de codificación que codifica un polipéptido en la que, sin embargo, dicha región de codificación que codifica un polipéptido en (a) no es una región de codificación que codifica un polipéptido alfa del citocromo b-245 (CYBA). Las regiones de codificación que codifican un polipéptido alfa del citocromo b-245 (CYBA), así como las secuencias de aminoácidos correspondientes son conocidas en la técnica. Se sabe que los polipéptidos alfa del citocromo b-245 son capaces de producir superóxido, y se sabe que participan en la fagocitosis. En la SEQ ID NO: 9, se muestra un ejemplo de una región de codificación que codifica un polipéptido alfa del citocromo b-245 (CYBA). Por tanto, en una realización preferida, la molécula de ARN de la presente invención es una molécula de ARN que comprende el módulo (a), es decir, una región de codificación que codifica un polipéptido, en la que dicha región de codificación que codifica un polipéptido en (a) no es una región de codificación que codifica un polipéptido alfa del citocromo b-245 (CYBA) como se muestra en SEQ ID NO: 9 o una región de codificación que muestra una secuencia de aminoácidos que es al menos un x % idéntica a SEQ ID NO: 9, siendo x un número entero entre 90 y 100, preferentemente 95, 96, 97, 98 o 99. Como ejemplo, en el ADN, una secuencia que representa la región de codificación de un polipéptido que codifica un polipéptido alfa del citocromo b-245 (CYBA) se muestra en la SEQ ID NO: 8.

Como se ha mencionado, el módulo (a) no se limita en particular, y puede ser cualquier región de codificación deseada que deba expresarse en una determinada célula. Por lo tanto, en el contexto de la presente invención, debe entenderse que una "región de codificación" significa cualquier molécula de polirribonucleótido que, si se introduce en una célula, se puede traducir a un polipéptido/una proteína o un fragmento de los mismos. Los términos "polipéptido" y "proteína", en el presente documento, abarcan cualquier clase de secuencia de aminoácidos, es decir, cadenas de dos o más aminoácidos que están unidas cada una a través de enlaces peptídicos, y también incluyen péptidos y proteínas de fusión. En una realización preferida, la "región de codificación que codifica un polipéptido" contiene una secuencia ribonucleotídica que codifica un polipéptido/una proteína o un fragmento de los mismos, cuya función en la célula o en el entorno de la célula es necesaria o beneficiosa, por ejemplo, una proteína que falta o una forma defectuosa a partir de la que se desencadena una enfermedad o dolencia, cuya provisión puede moderar o prevenir una enfermedad o dolencia, o una proteína que puede potenciar un proceso beneficioso para el organismo, en una célula o su entorno. La región de codificación puede contener la secuencia para la proteína completa o una variante funcional de la misma. Además, la secuencia de ribonucleótidos de la región de codificación puede codificar una proteína que actúe como un factor, inductor, regulador, estimulador o enzima, o un fragmento funcional de los mismos, siendo esta proteína aquella cuya función es necesaria para remediar un trastorno, en particular, un trastorno metabólico o para iniciar procesos *in vivo*, como la formación de nuevos vasos sanguíneos, tejidos, etc. En el presente documento, por variante funcional se entiende un fragmento que, en la célula, puede llevar a cabo la función de la proteína cuya función en la célula es necesaria, o cuya falta o forma defectuosa es patógena. En una realización preferida, la "región de codificación que codifica un polipéptido" codifica un polipéptido o una proteína terapéutica o farmacéuticamente activos, que tienen un efecto terapéutico o preventivo. Como tal, la molécula de ARN de la presente invención que comprende dicha "región de codificación que codifica un polipéptido" puede usarse en terapia de ácido nucleico y aplicaciones relacionadas. En este contexto, de acuerdo con la invención, una mayor eficacia de la traducción de una región de codificación de una molécula de ARN en un polipéptido o una proteína codificados por dicha región de codificación de una molécula de ARN exógeno introducida puede pretender compensar o complementar la expresión génica endógena, en particular, en casos en los que un gen endógeno sea defectuoso o silencioso, lo que conduce a un producto nulo, insuficiente o defectuoso o disfuncional de la expresión génica, como es el caso de muchas enfermedades metabólicas y hereditarias como la fibrosis quística, la hemofilia o la distrofia muscular, por nombrar algunas. Una mayor eficacia de la traducción de una región de codificación de una molécula de ARN en un polipéptido de moléculas de ARN exógeno introducidas de la presente invención también puede pretender que el producto de la expresión interactúe o interfiera con cualquier proceso celular endógeno tal como la regulación de la expresión génica, la transducción de señales y otros procesos celulares. La mayor eficacia de la traducción de una región de codificación de una molécula de ARN a un polipéptido de moléculas de ARN exógeno introducidas también puede pretender dar lugar a una respuesta inmune en el contexto del organismo en el que reside o se hace residir una célula transfectada o transducida. Los ejemplos son la modificación genética de células presentadoras de antígeno tales como células dendríticas para que presenten un antígeno con fines de vacunación. Otro ejemplo es la mayor eficacia de la traducción de una región de

codificación de una molécula de ARN en un polipéptido, en el que dicha región de codificación codifica citocinas. Esto puede, por ejemplo, ser deseable en tumores con el fin de provocar una respuesta inmune específica del tumor. Además, la mayor eficacia de la traducción de una región de codificación de una molécula de ARN en un polipéptido de una molécula de ARN exógeno también puede estar destinada a generar células genéticamente modificadas *in vivo* o *ex vivo* transitoriamente para terapias celulares tales como linfocitos T modificados o células precursoras o células madre u otras células para medicina regenerativa. En otras realizaciones preferidas, la "región de codificación que codifica un polipéptido" puede codificar proteínas que desempeñen un papel en los procesos de crecimiento y angiogénesis, que son necesarios, por ejemplo, en la regeneración controlada, y luego pueden formarse específicamente mediante la introducción de la molécula de ARN de acuerdo con 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 la implantación y el trasplante. Como se menciona, la molécula de ARN de la presente invención que comprende una "región de codificación que codifica un polipéptido" puede usarse apropiadamente en cualquier caso en el que un polipéptido o una proteína, que estarían presente de forma natural en el cuerpo, pero no están presente o están presentes de forma deficiente o en una cantidad demasiado baja debido a defectos genéticos o enfermedades, se debe proporcionar al cuerpo. Las proteínas y los genes que las codifican, cuya deficiencia o defecto están relacionados con una enfermedad, son conocidos. La respectiva versión intacta de la región de codificación que codifica el polipéptido o la proteína intactos se puede usar de acuerdo con la presente invención.

Numerosos trastornos genéticos, causados por la mutación de un solo gen son conocidos y candidatos para las metodologías terapéuticas con ARNm. Los trastornos causados por mutaciones de un solo gen, como la fibrosis quística, la hemofilia y muchos otros, pueden ser dominantes o recesivos con respecto a la probabilidad de que un determinado rasgo aparezca en la descendencia. Mientras que un alelo dominante manifiesta un fenotipo en individuos que tienen solo una copia del alelo, para un alelo recesivo, el individuo debe tener dos copias, una de cada padre para manifestarse. Por el contrario, los trastornos poligénicos están causados por dos o más genes, y la manifestación de la enfermedad respectiva suele ser fluida y estar asociada a factores ambientales. Son ejemplos de trastornos poligénicos la hipertensión, el nivel elevado de colesterol, el cáncer, los trastornos neurodegenerativos, las enfermedades mentales y otros. También en estos casos, el ARNm terapéutico que representa uno o más de estos genes puede ser beneficioso para esos pacientes. Además, un trastorno genético no debe haber sido transmitido desde los genes de los padres, sino que también puede estar causado por nuevas mutaciones. También en estos casos, el ARNm terapéutico que representa la secuencia genética correcta puede ser beneficioso para los pacientes. Un catálogo en línea con actualmente 22.993 entradas de Genes Humanos y Trastornos Genéticos junto con sus respectivos genes y una descripción de sus fenotipos se encuentra disponible en la página web de ONIM (Herencia Mendeliana Online en el Hombre) (<http://onim.org>); estando cada una de las secuencias disponibles en la base de datos de Uniprot (<http://www.uniprot.org>). Como ejemplos no limitantes, la siguiente **Tabla 2** enumera algunas enfermedades congénitas y el/los gen/es correspondiente/s. Debido al alto grado de interacción de las vías de señalización celular, la mutación de un determinado gen causa una multiplicidad de síntomas patógenos, de los que solo uno característico se enumera en la **Tabla 2**.

En algunas realizaciones de la presente invención, la proteína terapéutica se escoge entre las proteínas celulares enumeradas en la **Tabla 2**. De este modo, las composiciones de la invención pueden comprender un ARNm que codifica una proteína celular terapéutica, siendo la proteína terapéutica codificada una enumerada en la **Tabla 2** o un homólogo de la misma.

En otra realización de la presente invención, la proteína terapéutica se escoge entre las proteínas secretadas enumeradas en la **Tabla 2**. Por lo tanto, las composiciones de la invención pueden comprender un ARNm que codifica una proteína de fusión terapéutica, siendo la proteína terapéutica codificada, o un homólogo de la misma, una enumerada en la **Tabla 2**, y siendo la segunda proteína un péptido señal que permita la secreción de la proteína terapéutica. Un péptido señal es una secuencia de aminoácidos corta, normalmente de 5 a 30 aminoácidos de longitud, presente en el extremo N de dicha proteína terapéutica, y que dirige la proteína de fusión hacia la vía secretora de la célula a través de ciertos orgánulos (es decir, el retículo endoplásmico, el aparato de golgi o los endosomas). Por lo tanto, dicha proteína de fusión se secreta desde la célula o desde un orgánulo celular o se inserta en una membrana celular (por ejemplo, proteínas transmembrana) en un compartimento celular o en la superficie de la célula.

Por lo tanto, en realizaciones preferidas de la presente invención, la "región de codificación que codifica un polipéptido" (módulo (a)) puede codificar, pero sin limitación, los siguientes genes que causan, predisponen o protegen de enfermedades. Los ejemplos no limitantes de dichos trastornos que pueden tratarse (o prevenirse) incluyen aquellos en los que dicho polipéptido, proteína o péptido se selecciona del grupo que consiste en los que se describen en la siguiente **Tabla 2**. En algunas realizaciones, la "región de codificación que codifica un polipéptido" puede traducirse en una proteína de longitud parcial o completa que comprende actividad celular a un nivel igual o mayor que el de la proteína nativa. En algunas realizaciones, la "región de codificación que codifica un polipéptido" codifica un polipéptido, una proteína o un péptido terapéuticamente o farmacéuticamente activos que tienen un efecto terapéutico o preventivo, siendo dicho polipéptido, proteína o péptido seleccionado del grupo que consiste en los que se describen en la siguiente **Tabla 2**. La "región de codificación que codifica 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 se puede indicar la administración

de una molécula de ARN.

Tabla 2: Ejemplos no limitantes de genes humanos y trastornos genéticos

Enfermedad	Patología	Gen, herencia
Enfermedades de la sangre		
Anemia de Fanconi	Anemia y neutropenia, evidencia de que un mecanismo de reparación del ADN está afectado	FANCA, autosómica recesiva
Hemofilia A	Sangrado anormal	Factor de coagulación VIII, recesiva ligada al cromosoma X
Hemofilia B	Sangrado anormal	Factor de coagulación IX, recesiva ligada al cromosoma X
Esferocitosis hereditaria (varios tipos)	eritrocitos esféricos (esferocitos)	Anquirina (ANK1)
Hemoglobinuria paroxística nocturna	Anemia y presencia de sangre en la orina	PIG-A, cromosoma X
Porfiria cutánea tarda	Sobreproducción de hemo, sobrecarga de hierro	Uroporfirinógeno descarboxilasa (UROD), autosómica recesiva
Inmunodeficiencia combinada grave (SCID)	Debida a la alteración de la síntesis de ADN, inmunodeficiencia grave en la inmunidad humoral y celular	Adenosina desaminasa, autosómica recesiva, IL-2R- γ , JAK3, (IL-7R- α , RAG1/2, Artemis, CD3 δ , CD3 ϵ
Anemia falciforme	Hemoglobina anormal (HbS)	β - hemoglobina (HB), autosómica recesiva
Talasemia (forma α y β)	Falta de hemoglobina α o β que provoca anemia	Eliminación de HBA1 y/o HBA2,
Enfermedad de Von Willebrand (tres tipos conocidos, el Tipo III es más grave)	Sangrado anormal, hemorragia similar a la hemofilia A y B	Formas autosómica dominante y recesiva
Cáncer		
Melanoma maligno	La mutación P16 conduce a una proliferación descontrolada de los fibroblastos	Inhibidor de la quinasa dependiente de la ciclina 2 (CDKN2)
Neurofibromatosis (2 tipos)	Los tumores benignos en los nervios auditivos conducen a la sordera	NF1, NF2, autosómica dominante
Sordera (Oído)		
Sordera	Pérdida de la audición	Sordera 1A (DFNB1), autosómica recesiva
Síndrome de Pendred	Pérdida de la audición	Pendrina (PDS), autosómica recesivo
Corazón		
Ataxia telangiectasia	Reparación defectuosa del daño del ADN,	ATM,

Enfermedad	Patología	Gen, herencia
Corazón		
Aterosclerosis	Aumento del colesterol en sangre	apoE,
Síndrome de LQT (QT largo)	Defecto del canal de potasio	LQT1 y otros genes
Síndrome de Von-Hippel Lindau	El crecimiento anormal de los vasos sanguíneos puede provocar cáncer	VHL, autosómica dominante
Síndrome de William Beuren	La eliminación de elastina produce defectos vasculares, estenosis aórtica supra valvular	Eliminación de los genes elastina y LIM quinasa
Trastornos metabólicos y enfermedades de almacenamiento de glicógeno		
Adrenoleucodistrofia	Transporte y metabolismo defectuoso de los ácidos grasos	ABCD1, cromosoma X
Alcaptonuria	Defecto del metabolismo del nitrógeno, la orina se oscurece cuando se expone al oxígeno	Oxidasa homogentísica, autosómica recesiva
Diabetes de tipo I	Producción defectuosa de la insulina	IDDM1, IDDM2, GCK, etc.
Galactosemia	Trastorno del metabolismo de la galactosa	Gen de galactosa-1-fosfato uridiltransferasa (GALT), autosómica recesiva
Enfermedad de Gauche	Alteración del metabolismo de las grasas	Glucocerebrosidasa
Mala absorción de la glucosa galactosidasa	Transporte alterado de la glucosa y la galactosa fuera de la luz intestinal que produce diarrea	SGLT1, autosómica recesiva
Enfermedad de almacenamiento del glucógeno de tipo I, enfermedad de Von-Gierke	Acumulación de glucosa en hígado y riñón	Glucosa-6-fosfatasa, autosómica recesiva
Enfermedad de almacenamiento del glucógeno de tipo II, enfermedad de Pompe	Acumulación de glicógeno en hígado, corazón, músculo esquelético, cardiomegalia	α -1-Glucosidasa, autosómica recesiva
Enfermedad de almacenamiento del glucógeno de tipo III, enfermedad de Cori	Acumulación de glicógeno en hígado, corazón, músculo esquelético, hepatomegalia	Enzima de desramificación, autosómica recesiva
Enfermedad de almacenamiento del glucógeno de tipo V, enfermedad de McArdle	No se puede utilizar el glucógeno en las células musculares	Fosforilasa muscular, autosómica recesiva
Glucosa-6-Fosfato Deshidrogenasa	La incapacidad para mantener el glutatión conduce a la anemia hemolítica	G6PD, recesiva ligada al cromosoma X

Enfermedad	Patología	Gen, herencia
Trastornos metabólicos y enfermedades de almacenamiento de glicógeno		
Hemocromatosis hereditaria (4 tipos)	Exceso de hierro en el organismo (especialmente en el hígado) debido a la absorción excesiva de hierro en el intestino	Hemocromatosis (HFE)
Homocistinuria	Defecto del metabolismo del nitrógeno	Defecto de cistationa sintetasa, autosómica recesiva
Síndrome de Lesh Nyhan	Acumulación de ácido úrico que conduce a gota, piedras en la uretra y pérdida de masa muscular	HPRT1, cromosoma X
Enfermedad de la orina de jarabe de arce	El defecto del metabolismo de los aminoácidos conduce a la acumulación de α -cetoácidos y a la muerte en los primeros meses si no se trata	Alfa-deshidrogenasa de cadena ramificada (BCKDH)
Síndrome de Menkes	Capacidad reducida para absorber el cobre, conduce a la muerte en la infancia si no se trata	cromosoma X; ATP7A, recesivo ligado al cromosoma X
Obesidad	Peso corporal elevado	Los niveles elevados de leptina poligénicos pueden desempeñar un papel
Fenilcetonuria	La incapacidad de descomponer la fenilalanina en tirosina conduce a un retraso mental	Fenilalanina hidroxilasa (HAP), autosómica recesiva
Enfermedad de Tánger	niveles reducidos de lipoproteínas de alta densidad plasmáticas	Gen de casete 1 de unión a ATP (ABCA1)
Síndrome de Zellweger (conduce a la muerte en bebés)	Altos niveles de hierro y cobre en sangre	PXR1 (receptor en la superficie de los peroxisomas)
Enfermedad de Wilson	Acumulación de cobre en el cerebro y en el hígado	ATP7B (ATPasa de tipo P), autosómica recesiva
Sistema musculoesquelético		
Acondroplasia	Baja estatura con una gran cabeza debido a la lenta proliferación de los condrocitos	Receptor 3 del factor de crecimiento fibroblástico (FGF3R),
Síndrome de Charcot-Marie-Tooth y su forma más grave, el Síndrome de Dejerine-Sottas	Degeneración de los músculos en las extremidades	Diferentes formas causadas por diferentes mutaciones génicas, autosómico recesivo y ligado al cromosoma X
Síndrome de Cockayne (2 tipos)	Envejecimiento prematuro y baja estatura, pérdida de la reparación del ADN "sobre la marcha"	Proteína trans-complementaria de la reparación de la escisión del grupo 8 (ERCC8)

Enfermedad	Patología	Gen, herencia
Sistema musculoesquelético		
Displasia condroctodérmica	Malformación de huesos y polidactilia	EVC, autosómica recesiva
Displasia diastrófica (DTD)	Manos deformadas, defecto del transportador de sulfato	Gen DTDST
Distrofia muscular de Duchenne	Agrandamiento del tejido muscular con la consiguiente pérdida de función	DMD, recesiva ligada al cromosoma X
Fibrodisplasia osificante progresiva	Formación de hueso heterotópico	NOG, BMP, autosómica dominante
Ataxia de Friedreich	Agrandamiento del corazón y pérdida progresiva de la coordinación muscular	Frataxina, autosómica recesiva
Hipofosfatasa	Producción de una versión anormal de fosfatasa alcalina que afecta el proceso de mineralización	ALPL, autosómica recesiva
Síndrome de Marfan	Trastorno del tejido conjuntivo por deficiencia de fibrilina	Fibrilina 1 (FBN), autosómico dominante
Distrofia miotónica (aparición durante la edad adulta)	Defecto de la proteína quinasa en las células del músculo esquelético	Proteína quinasa de distrofia miotónica (DMPK), autosómica dominante
Osteogénesis imperfecta (varios tipos)	El defecto en la formación de colágeno de tipo I conduce a múltiples fracturas después del nacimiento	COL1A1, COL1A2
Síndrome de Prader-Willi	Disminución del tono muscular y retraso mental	SNRPN (ribonucleoproteína N pequeña) eliminada debido a una eliminación en el cromosoma 15
Neuronas y cerebro		
Enfermedad de Alzheimer	Aumento de la producción de amiloide, incapacidad progresiva para recordar hechos	poligénica, PS1, PS2, etc.
Esclerosis lateral amiotrófica (ELA) (varias formas)	Degeneración progresiva de las neuronas motoras (defecto en la eliminación de radicales de superóxido)	Superóxido dismutasa 1 (SOD1), varios genes implicados
Síndrome de Angelman	Retraso mental con una risa inadecuada	Impresión genómica en el cromosoma 15
Piruvato deshidrogenasa	Defectos neurológicos si no se trata	Piruvato deshidrogenasa, autosómica recesiva

Enfermedad	Patología	Gen, herencia
Neuronas y cerebro		
Enfermedad de Refsum	La acumulación de ácido fitánico conduce a neuropatía periférica	Fitanoil-CoA hidroxilasa (PHYH), autosómica recesiva
Síndrome de Rett	Retraso mental con detención del desarrollo entre los 6 y 18 meses de edad	Proteína 2 de unión a metil-CpG (MECP2), dominante ligada al cromosoma X
Enfermedad de Tay-Sachs (varias formas de severidad)	La descomposición afectada del gangliósido GM2 conduce a daño neurológico	HEXA (β -hexosaminidasa A), autosómica recesiva
Enfermedad de LaFora	Forma agresiva de epilepsia	EPM2A, autosómica recesiva
Temblor esencial (formas variables)	Agitación incontrolable	ETM1, ETM2, autosómico dominante
Síndrome X frágil	Falta de proteína de unión al ARN de FMR1, retraso mental	El gen FMR1 no se expresa debido a una amplificación de CGG en la región UTR 5'
Enfermedad de Huntington	Demencia progresiva con inicio en la edad adulta	HTT (huntingtina), autosómica dominante
Intestino		
Síndrome de Bartter (3 tipos)	Enfermedad renal	Gen del canal B del cloruro renal (CLCNKB), autosómico recesivo
Enfermedad renal poliquística (2 tipos)	Enfermedad renal	PDK1, PDK2, autosómica dominante, también hay una forma autosómica recesiva conocida (ARPKD)
Pulmón		
Alfa-1-antitripsina	Alvéolos defectuosos debido a la liberación incontrolada de elastasa	SERPINA1, codominante autosómica
Asma	Trastorno inflamatorio crónico de las vías respiratorias	Poligénico
Fibrosis quística	Mucosidad excesivamente viscosa debido al transporte defectuoso de iones de Cl ⁻	CFTR (regulador transmembrana de la conductancia de la fibrosis quística), autosómica recesiva
Disfunción del metabolismo de tensoactivos (varios tipos)	Los recién nacidos tienen un peso corporal normal, pero no se hinchan	Transportador de casetes de unión a ATP (ABCA3)
Discinesia clínica primaria	Mucosidad excesivamente viscosa debido a una función ciliar defectuosa/carente	CCNO, CCDC40 entre otros

Enfermedad	Patología	Gen, herencia
Enfermedades de almacenamiento lisosomal		
Enfermedad de Fabry	Además de otras, lesiones en la piel debido a la acumulación de ceramida trihexosidasa	α Galactosidasa A, recesiva ligada al cromosoma X
Enfermedad de Gaucher de tipo I: forma adulta (vida normal en tratamiento); Tipo II: forma infantil (muerte antes de cumplir 1 año) Tipo III: forma juvenil (inicio en la primera infancia, menos grave que el tipo II)	Acumulación de glucocerebrósidos (gangliósidos, esfingolípidos)	Glucocerebrosidasa, autosómica recesiva,
Síndrome de Hunter	Acumulación de mucopolisacáridos	L-iduronosulfato sulfatasa, recesivo ligado al cromosoma X
Síndrome de Hurler (muerte a la edad de 10 años)	Acumulación de mucopolisacáridos	α -L-iduronidasa, autosómico recesivo
Enfermedad de Niemann-Pick (tres formas distintas A, B, C)	Defecto al liberar el colesterol de los lisosomas, acumulación de esfingomielina	Esfingomielinasa, autosómica recesiva
Enfermedad de Tay-Sachs (muerte a la edad de 4 años)	Acumulación de gangliósido G_{M2} en células neuronales	Hexosaminidasa A, autosómica recesiva
Piel		
Albinismo	Defecto del metabolismo del nitrógeno	Deficiencia de tirosinasa, autosómico recesivo
Albinismo, oculocutáneo, tipo II	Biosíntesis reducida del pigmento de melanina	OCA2, autosómico recesivo
Síndrome de Ehlers-Danlos (varios tipos)	Hernia de diafragma, desprendimiento de retina común	Varios defectos en la síntesis del colágeno
Epidermólisis ampollosa (varios tipos incluyendo EB simple, EB juntural, EB distrófica y síndrome de Kindler)	Defectos en el mantenimiento de la estabilidad estructural de los queratinocitos o la adhesión del queratinocito a la dermis subyacente	Epidermólisis ampollosa de tipo macular (EBM), Epidermólisis ampollosa 3 progresiva (EBR3), Epidermólisis ampollosa 4 pseudojuntural (EBR4), Desmoplaquina (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 fotosensible	SLC6A19, autosómica recesiva
Telangiectasia hemorrágica hereditaria, síndrome de Osler-Weber-Rendu	Telangiectasia de la piel y las membranas mucosas	Endogлина (ENG), autosómica dominante

Enfermedad	Patología	Gen, herencia
Piel		
Hipercolesterolemia, familiar	Elevación del colesterol en suero unido a lipoproteínas de baja densidad, acumulación en la piel y arterosclerosis	Receptor de lipoproteínas de baja densidad (LDLR), apolipoproteína B (APOB), autosómica dominante
Xeroderma pigmentosa	Defecto de la piel y melanoma debido a la exposición a los rayos UV	Defecto de la reparación del ADN, autosómica recesiva
Calvicie de patrón masculino	Alteración de la conversión de la testosterona en dihidrotestosterona en la piel	5- α -reductasa
Enfermedades genéticas del hígado		
Trastornos del metabolismo de los aminoácidos	Interrupciones en el proceso de múltiples etapas que descompone el aminoácido tirosina y fenilalanina	FAH, TAT, HPD, autosómicos recesivos
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 disuelve en el agua	UGT1A1, autosómico recesivo
Trastornos de la oxidación de los ácidos grasos	Deficiencia en el procesamiento de ácidos grasos de cadena larga y ácidos grasos de cadena muy larga que producen letargo e hipoglucemia	HADHA, ACADVL autosómicos recesivos
Trastorno del metabolismo de la fructosa	Gluconeogénesis alterada que causa hipoglucemia	FBP1, ALDOB, autosómico recesivo
Galactosemia	Deficiencia en el procesamiento de galactosa	GALT, GALK1, GALE, autosómica recesiva
Enfermedades de almacenamiento de glicógeno	La alteración de la descomposición de la glucosa 6-fosfato y el glucógeno conduce a la acumulación de glucógeno, así como a moléculas de glucógeno anormales que causan daño celular	G6PC, SLC37A4, AGL, GBE1, autosómicas recesivas

Enfermedad	Patología	Gen, herencia
Enfermedades genéticas del hígado		
Trastorno de la biosíntesis de Heme	Disminución de la uroporfirinógeno descarboxilasa que da lugar a la acumulación de compuestos denominados porfirinas causantes de niveles tóxicos en el hígado	UROD autosómico dominante, ALAS2 dominante ligado al cromosoma X, ALAD autosómico recesivo
Trastornos del metabolismo lipídico (transporte)	Escasez de proteína funcional, que impide el movimiento del colesterol y otros lípidos, lo que lleva a su acumulación en las células	NPC1, NPC2 autosómico recesivo, LDLR autosómico dominante
Trastornos del metabolismo de los metales	Trastornos en el almacenamiento y transporte del hierro y cobre que producen acumulación en tejidos y órganos	ATP7B, HAMP, HFE, HFE2, autosómicos recesivos
Enfermedades genéticas del hígado		
Trastornos del ácido orgánico (Acidurias/Acidemias)	Alteración de la descomposición de varios constituyentes de proteínas (aminoácidos), ciertos lípidos y colesterol	BCKDHA, BCKDHB y DBT, PCCA y PCCB, MUT, MMAA, MMAB, MMADHC, MCEE, IVD, MCCC1 o MCCC2, autosómicos recesivos
Hiperoxaluria primaria de tipo 1	Alteración de la descomposición del glioxilato que causa daño renal	AGXT, GRHPR, autosómica recesiva
Colestasis intrahepática familiar progresiva	Acumulación de ácidos biliares en las células hepáticas que causan daño hepático	ATP8B1, autosómica recesiva
Trastorno de la actividad trombocítica	La falta de actividad enzimática interrumpe el equilibrio habitual entre sangrado y coagulación	ADAMTS13, autosómico recesivo
Trastornos del ciclo de la urea	Trastorno del ciclo de la urea que causa una forma de hiperamonemia	OTC (trastorno ligado a X), CPS1, ASS1 y SLC25A13, ASL, autosómicos recesivos

La **Tabla 2** anterior muestra ejemplos de genes en los que un defecto conduce a una enfermedad que puede tratarse con la molécula de ARN de la presente invención, en la que la molécula de ARN comprende una "región de codificación que codifica un polipéptido" que codifica una versión intacta de la proteína o un fragmento funcional de la misma del gen defectuoso desvelado anteriormente. En realizaciones particularmente preferidas, caben mencionar enfermedades hereditarias que, por ejemplo, afectan a los pulmones, tales como deficiencia de SPB (proteína tensoactiva B), deficiencia de ABCA3, fibrosis quística y deficiencia de α 1-antitripsina o que afectan proteínas plasmáticas (por ejemplo, hemocromatosis congénita (deficiencia de hepcidina)), púrpura trombocitopénica trombótica (TPP, deficiencia de ADAMTS 13) y causan defectos de coagulación (por ejemplo, hemofilia a y b) y defectos del complemento (por ejemplo, deficiencia de proteína C), defectos inmunes tales como, por ejemplo, SCID (causados por mutaciones en diferentes genes tales como: RAG1, RAG2, JAK3, IL7R, CD45, CD3 δ , CD3 ϵ) o por deficiencias debidas a la falta de adenosina desaminasa, por ejemplo, (ADA-SCID), granulomatosis séptica (por ejemplo, causada por mutaciones del gen gp-91-phox, el p47-phox, el gen p67-phox o el gen p33-phox) y enfermedades de almacenamiento como la enfermedad de Gaucher, enfermedad de Fabry, enfermedad de Krabbe, MPS I, MPS II (síndrome de Hunter), MPS VI, enfermedad de almacenamiento de glucógeno de tipo II o

mucopolisacaidosis. Otros trastornos para los que la presente invención que comprende una "región de codificación que codifica un péptido" puede ser útil incluyen trastornos tales como la atrofia muscular espinal (SMA) relacionada con SMN1; esclerosis lateral amiotrófica (ELA); galactosemia relacionada con GALT; fibrosis quística (CF); trastornos relacionados con SLC3A1 que incluyen cistinuria; trastornos relacionados con COL4A5 que incluyen el

5 síndrome de Alport; deficiencias de galactocerebrosidasa; adrenoleucodistrofia y adrenomieloneuropatía ligadas a 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; trastornos relacionados con FMR1 que incluyen el síndrome de X frágil, síndrome de temblor/ataxia asociado a X frágil y síndrome de insuficiencia ovárica prematura de X frágil; síndrome de Prader-Willi; telangiectasia hemorrágica hereditaria (AT); enfermedad de

10 Niemann-Pick de tipo C1; las enfermedades relacionadas con la lipofuscinosis ceroide neuronal incluyendo la lipofuscinosis ceroide neonatal 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/materia blanca desaparecida; ataxia episódica relacionada con CACNA1A y CACNB4 de tipo 2; los trastornos relacionados con MECP2 que incluyen el síndrome de Rett clásico, encefalopatía neonatal grave relacionada con MECP2 y el Síndrome de PPM-X; síndrome de Rett atípico relacionado con CDKL5; enfermedad de Kennedy (SBMA); arteriopatía cerebral autosómica dominante relacionada con Notch-3 con infartos subcorticales y leucoencefalopatía (CADASIL); trastornos convulsivos relacionados con SCN1A y SCN1B; los trastornos relacionados con la polimerasa G que incluyen el síndrome de Alpers-Huttenlocher, neuropatía atáxica sensorial relacionada con POLG, disartria y

20 oftalmoparesia, y la oftalmoplejía externa progresiva autosómica dominante y recesiva con eliminaciones de ADN mitocondrial; hipoplasia suprarrenal ligada a X; agammaglobulinemia ligada a X; enfermedad de Fabry; y enfermedad de Wilson. En todas estas enfermedades, una proteína, por ejemplo, una enzima, es defectuosa, lo que puede tratarse mediante el tratamiento con el ARN de acuerdo con la invención, que hace que la proteína codificada por el gen defectuoso o un fragmento funcional de la misma esté disponible. Las terapias de reemplazo de

25 transcripción/terapias de reemplazo de enzimas no afectan al defecto genético subyacente, pero aumentan la concentración de la enzima en la que el paciente es deficiente. Como ejemplo, en la enfermedad de Pompe, la terapia de reemplazo de la transcripción/terapia de reemplazo enzimático reemplaza la enzima lisosomal deficiente alfa-glucosidasa ácida (GAA). Por lo tanto, ejemplos no limitantes de proteínas que pueden ser codificadas por la "región de codificación que codifica un polipéptido" del módulo (a) de acuerdo con la invención son eritropoyetina

30 (EPO), hormona del crecimiento (somatotropina, hGH), regulador de la conductancia transmembrana de la fibrosis quística. (CFTR), factores de crecimiento tales como GM-SCF, G-CSF, MPS, proteína C, hepcidina, ABCA3 y proteína tensioactiva B. Otros ejemplos de enfermedades que pueden tratarse con el ARN de acuerdo con la invención son la hemofilia A/B, la enfermedad de Fabry, CGD, ADAMTS13, enfermedad de Hurler, A- γ -globulinemia mediada por el cromosoma X, inmunodeficiencia relacionada con la adenosina desaminasa y síndrome de dificultad respiratoria en el recién nacido, que está relacionado con SP-B. De forma particularmente preferida, la "región de codificación que codifica un polipéptido" de la molécula de ARN de acuerdo con la invención contiene la secuencia de la proteína tensioactiva B (SP-B) o de la eritropoyetina. Otros ejemplos de proteínas que pueden ser codificadas por la "región de codificación que codifica un polipéptido" de la molécula de ARN de acuerdo con la invención son factores de crecimiento tales como la hormona de crecimiento humano hGH, BMP-2 o factores de angiogénesis.

40 Como alternativa, los ácidos nucleicos pueden codificar anticuerpos de longitud completa o anticuerpos más pequeños (por ejemplo, cadenas pesadas y ligeras) para conferir inmunidad a un sujeto. En otra realización, la "región de codificación que codifica un polipéptido" puede codificar un anticuerpo monoclonal o policlonal funcional, que puede ser útil para dirigir y/o inactivar una diana biológica (por ejemplo, una citocina estimulante tal como el factor de necrosis tumoral). De forma similar, la "región de codificación que codifica un polipéptido" puede codificar,

45 por ejemplo, anticuerpos funcionales anti-factor nefrótico útiles para el tratamiento de la glomerulonefritis membranoproliferativa de tipo II o el síndrome urémico hemolítico agudo o, como alternativa, puede codificar anticuerpos del factor de crecimiento endotelial antivascolar (VEGF) útiles para el tratamiento de enfermedades mediadas por VEGF, tales como el cáncer.

50 El módulo (a), es decir, la "región de codificación que incluye un codón de inicio en su extremo 5' que codifica un polipéptido", puede ser una región de codificación que codifica 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 la que el ADN se inserta, elimina o reemplaza en el genoma de un organismo usando nucleasas. Estas nucleasas crean roturas específicas del sitio en lugares deseados del genoma. Las roturas inducidas se reparan mediante la unión de

55 extremos no homóloga o la recombinación homóloga, produciendo mutaciones dirigidas en el genoma, "editando" así el genoma. Las roturas pueden ser roturas monocatenarias o roturas bicatenarias (DSB), mientras que las roturas bicatenarias (DSB) son preferidas. En la técnica, se conocen numerosos sistemas de edición del genoma que utilizan diferentes polipéptidos o proteínas, es decir, por ejemplo, el sistema CRISPR-Cas, meganucleasas, nucleasas de dedos de cinc (ZFN) y nucleasas basadas en efectores de tipo activador de la transcripción (TALEN). Los métodos para la ingeniería del genoma se revisan en *Trends in Biotechnology*, 2013, 31 (7), 397-405.

60

Por tanto, en una realización preferida, la "región de codificación que incluye un codón de inicio en su extremo 5' que codifica un polipéptido" contiene una secuencia de nucleótidos que codifica un polipéptido o una proteína de la familia de proteínas Cas (proteína asociada a CRISPR), preferentemente Cas9 (proteína 9 asociada a CRISPR). Las

65 proteínas de la familia de proteínas Cas, preferentemente Cas9, se pueden usar en métodos basados en CRISPR/Cas9 y/o tecnologías de edición del genoma de CRISPR/Cas9. Los sistemas de CRISPR-Cas para la

edición, regulación y dirección del genoma se revisan en *Nat. Biotechnol.*, 2014, 32(4):347-355. En otra realización preferida, la "región de codificación que incluye un codón de inicio en su extremo 5' que codifica un polipéptido" contiene una secuencia de nucleótidos que codifica una meganucleasa. Las meganucleasas son endodesoxirribonucleasas que, a diferencia de las endodesoxirribonucleasas "convencionales", reconocen un sitio de reconocimiento grande (por ejemplo, una secuencia de ADN bicatenario de 12 a 40 pares de bases). Como resultado de ello, el respectivo sitio aparece solo unas cuantas veces, preferentemente solo una vez, en cualquier genoma dado. Por lo tanto, las meganucleasas se consideran las enzimas de restricción naturales más específicas y, por consiguiente, son herramientas adecuadas en las tecnologías de edición del genoma. En otra realización preferida, la "región de codificación que incluye un codón de inicio en su extremo 5' que codifica un polipéptido" contiene una secuencia de nucleótidos que codifica una nucleasa de dedo de cinc (ZFN). Las ZFN son enzimas de restricción artificiales generadas al fusionar un dominio de unión a ADN de dedo de cinc con un dominio de escisión del ADN. Los dominios de dedo de cinc pueden diseñarse para dirigirse a secuencias de ADN deseadas específicas y esto permite que las nucleasas de dedos de cinc se dirijan a secuencias únicas dentro de genomas complejos. Al aprovechar la maquinaria endógena de reparación del ADN, las ZFN se pueden usar para alterar con precisión el genoma de organismos superiores y, por lo tanto, son herramientas adecuadas en las tecnologías de edición del genoma. En otra realización preferida, la "región de codificación que incluye un codón de inicio en su extremo 5' que codifica un polipéptido" contiene una secuencia de nucleótidos que codifica una nucleasa efectora similar a un activador de la transcripción (TALEN). Las TALEN son enzimas de restricción que pueden modificarse para cortar secuencias específicas de ADN. Las TALEN son proteínas de fusión en las que un dominio de unión a ADN del efector TAL se fusiona con un dominio de escisión de ADN de una nucleasa. Los efectores de tipo activador de la transcripción (TALE) se pueden diseñar para unir prácticamente cualquier secuencia de ADN deseada. Por lo tanto, cuando se combina con una nucleasa, el ADN se puede cortar en ubicaciones específicas deseadas.

El segundo módulo (b) es la una o más UTR que comprenden la secuencia como se muestra en SEQ ID NO: 1 o una secuencia que muestra de 1 a 4 sustituciones en comparación con SEQ ID NO: 1 y que da lugar a una molécula de ARN que tiene la misma o mayor eficacia de traducción que una molécula de ARN que comprende una UTR que comprende la SEQ ID NO: 1. "Uno o más" en este contexto significa que el módulo (b) de la molécula de ARN puede albergar una UTR que comprende la secuencia mostrada en SEQ ID NO: 1 o una secuencia que muestra de 1 a 4 sustituciones en comparación con SEQ ID NO: 1 y que da lugar a una molécula de ARN que tiene la misma o mayor eficacia de traducción que una molécula de ARN que comprende una UTR que comprende la SEQ ID NO: 1 de la presente invención. La molécula de ARN también puede albergar dos, tres o cuatro de estas UTR de la presente invención. Como alternativa, la molécula de ARN también puede albergar cinco o incluso más de estas UTR de la presente invención.

El tercer módulo (c) es la una o más UTR que comprenden la secuencia mostrada en SEQ ID NO: 2 o una secuencia que muestra de 1 a 7 sustituciones en comparación con SEQ ID NO: 2 y que da lugar a una molécula de ARN que tiene la misma o mayor eficacia de traducción que una molécula de ARN que comprende una UTR que comprende la SEQ ID NO: 2 (es decir, el módulo (c) anterior). "Uno o más" en este contexto significa que el módulo (c) de la molécula de ARN puede albergar una UTR que comprende la secuencia mostrada en SEQ ID NO: 2 o una secuencia que muestra de 1 a 7 sustituciones en comparación con SEQ ID NO: 2 y que da lugar a una molécula de ARN que tiene la misma o mayor eficacia de traducción que una molécula de ARN que comprende una UTR que comprende la SEQ ID NO: 2 de la presente invención. La molécula de ARN también puede albergar dos, tres o cuatro de estas UTR de la presente invención. Como alternativa, la molécula de ARN también puede albergar cinco o incluso más de estas UTR de la presente invención.

La secuencia de longitud completa del ARNm del polipéptido alfa del citocromo b-245 humano (CYBA) se conoce en la técnica, y tiene la secuencia que se muestra en la SEQ ID NO: 7. En los ejemplos adjuntos, la secuencia de los nucleótidos 36 a 71 del ARNm del polipéptido alfa del citocromo b-245 humano (CYBA) se ha usado como el fragmento de UTR 5' del ARNm de CYBA (es decir, la secuencia de nucleótidos 5'-CGCGCCUAGCAGUGUCCAGCCGGUUCGUGUGCC-3' (SEQ ID NO: 1)), y la secuencia de los nucleótidos 657 a 723 del ARNm del polipéptido alfa del citocromo b-245 humano (CYBA) se ha usado como la UTR 3' del ARNm de CYBA (es decir, la secuencia de nucleótidos

5'-CCUCGCCCGGACCUGCCUCCCGCCAGGUGCACCC
ACCUGCAAUAAAUGCAGCGAAGCCGGGA-3' (SEQ ID NO:2)).

Sin embargo, las UTR como las usadas en la presente invención no se limitan en particular a la secuencia específica anterior de SEQ ID NO: 1, sino que también pueden ser una secuencia de UTR que comprende una secuencia que muestra de 1 a 4 sustituciones en comparación con SEQ ID NO: 1. Como alternativa, la secuencia de UTR también puede ser una secuencia que comprende una secuencia que muestra de 1 a 3 sustituciones en comparación con la SEQ ID NO: 1. La secuencia de UTR también puede ser una secuencia que comprende una secuencia que muestra de 1 a 2 sustituciones en comparación con la SEQ ID NO: 1. Más preferentemente, la secuencia de UTR también puede ser una secuencia que comprende una secuencia que muestra 1 sustitución en comparación con SEQ ID NO: 1. Preferentemente, la posición de la sustitución de nucleótidos anterior en comparación con SEQ ID NO: 1 se realiza en la posición 32 de secuencia de SEQ ID NO: 1. Preferentemente, el nucleótido "U" de esta posición está

sustituido por una "C". Se prefiere esta sustitución, ya que acerca el elemento Kozak de CYBA, que está (parcialmente) presente en la SEQ ID NO: 1, a la secuencia consenso de Kozak de los vertebrados. La secuencia de consenso de Kozak de los vertebrados tiene la secuencia de GCCRCCAUGG (el codón de inicio está subrayado, mientras que "R" indica cualquier purina) mientras que el elemento Kozak de CYBA tiene la secuencia de GuCGCCAUGG (el codón de inicio está subrayado mientras la desviación la secuencia de consenso de los vertebrados se indica con la letra minúscula "u").

La/s secuencia/s de UTR que tienen una o más de las sustituciones anteriores en comparación con la SEQ ID NO: 1 puede/n dar lugar a una molécula de ARN en la misma capacidad o capacidad similar en términos de eficacia de traducción que una molécula de ARN que comprende una UTR que comprende SEQ ID NO: 1, preferentemente una capacidad superior en términos de la eficacia de traducción como una molécula de ARN que comprende una UTR que comprende la SEQ ID NO: 1. La propiedad/capacidad de una secuencia de UTR modificada dada en comparación con la eficacia de traducción como una molécula de ARN que comprende una UTR que comprende la SEQ ID NO: 1 con respecto a la eficacia de traducción puede ser determinada por un experto en la materia mediante métodos conocidos en la técnica y como se describe en los ejemplos adjuntos. La eficacia de traducción es la velocidad de traducción de ARNm en polipéptidos o proteínas dentro de las células. La eficacia 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 los ribosomas celulares crean proteínas y es bien conocido por el experto. En resumen, en la traducción, el ARN mensajero (ARNm), que se produce mediante la transcripción del ADN, es descodificado por un ribosoma para producir una cadena de aminoácidos específica, o un polipéptido o una proteína. Por lo tanto, la eficacia de la traducción de una molécula de ARN dada que alberga una secuencia de UTR modificada es preferentemente mayor en comparación con una eficacia de traducción del mismo ARN dado, pero que alberga una UTR de SEQ ID NO: 1. Por consiguiente, el número de proteínas o polipéptidos codificados por la región de codificación de la molécula de ARN que alberga una secuencia de UTR modificada que se traduce por ARN por unidad de tiempo es mayor que el número de proteínas o polipéptidos codificados por la región de codificación de la molécula de ARN que alberga una UTR de SEQ ID NO: 1 que se traduce por ARN por unidad de tiempo. En caso de que la eficacia de traducción de una molécula de ARN dada que alberga una secuencia de UTR modificada sea similar o igual en comparación con una eficacia de traducción del mismo ARN dado, pero albergue una UTR de SEQ ID NO 1, la cantidad de proteínas o polipéptidos codificados por la región de codificación de la molécula de ARN que alberga una secuencia de UTR modificada que se traduce por ARN por unidad de tiempo es similar o igual al número de proteínas o polipéptidos codificados por la región de codificación de la molécula de ARN que alberga una UTR de la SEQ ID NO: 1 que se traducen por ARN por unidad de tiempo. La "eficacia de traducción" puede, por ejemplo, determinarse mediante métodos descritos en los ejemplos adjuntos y como se describe a continuación. La eficacia de traducción, en el contexto de la presente invención, es la velocidad de ARNm traducido a proteína dentro de una célula en un cierto punto de tiempo en relación con la cantidad de ARNm que codifica la proteína respectiva en dicha célula en el mismo punto de tiempo. Por lo tanto, la eficacia de traducción es el cociente del ARNm traducido a proteína dentro de una célula en un cierto punto de tiempo y la cantidad de ARNm que codifica la proteína respectiva. Ambos parámetros, es decir, el ARNm traducido a una proteína así como la cantidad de ARNm que codifica la proteína respectiva, se pueden determinar mediante métodos conocidos en la técnica. Como se ha hecho en los ejemplos adjuntos, como ejemplos no limitantes, la cantidad de ARNm traducido a proteína dentro de una célula puede determinarse, por ejemplo, mediante citometría de flujo (FC) mientras que la cantidad de ARNm que codifica la proteína respectiva puede, por ejemplo, medirse por qPCR.

La/s UTR que comprende/n la secuencia mostrada en SEQ ID NO: 1 o una secuencia que muestra de 1 a 4 sustituciones en comparación con SEQ ID NO: 1 y que da lugar a una molécula de ARN que tiene la misma o mayor eficacia de traducción que la molécula de ARN que comprende una UTR que comprende la SEQ ID NO 1 usada en la presente invención no se limita en particular a las secuencias específicas anteriores y a las sustituciones descritas anteriormente, sino que también puede referirse a (una) secuencia/s de UTR que comprende/n una secuencia que muestra (a) la/s adición/es de nucleótido/s en comparación con la SEQ ID NO: 1. La adición de (a) nucleótido/s puede ser flanqueante. Por lo tanto, el/los nucleótido/s adicional/es se puede/n añadir al extremo 3' o 5' de la/s UTR de la presente invención. Los nucleótidos adicionales comprenden cadenas de polinucleótidos de hasta 0 (sin cambios), 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10 nucleótidos, preferentemente de hasta 20 nucleótidos o incluso más preferentemente de hasta 30 nucleótidos. A la luz del razonamiento de que la adición de nucleótidos probablemente no modifique las propiedades funcionales anteriores de las 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 mayor, de hasta 200, 300, 400 o 500 nucleótidos, siempre que estas secuencias tengan una capacidad similar (en términos de la eficacia de traducción descrita anteriormente) a SEQ ID NO: 1, preferentemente eficacia de traducción superior a SEQ ID NO: 1 como se ha definido anteriormente.

Como alternativa o además de estas adiciones flanqueantes de (a) nucleótido/s, la adición de (a) nucleótido/s se puede intercalar. Por lo tanto, el/los nucleótido/s adicional/es puede/n añadirse/insertarse dentro de la secuencia de nucleótidos de la/s UTR de la presente invención. Estas inserciones nucleotídicas comprenden 1, 2 o 3 nucleótidos siempre que estas secuencias tengan una capacidad similar (en términos de la eficacia de traducción descrita anteriormente) a SEQ ID NO: 1, preferentemente una mayor eficacia de traducción a SEQ ID NO: 1 como se ha definido anteriormente.

Las UTR como las usadas en la presente invención no se limitan en particular a la anterior secuencia específica de SEQ ID NO: 1 y a modificaciones de la misma. En cambio, la secuencia específica de SEQ ID NO: 1 y las modificaciones de la misma definen simplemente la región del núcleo 5' de CYBA. Por lo tanto, en una realización preferida, la UTR mostrada en SEQ ID NO: 1 se extiende en el extremo 5' (es decir, cadena arriba) en al menos 1 nucleótido. En otra realización preferida, la UTR como la mostrada en SEQ ID NO: 1 se extiende en el extremo 5' (es decir, cadena arriba) en 1 a 20 nucleótidos. Por lo tanto, en una realización preferida, la secuencia de SEQ ID NO: 1 se extiende en 20 nucleótidos en el extremo 5' (es decir, cadena arriba) como se muestra en la secuencia de nucleótidos de SEQ ID NO: 10 (o la secuencia de ARN correspondiente de SEQ ID NO: 11) con respecto a SEQ ID NO: 1. En otras realizaciones preferidas, la secuencia de SEQ ID NO: 1 se extiende en 18, 15, 13, 10, 7 o 5 nucleótidos en el extremo 5' (es decir, cadena arriba) como se muestra en la secuencia de nucleótidos de SEQ ID NO: 10 (o la secuencia de ARN correspondiente de SEQ ID NO: 11) con respecto a SEQ ID NO: 1. En otras realizaciones preferidas, la secuencia de SEQ ID NO: 1 se extiende en 4, 5 o 2 nucleótidos en el extremo 5' (es decir, cadena arriba) como se muestra en la secuencia de nucleótidos de SEQ ID NO: 10 (o la secuencia de ARN correspondiente de SEQ ID NO: 11) con respecto a SEQ ID NO: 1. En otra realización preferida, la secuencia de SEQ ID NO: 1 se extiende en 1 nucleótido en el extremo 5' (es decir, cadena arriba) como se muestra en la secuencia de nucleótidos de SEQ ID NO: 10 (o la secuencia de ARN correspondiente de SEQ ID NO: 11) con respecto a SEQ ID NO: 1. La SEQ ID NO: 10 es una parte del código genético de la UTR 5' del gen CYBA humano mostrado anteriormente como SEQ ID NO: 5 (según lo definido en el nivel de ADN) mientras que SEQ ID NO: 11 es la secuencia de ARN correspondiente. Estas secuencias de UTR que se extienden en el extremo 5' (es decir, cadena arriba) también se pueden modificar según lo definido en el presente documento anteriormente para SEQ ID NO: 1. Por consiguiente, lo mismo se aplica, cambiando lo que corresponda, a las UTR que se extienden en el extremo 5' como se ha definido anteriormente como se ha expuesto anteriormente en el contexto de la UTR de SEQ ID NO: 1.

Además, las UTR como las usadas en la presente invención tampoco se limitan en particular a la secuencia específica anterior de SEQ ID NO: 2, sino que también pueden ser una secuencia de UTR que comprenda una secuencia que muestre de 1 a 7 sustituciones en comparación con SEQ ID NO: 2. Como alternativa, la secuencia de UTR también puede ser una secuencia que comprenda una secuencia que muestre de 1 a 6 sustituciones en comparación con la SEQ ID NO: 2. La secuencia de UTR también puede ser una secuencia que comprenda una secuencia que muestre de 1 a 5 sustituciones en comparación con la SEQ ID NO: 2. La secuencia de UTR también puede ser una secuencia que comprenda una secuencia que muestre de 1 a 4 sustituciones en comparación con la SEQ ID NO: 2. La secuencia de UTR también puede ser una secuencia que comprenda una secuencia que muestre de 1 a 3 sustituciones en comparación con la SEQ ID NO: 2. La secuencia de UTR también puede ser una secuencia que comprenda una secuencia que muestre de 1 a 2 sustituciones en comparación con la SEQ ID NO: 2. La secuencia de UTR también puede ser una secuencia que comprenda una secuencia que muestre de 1 a 3 sustituciones en comparación con la SEQ ID NO: 2. Lo más preferentemente, la secuencia de UTR también puede ser una secuencia que comprenda una secuencia que muestre 1 sustitución en comparación con la SEQ ID NO: 2.

La/s secuencia/s de UTR que tienen una o más de las sustituciones anteriores en comparación con SEQ ID NO: 2 puede/n dar lugar a una molécula de ARN en la misma capacidad o capacidad similar en términos de eficacia de traducción que una molécula de ARN que comprenda una UTR que comprenda SEQ ID NO: 2, preferentemente una capacidad superior en términos de la eficacia de traducción a una molécula de ARN que comprenda una UTR que comprenda la SEQ ID NO: 2. La propiedad/capacidad de una secuencia de UTR modificada determinada en comparación con la eficacia de traducción como una molécula de ARN que comprende una UTR que comprende la SEQ ID NO: 2 con respecto a la eficacia de traducción puede ser determinada por un experto en la materia mediante métodos conocidos en la técnica y como se describe en los ejemplos adjuntos. La eficacia de traducción es la velocidad de traducción de ARNm a polipéptidos o proteínas dentro de las células. La eficacia 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 los ribosomas celulares crean proteínas, y es bien conocido por el experto. En resumen, en la traducción, el ARN mensajero (ARNm), que se produce mediante la transcripción del ADN, es descodificado por un ribosoma para producir una cadena de aminoácidos específica, o un polipéptido o una proteína. Por lo tanto, la eficacia de traducción de una molécula de ARN dada que alberga una secuencia de UTR modificada es preferentemente mayor en comparación con una eficacia de traducción del mismo ARN dado pero que alberga una UTR de SEQ ID NO: 2. Por consiguiente, el número de proteínas o polipéptidos codificados por la región de codificación de la molécula de ARN que alberga una secuencia de UTR modificada que se traduce por ARN por unidad de tiempo es mayor que el número de proteínas o polipéptidos codificados por la región de codificación de la molécula de ARN que alberga una UTR de SEQ ID NO: 2 que se traduce por ARN por unidad de tiempo. En caso de que la eficacia de traducción de una molécula de ARN dada que alberga una secuencia de UTR modificada sea similar o igual en comparación con una eficacia de traducción del mismo ARN, dado pero que albergue una UTR de SEQ ID NO: 2, el número de proteínas o polipéptidos codificados por la región de codificación de la molécula de ARN que alberga una secuencia de UTR modificada que se traduce por ARN por unidad de tiempo es similar o igual al número de proteínas o polipéptidos codificados por la región de codificación de la molécula de ARN que alberga una UTR de SEQ ID NO: 2 que se traduce por ARN por unidad de tiempo. La "eficacia de traducción" puede, por ejemplo, determinarse mediante métodos descritos en los ejemplos adjuntos y como se describe anteriormente.

La/s UTR que comprende/n la secuencia mostrada en SEQ ID NO: 2 o una secuencia que muestra de 1 a 7 sustituciones en comparación con SEQ ID NO: 2 y que da lugar a una molécula de ARN que tiene la misma o mayor eficacia de traducción que la molécula de ARN que comprende una UTR que comprende la SEQ ID NO: 2 como se usa en la presente invención no se limita en particular a las secuencias específicas anteriores ni a las sustituciones descritas anteriormente, sino que también puede referirse a (una) secuencia/s de UTR que comprenda/n una secuencia que muestre (a) la/s adición/es de nucleótido/s en comparación con la SEQ ID NO: 2. La adición de nucleótido/s puede ser flanqueante o intercalada. Por lo tanto, el/los nucleótido/s adicional/es se puede/n añadir al extremo 3' o 5' de la/s UTR de la presente invención. Como alternativa o además de estos nucleótidos adicionales flanqueantes, el/los nucleótido/s adicional/es también pueden estar dentro de la secuencia de nucleótidos de las UTR de la presente invención. Los nucleótidos adicionales comprenden cadenas de polinucleótidos de hasta 0 (sin cambios), 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10 nucleótidos, preferentemente de hasta 20 nucleótidos o incluso más preferentemente de hasta 30 nucleótidos. A la luz del razonamiento de que la adición de nucleótidos probablemente no modifique las propiedades funcionales anteriores de las 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, de hasta 200, 300, 400 o 500 nucleótidos, siempre que estas secuencias tengan una capacidad similar (en términos de la eficacia de traducción descrita anteriormente) a SEQ ID NO: 2, preferentemente mayor eficacia de traducción que SEQ ID NO: 2 como se ha definido anteriormente.

Las UTR de la presente invención, así como las moléculas de ARN que contienen dichas UTR se pueden generar/sintetizar recombinante (por ejemplo, en un sistema *in vivo* o *in vitro*) o sintéticamente mediante métodos conocidos por los expertos en la técnica. Más concretamente, las UTR de la presente invención y las moléculas de ARN que contienen dichas UTR pueden producirse recombinantemente en sistemas *in vivo* mediante métodos conocidos por los expertos en la materia. Como alternativa, las UTR de la presente invención y las moléculas de ARN que contienen dichas UTR pueden producirse en un sistema *in vitro* usando, por ejemplo, un sistema de transcripción *in vitro*. Los sistemas de transcripción *in vitro* se conocen comúnmente, y normalmente requieren un molde de ADN lineal purificado que contiene una secuencia de ADN "que codifica" el módulo (b) y/o módulo (c) como se detalla más adelante, estando dicha secuencia de ADN bajo el control de un promotor. Además, un sistema de transcripción *in vitro* también requiere comúnmente ribonucleósidos trifosfato, un sistema tampón que incluye DTT e iones de magnesio, y una ARN polimerasa apropiada que proporcione la actividad enzimática para la transcripción *in vitro* de la secuencia de ADN que codifica los módulos (b) y/o (c) en las UTR de la presente invención. Además, las UTR de la presente invención y las moléculas de ARN que contienen dichas UTR se pueden sintetizar químicamente, por ejemplo, mediante síntesis química convencional en un sintetizador automático de secuencias de nucleótidos usando un soporte en fase sólida y técnicas convencionales o mediante síntesis química de las respectivas secuencias de ADN y la posterior transcripción *in vitro* o *in vivo* de las mismas.

De acuerdo con lo anterior, la presente invención proporciona moléculas de ARN/moléculas de ácido polirribonucleico, preferentemente moléculas de ácido polirribonucleico modificadas, en las que un módulo de dicha molécula de ARN, es decir, una región de codificación que incluye un codón de inicio en su extremo 5 (módulo a)), codifica un polipéptido. Los términos ácido nucleico y polinucleótido se usan indistintamente 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 indistintamente y, en ciertas 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 ciertas realizaciones, los polirribonucleótidos comprenden un polímero de nucleótidos en el que más del 60 %, 70 %, 75 %, 80 %, 90 %, más de 95 %, más de 99 % o 100 % de los nucleótidos son ribonucleótidos. Los polirribonucleótidos en los que uno o más nucleótidos son nucleótidos modificados se pueden denominar polirribonucleótidos modificados. Sin embargo, el término polirribonucleótidos puede incluir polirribonucleótidos modificados. La secuencia de las moléculas de ARN/polirribonucleótidos puede derivarse, por ejemplo, de cualquier ácido nucleico adecuado que comprenda la información genética de un gen de interés. Los ejemplos de ácidos nucleicos incluyen ADN genómico, ARN o ADNc de cualquier célula bacteriana o arqueática que comprenda el/los gen/es de interés. Los polinucleótidos pueden derivarse de ácidos nucleicos portadores de genes mutados y polimorfismos. Una molécula de ARN/polirribonucleótido de la presente invención comprende una secuencia que no se limita en particular y puede comprender, como módulo A, cualquier región de codificación deseada que se exprese en una célula dada. En una realización preferida, dicha secuencia puede ser una región de codificación que codifique un polipéptido/una proteína deseado como se ha descrito anteriormente. Preferentemente, en línea con lo anterior, la molécula de ARN/el polirribonucleótido comprende además una secuencia no traducida situada cadena arriba (5') del codón de inicio del módulo A, una secuencia no traducida situada cadena abajo (3') del codón de parada del módulo A o tanto una secuencia no traducida situada cadena arriba (5') del codón de inicio del módulo A como una secuencia no traducida situada cadena abajo (3') del codón de parada 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, en concreto, adenosina, guanosina, citidina y uridina, existen numerosos análogos de cada una de estas nucleobases. Algunas veces, a lo largo del presente documento y en la literatura, estos análogos o moléculas de ARN/polirribonucleótidos que incluyen uno o más de estos análogos, se denominan modificados (por ejemplo, nucleótidos modificados o ribonucleótidos modificados). Algunos análogos difieren de las nucleobases canónicas anteriores, pero aún pueden existir en la naturaleza. Otros análogos no

ocurren de manera natural. Se contempla cualquier tipo de análogo.

En ciertas realizaciones, las moléculas de ARN/polirribonucleótidos de la presente invención comprenden análogos de nucleótidos (por ejemplo, el polirribonucleótido comprende un polirribonucleótido modificado). A continuación, se proporcionan ejemplos de análogos de nucleótidos (por ejemplo, análogos de U, análogos de C, análogos de A, análogos de G). Además, en ciertas realizaciones, una molécula de ARN/polirribonucleótido u otro ácido nucleico de la divulgación también puede comprender (además de o como alternativa) modificaciones en la cadena principal de fosfodiéster o en el enlace entre nucleobases. Los ejemplos de ácidos nucleicos que pueden formar parte o la totalidad de una molécula de ARN/polirribonucleótido de la divulgación incluyen, pero sin limitación, ácidos ribonucleicos (ARN), ácidos desoxirribonucleicos (ADN), ácidos nucleicos de treosa (TNA), ácidos nucleicos de glicol (GNA), ácidos nucleicos peptídicos (PNA), ácidos nucleicos bloqueados (LNA, incluyendo LNA que tiene una configuración beta-D-ribo, alfa-LNA que tiene una configuración alfa-L-ribo (un diastereómero de LNA), 2'-amino-LNA que tiene una funcionalización 2'-amino y 2'-amino-alfa-LNA que tiene una funcionalización 2'-amino) o híbridos de los mismos. En ciertas realizaciones, una modificación puede ser en uno o más nucleósido/s o la cadena principal de la molécula de ácido nucleico/polinucleótido. En ciertas realizaciones, una modificación puede ser tanto en un nucleósido como en un enlace de cadena principal. En ciertas realizaciones, una modificación se puede diseñar en un polinucleótido *in vitro*. En ciertas realizaciones, también se puede sintetizar un ribonucleótido/nucleótido modificado después de la transcripción mediante modificación covalente de los nucleótidos/ribonucleótidos clásicos/naturales. Una molécula de ARN/polirribonucleótido de la presente invención puede ser una molécula de ARN/polirribonucleótido modificado y, en ciertas realizaciones, puede comprender análogos de purinas y/o análogos de pirimidinas. En ciertas realizaciones, una molécula de ARN modificada/polirribonucleótido de la presente invención comprende un análogo de pirimidina, como un análogo de uridina y/o un análogo de citidina. En ciertas realizaciones, una molécula de ARN/polirribonucleótido modificado de la presente invención comprende un análogo de uridina y un análogo de citidina. En ciertas realizaciones, la molécula de ARN/polirribonucleótido modificado no comprende análogos de adenosina y/o análogos de guanosina. En ciertas realizaciones, la molécula de ARN/el polirribonucleótido comprende un solo tipo de análogo de uridina y un solo tipo de análogo de citidina (por ejemplo, un tipo de análogo, no una sola molécula de análogo; el análogo único puede estar presente en cualquiera de varios porcentajes descritos en el presente documento). En otras realizaciones, la molécula de ARN/el polirribonucleótido comprende más de un tipo de análogo de uridina y/o citidina y, opcionalmente, y si está presente, uno o más análogos de adenosina y/o guanosina (o ninguno de ellos o ambos). En algunos casos, se selecciona una uridina modificada (por ejemplo, análogo de uridina) 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'NH₂), 2'-azido-2'-desoxiuridina (U2'N₃) y 2'-fluoro-2'-desoxiuridina (U2'F). En algunos casos, se selecciona una citidina modificada (por ejemplo, análogo de citidina) a partir de 5-metilcitidina, 3-metilcitidina, 2-tio-citidina, 2'-metil-2'-desoxicitidina (C2'm), 2'-amino-2'-desoxicitidina (C2'NH₂), 2'-fluoro-2'-desoxicitidina (C2'F), 5-yodocitidina (I5C), 5-bromocitidina (Br5C) y 2'-azido-2'-deoxicitidina (C2'N₃). Cabe señalar que, cuando se hace referencia a los análogos, lo anterior también se refiere a los análogos en su forma trifosfato 5'. En ciertas realizaciones, el análogo de citidina es 5-yodocitidina y el análogo de uridina es 5-yodouridina. En algunas realizaciones, la molécula de ARN/el polirribonucleótido es una molécula de ARN/polirribonucleótido modificado. En algunos casos, la molécula de ARN/polirribonucleótido modificado es al menos un 25 % más estable en comparación con una molécula de ARN/un polirribonucleótido no modificados (o sin modificar). En algunos casos, la molécula de ARN/el polirribonucleótido modificado puede ser al menos un 30 % más estable, al menos un 35 % más estable, al menos un 40 % más estable, al menos un 45 % más estable, al menos un 50 % más estable, al menos un 55 % más estable, al menos un 60 % más estable, al menos un 65 % más estable, al menos un 70 % más estable, al menos un 75 % más estable, al menos un 80 % más estable, al menos un 85 % más estable, al menos un 90 % más estable o al menos un 95 % más estable en comparación con una molécula de ARN/un polirribonucleótido sin modificar. En ciertas realizaciones, la estabilidad se mide *in vivo*. En ciertas realizaciones, la estabilidad se mide *in vitro*. En ciertas realizaciones, la estabilidad se cuantifica midiendo la semivida del polirribonucleótido. Una molécula de ARN/polirribonucleótido de la presente invención puede tener nucleótidos que se hayan modificado de la misma forma o también una mezcla de diferentes nucleótidos modificados. Los nucleótidos modificados pueden tener modificaciones que se producen de forma natural o no natural en el ARN mensajero. Se puede usar 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 el ARNm. Además, 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 bases o que todas las modificaciones sean modificaciones de azúcar o cualquier mezcla adecuada de las mismas. En algunos casos, la estabilidad de la molécula de ARN/del polirribonucleótido modificado puede optimizarse selectivamente cambiando la naturaleza de las bases modificadas dentro del polirribonucleótido modificado.

60

Tabla 2: Ejemplos no limitantes de análogos de U

Nombre	Modificación de base (posición 5')	Modificación de azúcar (posición 2')	De manera natural en ARNm
5-metiluridina (m5U)	CH ₃	-	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 ₃	Sí
2'-amino-2'-desoxiuridina (U2'NH ₂)	-	NH ₂	No
2'-azido-2'-desoxiuridina (U2'N ₃)	-	N ₃	No
2'-fluoro-2'-desoxiuridina (U2'F)	-	F	No

Tabla 3: Ejemplos no limitantes de los análogos de C

Nombre	Modificación de base (posición 5')	Modificación de azúcar (posición 2')	De manera natural en ARNm
5-metilcitidina (m5C)	CH ₃	-	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 ₃	Sí
2'-amino-2'-deoxicitidina (C2'NH ₂)	-	NH ₂	No
2'-azido-2'-desoxicitidina (C2'N ₃)	-	N ₃	No
2'-fluoro-2'-desoxicitidina (C2'F)	-	F	No

5 **Tabla 4:** Ejemplos no limitantes de análogos de A

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

Nombre	Modificación de base (posición 5')	Modificación de azúcar (posición 2')	De manera natural en ARNm
desoxiadenosina (A2'N3)			
2'-fluoro-2'-desoxiadenosina (A2'F)	-	F	No

Tabla 5: Ejemplos no limitantes de análogos de G

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

5 En ciertas realizaciones, un análogo (por ejemplo, un nucleótido modificado) se puede seleccionar 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-metiluridina, 1-metil-pseudouridina, 4-tio-1-metil-pseudouridina, 2-tio-1-metil-pseudouridina, 1-metil-1-deazapseudouridina, 2-tio-1-metil-1-deaza-pseudouridina, dihidrouridina, dihidropseudouridina, 2-tio-dihidrouridina, 2-tio-dihidropseudouridina, 2-metoxiuridina, 2-metoxi-4-tio-uridina, 4-metoxi-pseudouridina, 4-metoxi-2-tiopseudouridina, 5-aza-citidina, pseudoisocitidina, 3-metil-citidina, N4-acetilcitidina, 5-formilcitidina, 5-metilcitidina, N4-metilcitidina, 5-hidroximetilcitidina, 1-metil-pseudoisocitidina, pirrolo-citidina, pirrolo-pseudoisocitidina, 2-tio-citidina, 2-tio-5-metil-citidina, 4-tio-pseudoisocitidina, 4-tio-1-metil-pseudoisocitidina, 4-tio-1-metil-1-desaza-pseudoisocitidina, 1-metil-1-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-1-metil-pseudoisocitidina, 2-aminopurina, 2, 6-diaminopurina, 7-desaza-adenina, 7-desaza-8-za-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-metil-N6-(cis-hidroxiisopentenil)adenosina, N6-glicinilcarbamoiladenosina, N6-treonilcarbamoiladenosina, 2-metil-N6-treonil-carbamoiladenosina, N6,N6-dimetiladenosina, 7-metiladenina, 2-metiladenina, 2-metoxi-adenina, inosina, 1-metil-inosina, wiosina, wibutosina, 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-metilinosina, 6-metoxi-guanosina, 1-metilguanosa, N2-metilguanosa, N2,N2-dimetilguanosa, 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. En ciertas realizaciones, una molécula de ARN/un polirribonucleótido modificado de la presente invención no incluye pseudouridina. En ciertas realizaciones, una molécula de ARN/polirribonucleótido modificado de la presente invención no incluye 5-metil-uridina. En ciertas realizaciones, una molécula de ARN/polirribonucleótido modificado de la presente invención no incluye 5-metil-uridina. En ciertas realizaciones, una molécula de ARN/un polirribonucleótido modificado de la presente invención comprende análogos de U y análogos de C, en la que dichos análogos de U pueden ser todos del mismo análogo o pueden ser análogos diferentes (por ejemplo, más de un tipo de análogo) y en el que dichos análogos de C pueden ser todos el mismo análogo o pueden ser análogos diferentes (por ejemplo, más de un tipo de análogo). En ciertas realizaciones, una molécula de ARN/un polirribonucleótido modificado de la presente invención no incluye análogos de adenosina ni análogos de guanosina. 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 cierta proporción de los nucleótidos del compuesto (por ejemplo, un porcentaje dado de una nucleobase dada puede ser análogo, como se describe en el presente documento).

40 Una molécula de ARN/un polirribonucleótido que comprende al menos un nucleótido modificado es una molécula de ARN/polirribonucleótido modificado. En ciertas realizaciones, al menos aproximadamente el 5 % de la molécula de ARN/polirribonucleótido modificado incluye adenosina, citidina, guanosina o uridina modificada o que no se produce

de manera natural (por ejemplo, análogos o modificados), tal como los nucleótidos análogos descritos en el presente documento. En algunos casos, al menos aproximadamente el 5 %, 10 %, 15 %, 20 %, 25 %, 30 %, 40 %, 45 %, 50 % de la molécula de ARN/del polirribonucleótido modificado incluye adenosina, citidina, guanosina o uridina modificadas o no naturales (por ejemplo, análogos o modificadas). En algunos casos, como máximo, aproximadamente el 50 %, 45 %, 40 %, 35 %, 30 %, 25 %, 20 %, 15 %, 10 %, 5 %, 1 % de la molécula de ARN/del polirribonucleótido modificado incluye adenosina, citidina, guanosina o uridina modificadas o no naturales.

En una realización preferida, la molécula de ARN de la presente invención contiene una combinación de nucleótidos modificados y no modificados. Preferentemente, la molécula de ARN de la presente invención contiene una combinación de nucleótidos modificados y no modificados como se describe en el documento WO 2011/012316. Dichas moléculas de ARN también son conocidas y se comercializan como "SNIM[®]-RNA". Se informa que la molécula de ARN descrita en el documento WO 2011/012316 muestra una mayor estabilidad y una menor inmunogenicidad. En una realización preferida, en dicha molécula de ARN modificada está modificado del 5 al 50 % de los nucleótidos de citidina y del 5 al 50 % de los nucleótidos de uridina. Los nucleótidos que contienen adenosina y guanosina pueden no estar modificados. Los nucleótidos de adenosina y guanosina pueden estar no modificados o parcialmente modificados y están preferentemente presentes de forma no modificada. Preferentemente, 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 del 7,5 al 25 % y el contenido de los nucleótidos de uridina modificados en un intervalo del 7,5 al 25 %. Se ha encontrado que, de hecho, un contenido relativamente bajo, por ejemplo, de solo el 10 % de nucleótidos de citidina y uridina modificados puede alcanzar las propiedades deseadas. Se prefiere en particular que los nucleótidos de citidina modificados sean restos de 5-metilcitidina, y los nucleótidos de uridina modificados sean restos de 2-tiouridina. Lo más preferentemente, el contenido de nucleótidos de citidina modificados y el contenido de los nucleótidos de uridina modificados es del 25 %, respectivamente.

En otras ciertas realizaciones, en dicha molécula de ARN/molécula de polirribonucleótido modificada, del 5 al 50 % de las citidinas son análogos de C y del 5 al 50 % de las uridinas son análogos de U. En ciertas realizaciones, en dicha molécula de polirribonucleótido modificada, del 5 al 40 % de las citidinas son análogos de C y del 5 al 40 % de las uridinas son análogos de U. En ciertas realizaciones, en dicha molécula de ARN/molécula de polirribonucleótido modificada, del 5 al 30 % de las citidinas son análogos de C y del 5 al 30 % de las uridinas son análogos de U. En ciertas realizaciones, en dicha molécula de ARN/molécula de polirribonucleótido modificada, del 10 al 30 % de las citidinas son análogos de C y del 10 al 30 % de las uridinas son análogos de U. En ciertas realizaciones, en dicha molécula de polirribonucleótido modificada, del 5 al 20 % de las citidinas son análogos de C y del 5 al 20 % de las uridinas son análogos de U. En ciertas realizaciones, en dicha molécula de ARN/molécula de polirribonucleótido modificada, del 5 al 10 % de los nucleótidos de citidina y del 5 al 10 % de los nucleótidos de uridina están modificados. En ciertas realizaciones, en dicha molécula de ARN/molécula de polirribonucleótido modificada, el 25 % de los nucleótidos de citidina y el 25 % de los nucleótidos de uridina están modificados. En ciertas realizaciones, los nucleótidos que contienen adenosina y guanosina pueden no estar modificados. En ciertas realizaciones, los nucleótidos de adenosina y guanosina pueden no estar modificados o estar parcialmente modificados, y están preferentemente presentes en forma no modificada. Como se ha indicado anteriormente, en ciertas realizaciones, los análogos de U se refieren a un solo tipo de análogo de U. En ciertas realizaciones, los análogos de U se refieren a dos o más tipos de análogos de U. En ciertas realizaciones, los análogos de C se refieren a un solo tipo de análogo de C. En ciertas realizaciones, los análogos de C se refieren a dos o más tipos de análogos de C.

En ciertas realizaciones, 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. En ciertas realizaciones, el porcentaje de análogos de citidina es inferior al porcentaje de análogos de uridina. Como se ha indicado anteriormente, esto puede ser en presencia o en ausencia de análogos de adenosina y guanosina pero, en ciertas realizaciones, es en ausencia de análogos de adenosina y análogos de guanosina. En ciertas realizaciones, 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. En ciertas realizaciones, 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 son análogos de citidina y del 25 al 45 % de las uridinas son análogos de uridina. En otras palabras, 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. En otras realizaciones, 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 del 7 al 9 % de análogos de citidina, tal como aproximadamente el 7, 7,5 u 8 % y, tal como del 32 al 38 % de análogos de uridina, tal como aproximadamente el 33, 34, 35, 36 %. En ciertas realizaciones, puede usarse cualquiera de los análogos de uridina y análogos de citidina descritos en el presente documento, excluyendo opcionalmente la pseudouridina. En ciertas realizaciones, el análogo de citidina comprende o consiste en (por ejemplo, en el caso de que consista 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 que consista en, es el único tipo de análogo usado) 5-yodouridina. En ciertas realizaciones de cualquiera de las anteriores, 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 inicio). En ciertas realizaciones de cualquiera de las anteriores, el

porcentaje de análogos de un nucleótido dado se refiere a la producción (por ejemplo, al porcentaje en un compuesto sintetizado o transcrito). Las moléculas de ARN/moléculas de polirribonucleótido de la presente invención se pueden producir de forma recombinante en sistemas *in vivo* mediante métodos conocidos por los expertos en la materia que se describen con más detalle a continuación.

5 Como alternativa, las moléculas de polirribonucleótido modificadas de la presente invención pueden producirse en un sistema *in vitro* usando, por ejemplo, un sistema de transcripción *in vitro* que se describe con más detalle más adelante. Un sistema de transcripción *in vitro* capaz de producir moléculas de ARN/polirribonucleótidos requiere una mezcla de entrada de nucleósidos trifosfatos modificados y no modificados para producir moléculas de
 10 ARN/polirribonucleótidos modificados con las propiedades deseadas de la presente invención. En ciertas realizaciones, del 5 al 50 % de las citidinas son análogos de citidina en dicha mezcla de entrada y del 5 al 50 % de las uridinas son análogos de uridina en dicha mezcla de entrada. En ciertas realizaciones, del 5 al 40 % de las citidinas son análogos de citidina en dicha mezcla de entrada y del 5 al 40 % de las uridinas son análogos de uridina en dicha mezcla de entrada. En ciertas realizaciones, del 5 al 30 % de las citidinas son análogos de citidina en dicha mezcla de entrada y del 5 al 30 % de las uridinas son análogos de uridina en dicha mezcla de entrada. En ciertas realizaciones, del 5 al 30 % de las citidinas son análogos de citidina en dicha mezcla y del 10 al 30 % de las uridinas son análogos de uridina en dicha mezcla. En ciertas realizaciones, del 5 al 20 % de las citidinas son análogos de citidina en dicha mezcla de entrada y del 5 al 20 % de las uridinas son análogos de uridina en dicha mezcla de entrada. En ciertas realizaciones, del 5 al 10 % de las citidinas son análogos de citidina en dicha mezcla de entrada y del 5 al 10 % de las uridinas son análogos de uridina en dicha mezcla de entrada. En ciertas realizaciones, el 25% de las citidinas son análogos de citidina en dicha mezcla de entrada y el 25 % de las uridinas son análogos de uridina en dicha mezcla de entrada. En ciertas realizaciones, la mezcla de entrada no comprende análogos de adenosina y/o guanosina. En otras realizaciones opcionalmente, la mezcla de entrada comprende uno o más análogos de adenosina y/o guanosina (o ninguno de ellos o ambos). En ciertas realizaciones, 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. En ciertas realizaciones, el porcentaje de análogos de citidina en una mezcla de entrada es inferior al porcentaje de análogos de uridina en una mezcla de entrada. Como se ha indicado anteriormente, esto puede ser en presencia o en ausencia de análogos de adenosina y guanosina en la mezcla de entrada pero, en ciertas realizaciones, es en ausencia de análogos de adenosina y análogos de guanosina en la mezcla de entrada. En ciertas realizaciones, 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. En otras palabras, la mezcla de entrada comprende citidinas modificadas y no modificadas, y uridinas modificadas y no modificadas, y del
 35 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 del 7 al 9 % de análogos de citidina, tal como aproximadamente el 7, 7,5 u 8 % y, tal como del 32 al 38 % de análogos de uridina, tal como aproximadamente el 33, 34, 35, 36 %. En ciertas realizaciones, puede usarse cualquiera de los análogos de uridina y análogos de citidina descritos en el presente documento, excluyendo opcionalmente la pseudouridina. En ciertas realizaciones, 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. En las tablas anteriores, se describen ejemplos de análogos. Se ha de entender que para los polirribonucleótidos modificados que codifican el polipéptido deseado (módulo (a)), los análogos y el nivel de modificación se consideran, a menos que se indique lo contrario, en todo el polirribonucleótido que codifica el polipéptido deseado (módulo (a)), incluyendo las regiones no traducidas 5' y 3' (por ejemplo, el nivel de modificación se basa en las proporciones de entrada de los análogos en una reacción de transcripción *in vitro* de forma que los análogos puedan incorporarse en posiciones que se transcriban). Además, las moléculas de ARN/moléculas de polirribonucleótido modificadas pueden sintetizarse químicamente, por ejemplo, mediante síntesis química convencional en un sintetizador de secuencia de nucleótidos automatizado usando un soporte en fase sólida y técnicas convencionales o mediante síntesis química de las respectivas secuencias de ADN y la posterior transcripción *in vitro* o *in vivo* de las mismas.

55 En biología molecular y genética, cadena arriba y cadena abajo se refieren a una posición relativa en una molécula de ARN. En el contexto de la presente invención, cadena arriba es hacia el extremo 5' de la molécula de ARN y cadena abajo es hacia el extremo 3' de la molécula. Por consiguiente, en una realización, el módulo (b) de UTR (es decir, la una o más UTR que comprenden la secuencia como se muestra en SEQ ID NO: 1 o una secuencia que tiene de 1 a 4 sustituciones en comparación con SEQ ID NO: 1 y que lugar a una molécula de ARN que tiene la misma o mayor eficacia de traducción que una molécula de ARN que comprende una UTR que comprende la SEQ ID NO: 1 como se ha definido anteriormente) se sitúa cadena arriba de la región de codificación del módulo (a). Además, en una realización, el módulo (c) de UTR (es decir, la una o más UTR que comprenden la secuencia como se muestra en SEQ ID NO: 2 o una secuencia que muestra de 1 a 7 sustituciones en comparación con SEQ ID NO: 2 y que da lugar a una molécula de ARN que tiene la misma o mayor eficacia de traducción que una molécula de ARN que comprende una UTR que comprende la SEQ ID NO: 2 como se ha definido anteriormente) se sitúa cadena
 60 abajo de la región de codificación del módulo (a). Sin embargo, preferentemente, la región de codificación que codifica un polipéptido (es decir, el módulo (a)) se ubica entre el módulo (b) de UTR y el módulo (c) de UTR y, por

consiguiente, la molécula de ARN preferentemente tiene la disposición de 5'-(b)-(a)-(c)-3'. En caso de que la molécula de ARN solo contenga un módulo de UTR (es decir, cualquiera de los módulos (b) (es decir, una o más UTR que comprenden la secuencia como se muestra en SEQ ID NO: 1 o una secuencia que muestra de 1 a 4 sustituciones en comparación con SEQ ID NO: 1 y que da lugar a una molécula de ARN que tiene la misma o mayor eficacia de traducción que una molécula de ARN que comprende una UTR que comprende la SEQ ID NO: 1 como se ha definido anteriormente) o el módulo (c) (es decir, la una o más UTR que comprenden la secuencia mostrada en SEQ ID NO: 2 o una secuencia que muestra de 1 a 7 sustituciones en comparación con SEQ ID NO: 2 y que da lugar a una molécula de ARN que tiene la misma o mayor eficacia de traducción que molécula de ARN que comprende una UTR que comprende la SEQ ID NO: 2 como se ha definido en el presente documento anteriormente)) la molécula de ARN tiene preferentemente la disposición de 5'-(b)-(a)-3' o 5'-(a)-(c)-3'.

La molécula de ARN puede estar presente en forma de secuencias de ARN fusionadas de los módulos (a), (b) y/o (c), es decir, una molécula de ARN (de fusión) que se forma mediante la expresión de un gen híbrido producido mediante la combinación de al menos dos secuencias de nucleótidos que codifican dichos módulos. Por lo general, como se explicará con más detalle más adelante, esto se puede lograr clonando un ADNc en un vector de expresión que permita la traducción de la molécula de ARN. Por consiguiente, la molécula de ADN que codifica la molécula de ARN de la presente invención puede ser una secuencia de ADN fusionada, es decir, una molécula quimérica que se forma uniendo dos o más polinucleótidos a través del grupo fosfato de un nucleótido unido al átomo de carbono 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 dichos al menos dos módulos, preferentemente los tres módulos, se unen entre sí en forma de una molécula de ADN en términos de la presente invención. Una vez clonado en fase, dicha molécula de ADN recombinante se transcribe luego en su secuencia de ácido nucleico de ARN correspondiente que codifica dicha molécula de proteína, polipéptido o enzima. Como alternativa, los al menos dos módulos, preferentemente los tres módulos, también pueden acoplarse covalentemente mediante un conjugado químico. Por lo tanto, como se describirá con más detalle más adelante, los módulos de la molécula de ARN pueden sintetizarse químicamente individualmente, y posteriormente acoplarse en un enlace covalente mediante un enlace fosfodiéster como se ha descrito anteriormente.

A continuación, se describen disposiciones preferidas de los módulos (b) y/o (c) de UTR de la presente invención en relación con la región de codificación (a) en las que el módulo (b) de UTR (que corresponde al fragmento de UTR 5' definido anteriormente del ARNm de CYBA) se sitúa cadena arriba de la región de codificación (es decir, en el extremo 5' de la región de codificación) y/o el módulo (c) de UTR (que corresponde a la UTR 3' anteriormente definida del ARNm de CYBA) se sitúa cadena abajo de la región de codificación (es decir, en el extremo 3' de la región de codificación).

Por lo tanto, en una realización preferida y de acuerdo con lo anterior, la presente invención se refiere a una molécula de ARN que comprende (a) una región de codificación que codifica un polipéptido; y (b) una o más UTR que comprenden la secuencia mostrada en SEQ ID NO: 1 o una secuencia que muestra de 1 a 4 sustituciones en comparación con SEQ ID NO: 1 y que da lugar a una molécula de ARN que tiene la misma o una mayor eficacia de traducción que una molécula de ARN que comprende una UTR que comprende la SEQ ID NO: 1, en la que dicha región de codificación que codifica un polipéptido en (a) no es una región de codificación que codifica un polipéptido alfa del citocromo b-245 (CYBA) como se ha definido anteriormente en el presente documento y en la que dicha/s UTR como se define/n en (b) está/n situada/s en el extremo 5' de la región de codificación según lo definido en (a).

En una realización preferida y de acuerdo con lo anterior, la presente invención se refiere a una molécula de ARN que comprende (a) una región de codificación que codifica un polipéptido; y (c) una o más UTR que comprenden la secuencia mostrada en SEQ ID NO: 2 o una secuencia que muestra de 1 a 7 sustituciones en comparación con SEQ ID NO: 2 y que da lugar a una molécula de ARN que tiene la misma o una mayor eficacia de traducción que una molécula de ARN que comprende una UTR que comprende la SEQ ID NO: 2, en la que dicha región de codificación que codifica un polipéptido en (a) no es una región de codificación que codifica un polipéptido alfa del citocromo b-245 (CYBA) como se ha definido anteriormente en el presente documento y en la que dicha/s UTR según lo definido en (c) está/n ubicada/s en el extremo 3' de la región de codificación según lo definido en (a).

En una realización preferida y de acuerdo con lo anterior, la presente invención se refiere a una molécula de ARN que comprende (a) una región de codificación que codifica un polipéptido; y (b) una o más UTR que comprenden la secuencia mostrada en SEQ ID NO: 1 o una secuencia que muestra de 1 a 4 sustituciones en comparación con SEQ ID NO: 1 y que da lugar a una molécula de ARN que tiene la misma o una mayor eficacia de traducción que una molécula de ARN que comprende una UTR que comprende la SEQ ID NO: 1; y (c) una o más UTR que comprenden la secuencia mostrada en SEQ ID NO: 2 o una secuencia que muestra de 1 a 7 sustituciones en comparación con SEQ ID NO: 2 y que da lugar a una molécula de ARN que tiene la misma o una mayor eficacia de traducción que una molécula de ARN que comprende una UTR que comprende la SEQ ID NO: 2, en la que dicha región de codificación que codifica un polipéptido en (a) no es una región de codificación que codifica un polipéptido alfa del citocromo b-245 (CYBA) según lo definido anteriormente en el presente documento y en la que dicha/s UTR según lo definido en (b) están ubicadas en el extremo 5' de la región de codificación según lo definido en (a) y en la que dichas UTR están ubicadas según lo definido en (c) en el extremo 3' de la región de codificación según lo definido en (a).

En una realización preferida y de acuerdo con lo anterior, la presente invención se refiere a una molécula de ARN que comprende (a) una región de codificación que codifica un polipéptido; y (b) una UTR que comprende la secuencia mostrada en SEQ ID NO: 1 o una secuencia que muestra de 1 a 4 sustituciones en comparación con SEQ ID NO: 1 y que da lugar a una molécula de ARN que tiene la misma o mayor eficacia de traducción que una molécula de ARN que comprende una UTR que comprende la SEQ ID NO: 1; y (c) dos UTR que comprenden la secuencia mostrada en SEQ ID NO: 2 o una secuencia que muestra de 1 a 7 sustituciones en comparación con SEQ ID NO: 2 y que da lugar a una molécula de ARN que tiene la misma o mayor eficacia de traducción que una molécula de ARN que comprende una UTR que comprende la SEQ ID NO: 2; en la que dicha región de codificación que codifica un polipéptido en (a) no es una región de codificación que codifica un polipéptido alfa del citocromo b-245 (CYBA) como se ha definido anteriormente en el presente documento y en la que dicha molécula de ARN comprende dicha UTR según lo definido en (b) en el extremo 5' de la región de codificación según lo definido en (a) y que comprende dichas dos UTR según lo definido en (c) en el extremo 3' de la región de codificación según lo definido en (a).

En una realización preferida y de acuerdo con lo anterior, la presente invención se refiere a una molécula de ARN que comprende (a) una región de codificación que codifica un polipéptido; y (c) dos UTR que comprenden la secuencia mostrada en SEQ ID NO: 2 o una secuencia que muestra de 1 a 7 sustituciones en comparación con SEQ ID NO: 2 y que da lugar a una molécula de ARN que tiene la misma o mayor eficacia de traducción que una molécula de ARN que comprende una UTR que comprende la SEQ ID NO: 2, en la que dicha región de codificación que codifica un polipéptido en (a) no es una región de codificación que codifica un polipéptido alfa del citocromo b-245 (CYBA) como se ha definido anteriormente y en la que dicha molécula de ARN comprende dichas dos UTR según lo definido en (c) en el extremo 3' de la región de codificación según lo definido en (a).

Como se ha mencionado anteriormente, la molécula de ARN de la presente invención también puede albergar una cola de poli-A. Como se usa en el presente documento, una cola de poli-A se refiere a una secuencia de nucleótidos de adenina situados en el extremo 3' del ARN. Por lo general, se añade una cola de poli-A al extremo 3' del ARN mediante un proceso denominado poliadenilación. Por lo tanto, la presente invención se refiere a cualquiera de los ARN descritos anteriormente, en la que la molécula de ARN comprende una cola de poli-A en el extremo 3'. La longitud de la cola de poli-A no se limita en particular. 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 o 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 o 1.000 nucleótidos. En caso de que la molécula de ARN de la presente invención se produzca mediante un método de transcripción *in vitro* como se describe en el presente documento más adelante, la cola de poli-A está situada en el extremo 3' del ARN adyacente a la UTR en el extremo 3' de la construcción de ARN, mientras que el plásmido que alberga la molécula de ARN de la presente invención se linealiza antes de la transcripción *in vitro* cadena abajo de la cola de poli-A para garantizar que la molécula de ARN transcrita *in vitro* contenga dicha cola de poli-A.

La construcción de acuerdo con la presente invención puede comprender no solo los tres módulos principales anteriores (a), (b) y/o (c). En cambio, puede ser deseable que, entre los módulos (a) individuales se coloquen fracción/fracciones enlazadora/s y/o sitio/s de clonación múltiple (a) que pueden, por ejemplo, facilitar la creación de la construcción. Las fracciones enlazadoras adecuadas y los sitios de clonación múltiples son conocidos por los expertos. Preferentemente, la construcción de la presente invención alberga un sitio de clonación múltiple que se deriva del plásmido pVAX1 (Invitrogen). Todas las construcciones como se esbozan en el apartado de ejemplos se originan a partir de la construcción pVAX A120 que se ha descrito previamente en el documento WO2013/182683 A1.

La posición de los módulos de UTR (b) y/o (c) dentro de la molécula de ARN de la presente invención en relación con el módulo (a) (es decir, la región de codificación) no se limita en particular y, por consiguiente, entre los módulos individuales de la molécula de ARN de la presente invención, puede haber un espacio o un hueco relleno con uno o más nucleótidos G, A, U y/o C que no son parte de los módulos principales (a), (b) y/o (c). "Uno o más nucleótidos G, A, U y/o C", en este contexto, significa que el espacio o hueco entre los módulos individuales de la molécula de ARN de la presente invención está/n lleno/s de 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10 nucleótidos G, A, U y/o C. En otras realizaciones preferidas, el espacio o hueco entre los módulos individuales de la molécula de ARN de la presente invención se llena con 20, 30, 40, 50, 60, 70, 80, 90, 100 o 110 o más nucleótidos G, A, U y/o C. Sin embargo, en una realización preferida, el módulo (b) o (c) de UTR, dentro de la molécula de ARN de la presente invención en relación con el módulo (a) (es decir, la región de codificación), está situado directamente adyacente al codón de inicio de la región de codificación del módulo (a) sin ningún espacio o hueco entre ellos, es decir, directamente cadena arriba del codón de inicio de la región de codificación del módulo (a). En otra realización preferida, el módulo (b) o (c) de UTR, dentro de la molécula de ARN de la presente invención en relación con el módulo (a) (es decir, la región de codificación), está situado directamente adyacente al codón de terminación (es decir, el codón de parada) de la región de codificación del módulo (a) sin ningún espacio o separación, es decir, directamente cadena abajo del codón de terminación/codón de parada de la región de codificación del módulo (a).

En una realización preferida, el módulo (b) de UTR, dentro de la molécula de ARN de la presente invención en relación con el módulo (a) (es decir, la región de codificación), está situado directamente adyacente al codón de inicio de la región de codificación del módulo (a) sin ningún espacio o hueco entre ellos, es decir, directamente
 5 molécula de ARN de la presente invención en relación con el módulo (a) (es decir, la región de codificación), está situado directamente adyacente al codón de terminación (es decir, el codón de parada) de la región de codificación del módulo (a) sin ningún espacio o hueco intermedio, es decir, directamente cadena abajo del codón de terminación/codón de parada de la región de codificación del módulo (a).

10 Como se ha mencionado anteriormente, la molécula de ARN puede estar presente en forma de secuencias de ARN fusionadas de los módulos (a), (b) y/o (c), es decir, una molécula de ARN (de fusión) que se forma mediante la transcripción de un gen híbrido realizada mediante la combinación de al menos dos secuencias de nucleótidos que codifican dichos módulos. Por lo general, esto se lleva a cabo clonando un ADNc en un vector de expresión que permita la transcripción de la molécula de ARN completa. Se conoce una variedad de métodos para fabricar
 15 construcciones de fusión, que incluyen la síntesis, hibridación y/o amplificación de ácidos nucleicos para producir una molécula de ácido nucleico bicatenario sintético que codifique la molécula de ARN de la presente invención. Dicha molécula de ácido nucleico bicatenario (es decir, molécula de ADN) alberga en una cadena (es decir, en la cadena de codificación) la secuencia de ADN correspondiente a la molécula de ARN de la presente invención y, por consiguiente, "codifica" la molécula de ARN de la presente invención. En otras palabras, dicha molécula de ácido
 20 nucleico/ADN bicatenario comprende, en una cadena, la información genética, cuando se transcribe la molécula de ARN de la presente invención como se ha definido anteriormente en el presente documento. La expresión "que codifica" en el contexto de la presente invención no solo se usa en su sentido convencional, es decir, para referirse al ADN de un gen que codifica una proteína (y, por consiguiente, la información genética que puede traducirse a un polipéptido o una secuencia de aminoácidos de proteína). En cambio, en términos de la presente invención, en una
 25 construcción en la que las secuencias de ADN individuales que codifican los módulos (a), (b) y/o (c) se "fusionan" o se unen en una sola molécula de ADN (quimérico), la construcción también comprende componentes (es decir, módulo (b) y/o módulo (c)) que no están traducidos a una proteína. Sin embargo, la secuencia de ADN correspondiente al módulo (b) y/o módulo (c) proporciona la información, es decir, el "código", para la estructura de las UTR de la presente invención y, por consiguiente, la expresión "que codifica" en el presente invención también se
 30 refiere a la información genética para las UTR que pueden expresarse, es decir, transcribirse, si, por ejemplo, está presente en una molécula de ácido nucleico bicatenario que alberga en una cadena la molécula de ARN de la presente invención. Por lo tanto, la expresión "que codifica" en el contexto de la presente invención, aunque se usa comúnmente solo para referirse a la codificación/expresión de una proteína, debe entenderse de forma que la molécula de ácido nucleico pueda transcribirse en la molécula de ARN de la presente invención que alberga partes
 35 que codifican una proteína o un polipéptido (es decir, módulo (a)) y partes que "codifican" las UTR (es decir, módulos (b) y/o (c)) en la que estos últimos representan el producto final cuando se expresa, ya que las UTR no se traducen a proteínas o polipéptidos. Dicho ácido nucleico bicatenario se puede insertar en vectores de expresión para la producción de proteínas de fusión mediante técnicas convencionales de Biología molecular (véase, por ejemplo, Sambrook *et al.*, "Molecular Cloning, A laboratory manual", 2ª Ed, 1989). El término "vector" tal como en "vector de expresión" o "vector de clonación" en el sentido de la presente invención se entiende como una unidad de ADN bicatenario circular que se replica dentro de una célula independientemente del ADN cromosómico y que se usa
 40 como un vehículo para transportar material genético a una célula, en la que puede replicarse y/o expresarse (es decir, transcribirse a ARN y traducirse a una secuencia de aminoácidos). Un vector que contiene ADN foráneo se denomina ADN recombinante. En general, el propio vector es una secuencia de ADN que normalmente consiste en una inserción (es decir, módulo (b) y/o módulo (c) que no se traduce a una proteína y módulo (a) la región de codificación) y una secuencia mayor que sirve como la "estructura principal" del vector. Los plásmidos en el sentido de la presente invención se encuentran con mayor frecuencia en bacterias, y se usan en la investigación del ADN recombinante para transferir genes entre células y son, como tales, una subpoblación de "vectores" tal como se usa
 45 en el sentido de la presente invención.

50 Por lo tanto, la presente invención también se refiere a una molécula de ácido nucleico que codifica la molécula de ARN de la presente invención. El ácido nucleico es, por ejemplo, un ADN, que codifica dos de los tres módulos principales (es decir, módulo (a) y módulo (b) o módulo (c)) de la molécula de ARN de la presente invención. Como alternativa, el ácido nucleico, preferentemente un ADN, codifica los tres módulos principales (es decir, el módulo (a) y el módulo (b) y el módulo (c)). La molécula de ácido nucleico anterior de la presente invención preferentemente es una molécula de ácido nucleico recombinante, pero también puede comprender moléculas de ácido nucleico de origen natural. La molécula de ácido nucleico de la invención puede, por lo tanto, ser de origen natural, sintético o semisintético. Puede comprender ADN, ARN, ácido nucleico bloqueado, así como PNA, y puede ser un híbrido de los mismos.
 60

65 Es evidente para el experto en la materia que se pueden añadir secuencias reguladoras a la molécula de ácido nucleico de la invención que codifica la molécula de ARN. Por ejemplo, se pueden emplear promotores, potenciadores de la transcripción y/o secuencias que permiten la expresión inducida del polinucleótido, es decir, la molécula de ARN, de la invención. Un sistema inducible adecuado es, por ejemplo, la expresión génica regulada por tetraciclina según lo descrito, por ejemplo, por Gossen y Bujard, *Proc. Natl. Acad. Sci. EE.UU.* 89 (1992), 5547-5551) y Gossen, *Trends Biotech.* 12 (1994), 58-62 o un sistema de expresión génica inducible por dexametasona según lo

descrito, por ejemplo, por Crook, *EMBO J.* 8 (1989), 513-519.

Además, dicha molécula de ácido nucleico puede contener, por ejemplo, enlaces de tioéster y/o análogos de nucleótidos. Dichas modificaciones pueden ser útiles para la estabilización de la molécula de ácido nucleico contra endonucleasas y/o exonucleasas en la célula. Dichas moléculas de ácido nucleico pueden transcribirse a partir de un vector apropiado que contenga un gen quimérico que permita la transcripción de dicha molécula de ácido nucleico en la célula. En el contexto de la presente invención, dichas moléculas de ácido nucleico también pueden marcarse. Los métodos de detección de ácidos nucleicos son bien conocidos en la técnica, por ejemplo, la transferencia Southern y Northern, PCR o extensión de cebador.

La/s molécula/s de ácido nucleico de la invención pueden ser una molécula de ácido nucleico quimérico producida recombinantemente que comprenda cualquiera de las moléculas de ácido nucleico mencionadas anteriormente, ya sea solas o en combinación. Preferentemente, la molécula de ácido nucleico de la invención es parte de un vector.

La presente invención, por lo tanto, también se refiere a un vector que comprende la molécula de ácido nucleico de la presente invención. Por consiguiente, la presente invención se refiere a vectores, preferentemente vectores de expresión que comprenden los ácidos nucleicos de la 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 marcadores que permitan la selección de dicho vector en una célula hospedadora adecuada y en condiciones adecuadas.

Además, el vector de la presente invención puede, además de las secuencias de la molécula de ácido nucleico que codifica la molécula de ARN de la invención, comprender elementos de control de la expresión, permitiendo la expresión apropiada de las regiones de codificación en células hospedadoras adecuadas.

Dichos elementos de control son conocidos por el experto en la materia, y pueden incluir un promotor, un casete de corte y empalme, un codón de inicio de la traducción, un sitio de traducción y de inserción para introducir una inserción en el vector. Preferentemente, la molécula de ácido nucleico de la invención está operativamente unida a dichas secuencias de control de la expresión que permiten la expresión en células eucariotas o procariotas. Por consiguiente, la presente invención se refiere a un vector que comprende la molécula de ácido nucleico de la presente invención, en la que la molécula de ácido nucleico está operativamente unida a secuencias de control que son reconocidas por una célula hospedadora cuando la célula eucariota y/o procariota (hospedadora) es transfectada con el vector. Los elementos de control que garantizan la expresión en células eucariotas y procariotas (hospedadoras) son bien conocidos por los expertos en la materia. Como se ha mencionado anteriormente en el presente documento, habitualmente comprenden secuencias reguladoras que garantizan el inicio de la transcripción y, opcionalmente, señales de poli-A que garantizan la terminación de la transcripción y la estabilización del producto de transcripción. Sin embargo, de acuerdo con la presente invención, no es crucial que el vector en sí contenga una secuencia para una cola de poli-A. Como se ha mencionado anteriormente, en caso de que la molécula de ARN de la presente invención se produzca mediante un método de transcripción *in vitro* como se describe en el presente documento más adelante, la cola de poli-A anterior es parte de la construcción de la presente invención (y no necesariamente se ubica originalmente en el vector de clonación) y está ubicada en el extremo 3' del ARN adyacente a la UTR en el extremo 3' de la construcción de ARN. En caso de que la molécula de ARN de la presente invención se produzca 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* cadena abajo de la cola de poli-A para garantizar que el la molécula de ARN transcrita contenga dicha cola de poli-A. Los elementos reguladores adicionales pueden incluir potenciadores de la transcripción, así como de la traducción y/o regiones promotoras heterólogas o asociadas de forma natural. Los posibles elementos reguladores que permiten la expresión en, por ejemplo, células hospedadoras de mamíferos comprenden el promotor de timidina quinasa CMV-HSV, SV40, promotor RSV (virus del sarcoma de Rous), promotor 1 α del factor de elongación humano, promotor de MMTV inducible por glucocorticoides, virus del tumor de mama de ratón), promotores o potenciadores inducibles por metalotioneína o tetraciclina, como potenciador de CMV o potenciador de SV40. Para la expresión en células neuronales, se prevé que pueden emplearse promotores de neurofilamentos, PGDF, NSE, PrP o thy-1. Dichos promotores son conocidos en la técnica y, entre otros, se describen en Charron, *J. Biol. Chem.* 270 (1995), 25739-25745. Para la expresión en células procariotas, se ha descrito una multitud de promotores que incluyen, por ejemplo, el promotor tac-lac o el promotor trp. Además de los elementos que son responsables del inicio de la transcripción, dichos elementos reguladores también pueden comprender señales de terminación de la transcripción, tales como el sitio poli-A de SV40 o el sitio poli-A de tk, cadena abajo del polinucleótido. En este contexto, los vectores de expresión adecuados son conocidos en la técnica, tales como el vector de expresión de ADNc de Okayama-Berg pcDV1 (Pharmacia), pRc/CMV, pcDNA1, pcDNA3 (In-vitrogene), pSPORT1 (GIBCO BRL), pX (Pagano, *Science* 255 (1992), 1144-1147), vectores de dos híbridos de levadura, tales como pEG202 y dpJG4-5 (Gyuris, *Cell* 75 (1995), 791-803) o vectores de expresión procariotas, tales como lambda gt11 o pGEX (Amersham-Pharmacia). Además, el vector de la presente invención también puede ser un vector de expresión. Las moléculas de ácido nucleico y los vectores de la invención pueden diseñarse para la introducción directa o para la introducción a través de liposomas, vectores víricos (por ejemplo, adenovíricos, retrovíricos), electroporación, balística (por ejemplo, pistola de genes) u otros sistemas de

suministro a la célula. Además, se puede usar un sistema de baculovirus como sistema de expresión eucariota para las moléculas de ácido nucleico de la invención. La presente invención también se refiere a una célula hospedadora que comprende el vector de la presente invención. Por lo tanto, la presente invención se refiere a una célula hospedadora transfectada o transformada con el vector de la invención, es decir, a una célula hospedadora que está modificada genéticamente con una molécula de ácido nucleico de acuerdo con la invención o con un vector que comprende dicha molécula de ácido nucleico. La expresión "genéticamente modificada" significa que la célula hospedadora comprende, además de su genoma natural, una molécula o un vector de ácido nucleico de acuerdo con la invención que se introdujo en la célula o en uno de sus predecesores. La molécula o el vector de ácido nucleico pueden estar presentes en la célula hospedadora genéticamente modificada como una molécula independiente fuera del genoma, preferentemente como una molécula que es capaz de replicarse, o puede integrarse de forma estable en el genoma de la célula hospedadora. La transformación de la célula hospedadora con un vector de acuerdo con la invención se puede llevar a cabo mediante métodos convencionales, 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 hospedadora se cultiva en medios de nutrientes que cumplen los requisitos de la célula hospedadora usada en particular, en concreto, con respecto al valor del pH, la temperatura, la concentración de sal, la aireación, los antibióticos, las vitaminas, los oligoelementos, etc. La célula hospedadora de la presente invención puede ser cualquier célula procariota o eucariota. Las células procariotas adecuadas son las que se usan en general para la clonación como *E. coli* o *Bacillus subtilis*. Además, las células eucariotas comprenden, por ejemplo, células fúngicas o animales. Los ejemplos de células fúngicas adecuadas son células de levadura, preferentemente las del género *Saccharomyces*, y más preferentemente aquellas de la especie *Saccharomyces cerevisiae*. Las células animales adecuadas son, por ejemplo, células de insectos, células de vertebrados, preferentemente 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. Se pueden obtener otras estirpes celulares adecuadas conocidas en la técnica a partir de depósitos de estirpes celulares, como, por ejemplo, la Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH (DSMZ) o la Colección Americana de cultivos tipo (ATCC). De acuerdo con la presente invención, se contempla además que las células/los cultivos celulares primarios puedan funcionar como células hospedadoras. Dichas células derivan, en particular, de insectos (como insectos de la especie *Drosophila* o *Blatta*) o mamíferos (como seres humanos, cerdos, ratones o ratas). Dichas células hospedadoras también pueden comprender células de y/o derivadas de estirpes celulares como líneas celulares de neuroblastoma. Las células primarias mencionadas anteriormente son bien conocidas en la técnica, y comprenden, entre otros, astrocitos primarios, cultivos espinales (mixtos) o cultivos de hipocampo. La presente invención también se refiere a métodos de producción de la molécula de ARN de la presente invención cultivando una célula hospedadora que alberga un vector de expresión que codifica 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 de la célula hospedadora o del medio de cultivo. La presente invención también se puede referir a un método de producción de una molécula de ARN de la presente invención que comprende el cultivo de la célula hospedadora de la presente invención y, opcionalmente, la recuperación de la molécula de ARN del cultivo. Los métodos de recuperación y/o, posteriormente, purificación de la molécula de ARN de la presente invención son conocidos por los expertos en la materia.

La presente invención también se refiere a métodos de producción, en una reacción *in vitro*, de la molécula de ARN de la presente invención mediante métodos conocidos por los expertos en la materia. Más concretamente, la molécula de ARN de la presente invención se puede producir *in vitro* usando un sistema de transcripción *in vitro*. Los sistemas de transcripción *in vitro* son comúnmente conocidos y normalmente requieren un molde de ADN lineal purificado que contenga una secuencia de ADN "que codifique" el módulo (b) y/o módulo (c) como se ha descrito anteriormente, en los que dicha secuencia de ADN está bajo el control de un promotor apropiado. Además, un sistema de transcripción *in vitro* también requiere comúnmente ribonucleótidos trifosfato, un sistema tampón que incluye DTT e iones de magnesio y una ARN polimerasa apropiada que proporcione la actividad enzimática para la transcripción *in vitro* de la secuencia de ADN a la molécula de ARN de la presente invención. Los métodos que se usan comúnmente para producir moléculas de ARN usando la transcripción *in vitro* son bien conocidos por los expertos en la materia y se describen, por ejemplo, en *Methods Mol. Biol.* 703 (2011):29-41. Como se ha mencionado anteriormente, en caso de que la molécula de ARN de la presente invención se produzca mediante un método de transcripción *in vitro* como se describe en el presente documento más adelante, la cola de poli-A anterior puede formar parte de la construcción de la presente invención (y no necesariamente situada originalmente en el vector de clonación) y estar situada en el extremo 3' del ARN adyacente a la UTR en el extremo 3' de la construcción de ARN. En caso de que la molécula de ARN de la presente invención se produzca 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* cadena abajo de la cola de poli-A para garantizar que el la molécula de ARN transcrita contenga dicha cola de poli-A. Como alternativa, la molécula de ARN de la presente invención también se puede sintetizar químicamente, por ejemplo, mediante síntesis química convencional en un sintetizador automático de secuencia de nucleótidos usando un soporte en fase sólida y técnicas convencionales. La presente invención también se refiere a 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 materia y como se ha descrito anteriormente, y recuperando la molécula de ARN de la reacción. Los métodos de recuperación y/o, posteriormente, purificación de la molécula de ARN de la presente invención son conocidos por los expertos en la materia.

Las moléculas de ARN definidas anteriormente son particularmente útiles en entornos médicos y en el tratamiento de una determinada enfermedad y, en particular, en terapias basadas en ARN. Por lo 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 hospedadora de la presente invención y, opcionalmente, un vehículo farmacéuticamente aceptable.

El término "tratamiento" y similares se usan en el presente documento para referirse, en general, a la obtención de 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 términos de prevenir completa o parcialmente una enfermedad o un síntoma de la misma. Preferentemente, el término "tratamiento" debe entenderse como terapéutico en términos de curar parcial o completamente una enfermedad y/o un efecto adverso y/o síntomas atribuidos a la enfermedad. "Agudo", en este sentido, significa que el sujeto muestra síntomas de la enfermedad. En otras palabras, el sujeto que se va a tratar necesita realmente un tratamiento, y la expresión "tratamiento agudo" en el contexto de la presente invención se refiere a las medidas tomadas para tratar realmente la enfermedad tras la aparición de la enfermedad o la manifestación de la enfermedad. El tratamiento también puede ser un tratamiento profiláctico o preventivo, es decir, medidas tomadas para la prevención de enfermedades, por ejemplo, para prevenir la infección y/o la aparición de la enfermedad.

La composición farmacéutica de la presente invención se puede administrar a través de una amplia selección de clases de formas de administración conocidas por el experto. La administración puede ser sistémica, local, oral, a través de aerosoles que incluyen, pero sin limitación, comprimidos, inyección de aguja, el uso de inhaladores, cremas, espumas, geles, lociones y pomadas.

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 hospedadora) de la presente invención de acuerdo con lo anterior y al menos un excipiente o vehículo farmacéuticamente aceptable.

Un excipiente o vehículo es una sustancia inactiva formulada junto con el principio activo, es decir, una construcción de la presente invención de acuerdo con lo anterior, con el fin de formulaciones de aumento de volumen que contengan principios activos potentes. Los excipientes se suelen denominar "agentes formadores de volumen", "cargas" o "diluyentes". El aumento de volumen permite una dispensación conveniente y precisa de una sustancia farmacológica cuando se produce una forma farmacéutica. También pueden servir para diversos fines terapéuticos de mejora, tales como facilitar la absorción o solubilidad del fármaco u otras consideraciones farmacocinéticas. Los excipientes también pueden ser útiles en el proceso de fabricación, para ayudar a la manipulación de la sustancia activa en cuestión, facilitando la fluidez del polvo o propiedades antiadherentes, además de ayudar a la estabilidad *in vitro*, tal como la prevención de la desnaturalización durante la vida útil esperada. La selección de los excipientes apropiados también depende de la vía de administración y a la forma farmacéutica, así como del principio activo y de otros factores.

Por lo tanto, de acuerdo con lo anterior, la composición farmacéutica que comprende una cantidad eficaz del ácido nucleico de la presente invención puede estar en forma sólida, líquida o gaseosa y puede estar, entre otras, en forma de (a) polvo/s, (a) comprimido/s, (a) solución/es o (a) aerosol/es. Se prefiere que dicha composición farmacéutica comprenda opcionalmente un vehículo y/o diluyente farmacéuticamente aceptable.

Los ejemplos de vehículos, excipientes y/o diluyentes farmacéuticos adecuados son bien conocidos en la técnica e incluyen soluciones salinas tamponadas con fosfato, agua, emulsiones, tales como emulsiones de aceite/agua, diversos tipos de agentes humectantes, soluciones estériles, etc. Las composiciones que comprenden dichos vehículos pueden formularse mediante métodos convencionales bien conocidos. Estas composiciones farmacéuticas se pueden administrar al sujeto a una dosis adecuada, es decir, en "una cantidad eficaz" que puede ser determinada fácilmente por el experto mediante métodos conocidos en la técnica. La pauta posológica será determinada por el médico tratante y los factores clínicos. Como es bien sabido en las técnicas médicas, las dosis para cualquier paciente dependen de muchos factores, incluyendo el tamaño del paciente o del sujeto, el área de superficie corporal, la edad, el compuesto en particular por administrar, sexo, tiempo y vía de administración, el estado de salud general y otros fármacos administrados simultáneamente.

Por lo tanto, preferentemente, la construcción de la presente invención se incluye en una cantidad eficaz. La expresión "cantidad eficaz" se refiere a una cantidad suficiente para inducir una respuesta terapéutica detectable en el sujeto a quien se va a administrar la composición farmacéutica. De acuerdo con lo anterior, el contenido de la construcción de la presente invención en la composición farmacéutica no se limita en la medida en que sea útil para el tratamiento como se ha descrito anteriormente, pero preferentemente contiene del 0,0000001 al 10 % en peso por composición total. Además, la construcción descrita en el presente documento se emplea preferentemente en un vehículo. En general, se usa una cantidad apropiada de una sal farmacéuticamente aceptable en el vehículo para hacer que la composición sea isotónica. Los ejemplos del vehículo incluyen, pero sin limitación, solución salina, solución de Ringer y solución de dextrosa. Preferentemente, los excipientes, vehículos o estabilizantes aceptables no son tóxicos a las dosis y concentraciones empleadas, incluyendo tampones tales como citrato, fosfato y otros

ácidos orgánicos; contraiones formadores de sales, por ejemplo, de sodio y potasio; polipéptidos de bajo peso molecular (> 10 restos de aminoácidos); 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 que incluyen 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 que incluyen metionina, ácido ascórbico y tocoferol; y/o conservantes, por ejemplo, cloruro de octadecildimetilbencilamonio; cloruro de hexametonio; cloruro de benzalconio, cloruro de bencetonio; alcohol fenólico, butílico o bencílico; alquilparabenos, por ejemplo, metil- o propil-parabeno; catecol; resorcinol; ciclohexanol; 3-pentanol; y m-cresol). Los vehículos adecuados y sus formulaciones se describen con mayor detalle en "Remington's Pharmaceutical Sciences", 17ª ed., 1985, Mack Publishing Co.

El progreso terapéutico se puede controlar mediante una evaluación periódica. La molécula de ARN de la presente invención o la composición farmacéutica de la invención pueden estar en soluciones, suspensiones y emulsiones acuosas o no acuosas estériles, así como en cremas y supositorios. Los ejemplos de disolventes no acuosos son propilenglicol, polietilenglicol, aceites vegetales tales como aceite de oliva y ésteres orgánicos tales como oleato de etilo. Los vehículos acuosos incluyen agua, soluciones, emulsiones o suspensiones alcohólicas/acuosas, que incluyen solución salina y medios tamponados. También pueden estar presentes conservantes y otros aditivos tales como, por ejemplo, antimicrobianos, antioxidantes, agentes quelantes y gases inertes y similares. Además, la composición farmacéutica de la invención puede comprender otros agentes dependiendo del uso previsto de la composición farmacéutica. Dichos agentes pueden ser, por ejemplo, monolaurato de polioxietilensorbitán, disponibles en el mercado con el nombre comercial Tween, propilenglicol, EDTA, citrato, sacarosa, así como otros agentes que son adecuados para el uso previsto de la composición farmacéutica que son bien conocidos por el experto en la materia.

De acuerdo con la presente invención, la expresión "composición farmacéutica" se refiere a una composición para administración a un paciente, preferentemente un paciente humano.

La composición farmacéutica de la presente invención puede ser para uso en terapias basadas en ARN. Como se ha mencionado anteriormente, la molécula de ARN de la presente invención que comprende una "región de codificación que codifica un polipéptido" puede usarse en terapias basadas en ARN, en la que la "región de codificación que codifica un polipéptido" codifica un polipéptido o una proteína terapéutica o farmacéuticamente activa que tiene un efecto terapéutico o preventivo. Por lo tanto, en realizaciones preferidas, 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 como se ha mencionado en la **Tabla 2** anterior. Por consiguiente, las terapias basadas en ARN de acuerdo con la presente invención pueden ser para uso en el tratamiento o en la prevención de una enfermedad como las mencionadas en la **Tabla 2** anterior. Por lo tanto, la composición farmacéutica de la presente invención puede usarse en terapias basadas en ARN en casos en los que los defectos genéticos descritos en la **Tabla 2** anterior conduzcan a una enfermedad que luego pueda tratarse o prevenirse mediante una terapia de reemplazo de transcripciones/terapia de reemplazo de enzimas con la molécula de ARN de la presente invención, en las que la molécula de ARN comprende una "región de codificación de un polipéptido" que codifica una versión intacta de la proteína o un fragmento funcional de la misma que compensa el gen defectuoso desvelado. En realizaciones particularmente preferidas, la composición farmacéutica de la presente invención puede usarse en terapias basadas en ARN en el tratamiento o en la prevención de enfermedades lisosomales tales como la enfermedad de Gaucher, enfermedad de Fabry, MPS I, MPS II (síndrome de Hunter), MPS VI y enfermedades de almacenamiento de glicógeno tales como, por ejemplo, enfermedad de almacenamiento de glicógeno de tipo I (enfermedad de von Giercke), de tipo II (enfermedad de Pompe), de tipo III (enfermedad de Cori), de tipo IV (enfermedad de Andersen), de tipo V (enfermedad de McArdle), de tipo VI (enfermedad de Hers), de tipo VII (enfermedad de Tauri), tipo VII, tipo IX, tipo X, tipo XI (síndrome de Fanconi-Bickel), tipo XI o tipo 0. Las terapias de reemplazo de la transcripción/terapias de reemplazo de enzimas beneficiosamente no afectan al defecto genético subyacente, pero aumentan la concentración de la enzima en la que el paciente es deficiente. Como ejemplo, en la enfermedad de Pompe, la terapia de reemplazo de la transcripción/terapia de reemplazo de enzima reemplaza la enzima lisosomal deficiente alfa-glucosidasa ácida (GAA). En otras realizaciones preferidas, la composición farmacéutica de la presente invención puede usarse en terapias basadas en ARN de acuerdo con la presente invención, en las que la "región de codificación que codifica un polipéptido" codifica un polipéptido, una proteína o un péptido terapéutica o farmacéuticamente activo que tiene un efecto terapéutico o preventivo, en las que dicho polipéptido, proteína o péptido se selecciona del grupo codificado por los genes descritos en la **Tabla 2**.

En otras realizaciones preferidas, las terapias basadas en ARN de acuerdo con la presente invención pueden usarse para tratar el cáncer, una enfermedad cardiovascular, una infección vírica, una disfunción inmune, una enfermedad autoinmune, un trastorno neurológico, trastornos metabólicos hereditarios o un trastorno genético o cualquier enfermedad en la que una proteína o fragmento de proteína producida en una célula pueda 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 hueso, cáncer de cerebro, cáncer de cuello uterino, cáncer anal, cáncer de colon, cáncer colorrectal, cáncer de apéndice, cáncer de ojo, cáncer gástrico, leucemia, linfoma, cáncer de hígado, cáncer de piel, cáncer de ovario, cáncer de pene, cáncer de páncreas, cáncer testicular, cáncer de tiroides, cáncer vaginal, cáncer vulvar, cáncer endometrial, cáncer cardíaco y sarcoma. Los

ejemplos de enfermedades cardiovasculares incluyen aterosclerosis, enfermedad coronaria, enfermedad cardíaca pulmonar y cardiomiopatía. Los ejemplos de disfunciones inmunes y enfermedades autoinmunes incluyen, pero sin limitación, enfermedades reumáticas, esclerosis múltiple y asma. Los ejemplos de infecciones virales incluyen, pero sin limitación, infecciones con virus de inmunodeficiencia humana, virus del herpes simple, virus del papiloma humano, así como virus de la hepatitis B y C.

Los ejemplos de trastornos neurológicos incluyen, pero sin limitación, enfermedad de Parkinson, esclerosis múltiple y demencia. Los ejemplos de trastornos metabólicos hereditarios incluyen, pero sin limitación, la enfermedad de Gaucher y fenilcetonuria.

La invención también se refiere a un método de una terapia basada en ARN. Por lo tanto, la presente invención se refiere a un método de tratamiento de una enfermedad tal como cáncer, una enfermedad cardiovascular, una infección vírica, una disfunción inmune, una enfermedad autoinmune, un trastorno neurológico, trastornos metabólicos hereditarios o un trastorno genético mediante una terapia basada en ARN. Con respecto a las realizaciones preferidas del método para el tratamiento, lo mismo se aplica, cambiando lo que corresponda, como se ha expuesto anteriormente en el contexto de la molécula de ARN o la composición farmacéutica para su uso en la terapia basada en ARN como se ha definido anteriormente.

En la presente invención, el sujeto es, en una realización preferida, 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 más preferida, el sujeto es un ser humano.

La presente invención también se refiere a un kit 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 hospedadora de la presente invención. Con respecto a las realizaciones preferidas, lo mismo se aplica, cambiando lo que corresponda, como se ha expuesto anteriormente en el contexto de la molécula de ARN, molécula de ácido nucleico, vector o célula hospedadora de acuerdo con la presente invención. Ventajosamente, el kit de la presente invención comprende además, opcionalmente, (a) tampón/es, soluciones de almacenamiento y/o reactivos restantes o materiales requeridos para llevar a cabo los usos y métodos anteriores y posteriores. Además, las partes del kit de la invención se pueden envasar individualmente en viales o botes o en combinación en recipientes o unidades de múltiples recipientes. El kit de la presente invención se puede usar ventajosamente, entre otras cosas, para llevar a cabo los métodos de la invención, la preparación de la molécula de ARN de la invención y se podría emplear en una variedad de aplicaciones a las que se hace referencia en el presente documento, por ejemplo, en los usos resumidos anteriormente y a continuación. Otro componente que se puede incluir en el kit son las instrucciones de uso a una persona que usa un kit. La fabricación de los kits sigue preferentemente procedimientos convencionales que son conocidos por los expertos en la materia.

Finalmente, la presente invención también se refiere al uso de una o más UTR que comprenden la secuencia mostrada en SEQ ID NO: 1 o una secuencia que muestra de 1 a 4 sustituciones en comparación con SEQ ID NO: 1 y que da lugar a una molécula de ARN que tiene la misma o mayor eficacia de traducción que una molécula de ARN que comprende una UTR que comprende la SEQ ID NO: 1; y/o de una o más UTR que comprenden la secuencia mostrada en SEQ ID NO: 2 o una secuencia que muestra de 1 a 7 sustituciones en comparación con SEQ ID NO: 2 y que da lugar a una molécula de ARN que tiene la misma o una mayor eficacia de traducción que una molécula de ARN que comprende una UTR que comprende la SEQ ID NO: 2 para aumentar la eficacia de traducción de una región de codificación de una molécula de ARN a un polipéptido o una proteína codificados por dicha región de codificación como se ha definido anteriormente. Con respecto a las realizaciones preferidas del uso, lo mismo se aplica, cambiando lo que corresponda, como se ha expuesto anteriormente, en el contexto de la molécula de ARN de la presente invención.

Figura 1: Microscopía de fluorescencia y datos de citometría de flujo de células A549.

(A) Ilustración esquemática del ARNm terapéutico, que consiste en un capuchón 5', una UTR 5', una región de codificación, una UTR 3' y una cola de poli-A.

(B) Imágenes de microscopía de fluorescencia tomadas con 4 aumentos (JULY™) a las 24 h de la transfección. Todas las construcciones mostraron niveles de expresión proteica mejorados en comparación con el control.

(C) El porcentaje de células positivas para d2EGFP determinado por FC es similar para todas las construcciones. Se usó yoduro de propidio para detectar las células muertas. Las ventanas de análisis aplicadas aseguraron la exclusión de las células muertas y las células no transfectadas.

(D) A las 48 h de la transfección, la expresión de proteína sostenida fue mayor para las construcciones estabilizadas en comparación con el control.

Figura 2: Intervalos de tiempo de expresión de proteína según lo determinado por FC para células A549 (A) y células Huh7 (B). Las intensidades medias de fluorescencia normalizadas para el control se representan gráficamente frente al tiempo en una representación logarítmica-lineal. Con el aumento del tiempo tras la transfección, los niveles elevados de expresión proteica de las construcciones estabilizadas se vuelven cada vez más evidentes. Las barras correspondientes al control, las construcciones de UTR 5' y UTR 3', respectivamente, así

como a las construcciones 5'+3', 5'+2x3' y 2x3' están sombreadas de diferente manera, como se muestra en el lado derecho de la figura.

Figura 3: Portaobjetos multicanal microestructurados para ensayos paralelos de células individuales para ensayar construcciones de ARNm estabilizadas de forma diferente.

(A) Los patrones de proteínas microestructuradas y adhesivas a las células con áreas de PEG repelentes de las células permiten la disposición ordenada de las células. Se usó fibronectina marcada fluorescentemente para visualizar el micropatrón.

(B) células fluorescentes A549 que se adhieren a los patrones de fibronectina dentro de un microcanal (tres horas después de la siembra).

(C) Dibujo esquemático de la lipofección de ARNm (a la izquierda) y esquema de reacción subyacente a la presente solución analítica (a la derecha).

(D) Cursos de tiempo ilustrativos de expresión de d2EGFP mediada por ARNm en células A549. Las líneas negras son ajustes representativos del modelo teórico de traducción.

Figura 4: Distribuciones de las velocidades de expresión K, tiempos de vida del ARNm y tiempos de vida de d2EGFP y los valores medios correspondientes con representaciones esquemáticas de las construcciones.

(A) Distribuciones de la velocidad de expresión K, que es el producto del número inicial de moléculas de ARNm y las velocidades de traducción. El hecho de que las distribuciones tengan una forma similar indica que la cinética de transfección y las velocidades de traducción son muy similares.

(B) Las distribuciones de las semividas de ARNm muestran grandes variaciones en su amplitud. Como una guía para el ojo, las líneas de puntos indican la semivida del control.

(C) Distribuciones de las semividas de d2EGFP. Como era de esperar, las distribuciones de las diferentes construcciones tienen una forma similar y muestran valores medios comparables. Como guía para el ojo, la semivida global de d2EGFP basada en todas las semividas medidas se muestra como una línea de puntos.

(D) Valores medios y las desviaciones típicas correspondientes (DT) de las tasas ajustadas. Aunque la construcción de control produce altos valores medios de K en ambos tipos de células, la corta semivida del ARNm de esta construcción conduce a valores de AUC bajos en comparación con las construcciones estabilizadas. Esto se puede ver en la **Figura 6**. Las representaciones esquemáticas de las construcciones se pueden ver en el lado de la derecha. Todas las construcciones tienen el mismo capuchón 5' y una cola de poli-A. Se analizaron los datos de 895 células A549 y 1.355 células Huh7 individuales.

Figura 5: Curvas maestras de las diferentes construcciones. Medias de la población de células A549 (A) y Huh7 (B) con el tiempo de inicio cambiado a cero. Las curvas gris oscuro, gris medio y gris claro corresponden a las construcciones de control/UTR 5'/UTR 3', respectivamente. Las curvas corresponden a las construcciones indicadas correspondientemente en el lado de la derecha.

Figura 6: AUC y factores de prolongación del tiempo de vida de ARNm de las diferentes construcciones.

(A) Representación esquemática de la AUC para ilustrar la interacción entre la traducción de ARNm y la degradación del ARNm y de la proteína.

(B) y (C) AUC de las diferentes construcciones analizadas para $t \rightarrow \infty$. Las cruces muestran las AUC relativas de diferentes experimentos, las barras corresponden a la media de todas las AUC de células individuales.

(D) y (E) factores de prolongación del tiempo de vida del ARNm. Todas las modificaciones producen tiempos de vida del ARNm prolongados en comparación con el control. Se observan tendencias similares en las células A549 (D) y Huh7 (E). Las barras de error en (D) y (E) indican una desviación típica.

Figura 7: Microscopía de fluorescencia y datos de citometría de flujo de células Huh7.

(A) Imágenes de microscopía de fluorescencia tomadas con 4 aumentos (JULY™) a las 24 h de la transfección. Todas las construcciones mostraron niveles de expresión de proteína mejorados en comparación con el control.

(B) El porcentaje de células positivas para d2EGFP de acuerdo con lo determinado por FC es similar para todas las construcciones. Se usó yoduro de propidio para detectar las células muertas. Las ventanas de análisis aplicadas garantizaron la exclusión de las células muertas y las células no transfectadas.

(C) A las 48 h de la transfección, la expresión de proteína sostenida fue mayor para las construcciones estabilizadas en comparación con el control.

Figura 8: Determinación de la semivida del ARNm mediante qRT-PCR en células A549 y Huh7. Las células se transfectaron de acuerdo con el protocolo descrito en la parte de Materiales y métodos. Se determinó la cuantificación absoluta del ARNm a las 4, 8, 24, 36, 48, 60, 72 horas para todas las construcciones de ARNm en A549 (véase la Figura 8 A) y en Huh7 (véase la Figura 8 B). De estos datos, se calculó la semivida del ARNm. La semivida física se normalizó para el control.

Figura 9: Eficacias de transfección en sustratos microestructurados. El porcentaje de células transfectadas y las

desviaciones típicas correspondientes para células A549 y células Huh7 transfectadas con ARN de SNIM con ayuda de Lipofectamine™ 2000 o DOGTOR. Se encontraron eficacias de transfección más altas para las células transfectadas con Lipofectamine™ 2000.

5 **Figura 10:** Distribuciones de semividas de d2EGFP medidas directamente.

(A) Cursos de tiempo ilustrativos de la degradación de d2EGFP inducida por cicloheximida en células Huh7. Las líneas negras son ajustes exponenciales simples para la degradación de proteínas.

10 (B) Distribución de semividas de d2EGFP medidas en células A549, produciendo una semivida media de 2,46 h (desviación típica de 0,71 h).

(C) Distribución de semividas de d2EGFP medidas en células Huh7, produciendo una semivida media de 4,04 h (desviación típica de 1,82 h).

15 **Figura 11:** Distribución de las AUC de células individuales. Las AUC se calcularon de acuerdo con la ecuación 3 que se presenta más adelante. Los datos de A549 se muestran en la columna de la izquierda, los datos de Huh7 se muestran en la columna de la derecha.

20 **Figura 12:** Comparación de las construcciones n.º 2 o n.º 5 que tienen UTR de diferentes genes como se indica en la **Tabla 5** con la construcción n.º 1 de UTR de CYBA.

Otros aspectos y ventajas de la invención se describirán en los siguientes ejemplos, que se proporcionan con fines ilustrativos y no a modo de limitación. Cada publicación, patente, solicitud de patente u otro documento citado en la presente solicitud se incorpora en el presente documento por referencia en su totalidad.

25 **Ejemplos**

I. Materiales y métodos

Vectores plasmídicos

30 Se escindió la Proteína Fluorescente Verde Mejorada Desestabilizada (d2EGFP) de pd2EGFP-N1 (Clontech) y se clonó en pVAXA120 (3) para generar pVAXA120-d2EGFP. Basándose en datos previamente publicados con respecto a la estabilidad del ARNm, Eurofins MWG (Alemania) sintetizaron secuencias de UTR 5' y 3' previamente seleccionadas del gen CYBA y se clonaron cadena arriba (UTR 5') y/o cadena abajo (UTR 3' o 2 x UTR 3') de d2EGFP en pVAXA120-d2EGFP, generando de este modo las construcciones con combinaciones de UTR respectivas.

Producción de ARNm

40 Para generar ARNm transcrito *in vitro* (ARNm de IVT), se linealizaron los plásmidos cadena abajo de la cola de poli-A mediante digestión con NotI y se purificaron mediante extracción con cloroformo y precipitación con etanol. Los plásmidos lineales purificados se usaron como molde para la transcripción *in vitro* usando el sistema de T7 de producción de ARN a gran escala RiboMax (Promega, Alemania). Se añadió Análogo de capuchón anti-inverso (ARCA) a la mezcla de reacción para generar ARNm protegido con capuchón en 5'. Además, para la producción de ARNm de SNIM, se añadieron nucleótidos químicamente modificados, en concreto, metil-CTP y tio-UTP (Jena Bioscience, Alemania) a 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 toda la mezcla de IVT a 37 °C durante 2 horas seguido de una digestión de ADN con DNaseI durante 20 minutos a 37 °C. El ARN se precipitó 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, el sedimento de ARN se volvió a suspender en agua exenta de ARNasa. Todos los ARNm se verificaron en geles de agarosa al 1 %. En la **Figura 1A**, se puede ver una representación esquemática de una construcción de ARNm ilustrativa. Las secuencias exactas de las UTR se dan en el texto debajo de la **Tabla 1** anterior.

d2EGFP	ΔG	Extremo 5'	Extremo 3'	UTR 5'	UTR 3'
control	-358,9	unión parcial con CDS (8/8)	suelto (8/8)	ninguna	ninguna
CYBA 5'	-375	unión parcial con UTR de CYBA 5' (7/8)	suelto (8/8)	se une con CDS (6/8)	ninguna
CYBA 3'	-411,6	unión parcial con UTR de CYBA 3' (8/8)	se une con el extremo 5' (4/4)	ninguna	forma una horquilla (7/8)
CYBA 5'+3'	-405,7	se une con UTR de CYBA (3/8)	se une con el extremo 5' (4/4)	se une con UTR de CYBA (4/8)	forma una horquilla (7/8)
CYBA 5'+2x3'	-437,7	se une con UTR	suelto (8/8)	se une con UTR	primera UTR 3':

d2EGFP	ΔG	Extremo 5'	Extremo 3'	UTR 5'	UTR 3'
		3' (8/8)		3' y el gen (6/8)	horquilla; segunda UTR 3': horquilla (7/8)
CYBA 2x3'	-444,1	se une consigo mismo y forma horquilla (8/8)	suelto (7/8)	ninguna	primera UTR 3': horquilla segunda UTR 3': dos horquillas (3/8)

En la **Tabla 3**, se enumeran las características de las construcciones de ARNm tales como la energía mínima libre (ΔG) y las estructuras secundarias encontradas en ambos extremos y dentro de las UTR. Se usó la plataforma plegable de m veces para predecir estructuras secundarias de ARNm (40). Para cada construcción, se compararon las ocho estructuras secundarias que tienen la mayor energía libre. Los valores más altos de energía libre se pronostican para las construcciones 2 x UTR 3' y UTR 3'. El extremo 5' de cada construcción de ARNm se une parcialmente con la UTR 3' o la UTR 5', excepto para la construcción de control, que se une a la secuencia de codificación (cds). Curiosamente, el extremo 5' de la construcción de ARNm 2x3' forma una horquilla estabilizadora consigo misma. Sin embargo, los bucles de la horquilla cerca del extremo 5' también pueden dificultar la traducción de proteínas (41). Se encontró otra característica en el extremo 3' de las construcciones de ARNm de UTR 3' y UTR 5'+ 3': allí, el extremo 3' se une con el extremo 5', minimizando la distancia entre sí y permitiendo así un inicio más rápido de la traducción. A diferencia de las UTR 5', la UTR 3' de cada construcción de ARNm forma al menos una horquilla consigo misma.

15 Citometría de flujo (FC)

La configuración experimental es la siguiente: se sembraron 20.000 células en 150 μ l de medio por pocillo en placas de 96 pocillos y se transfectaron 24 horas después de la siembra. Las células se transfectaron a una dosis de 5 μ g de ARNm/célula usando el reactivo de transfección comercial Lipofectamine™ 2000 a una proporción de 2,5 μ l de Lipofectamine™ 2000 por 1 μ g de ARNm. Para la formación de lipoplexos, se diluyeron Lipofectamine™ 2000 y ARNm por separado en medio de transfección OptiMEM en un volumen total de 50 μ l, cada uno. Estas mezclas se incubaron a temperatura ambiente durante 5 minutos. Luego, se mezcló la solución de ARNm con la solución Lipofectamine™ 2000, seguido de otros 20 minutos de incubación a temperatura ambiente. Tras la incubación, se añadieron 900 μ l de OptiMEM a la solución de lipoplex. Finalmente, se añadieron 50 μ l de la solución compleja a las células y se incubaron durante 1 hora. Para cada construcción de ARNm, se prepararon triplicados biológicos. Tras la incubación, se desechó la solución de lipoplex y se añadieron 150 μ l de medio nuevo a cada pocillo. La expresión de d2EGFP se midió después de 8, 24, 36, 48, 60 y 72 horas usando FC. Se tomaron imágenes de microscopía de fluorescencia en cada uno de estos puntos de tiempo. Para las mediciones de FC, se desechó el medio de cultivo celular y las células se lavaron con 1 x DPBS (Gibco Life Technology). Posteriormente, se añadieron 20 μ l de TrypLE Express (tecnología Gibco Life) por pocillo y se incubaron durante 5 minutos a 37 °C. La reacción se neutralizó añadiendo 80 μ l de 1 x PBS, complementado con FBS al 2 %. Las células se mezclaron pipeteando y se transfirieron a una placa de 96 pocillos apropiada para mediciones de citometría de flujo. Finalmente, se añadieron 5 μ l de yoduro de propidio (concentración final de 1 μ g/ml) por pocillo y se midieron con el muestreador automático Attune (Applied Biosystems). Las imágenes de fluorescencia se tomaron antes del análisis de FC con un microscopio JULY™.

PCR cuantitativa en tiempo real

Se usó un análisis de qRT-PCR para determinar la cantidad de ARNm de d2EGFP a intervalos de tiempo de 4, 8, 24, 36, 48, 60 y 72 horas en células A549 y Huh7. Además, se usó la cinética de expresión de ARNm en sí para calcular la semivida del ARNm de cada UTR. En el presente documento, las células se transfectaron de forma similar al protocolo descrito anteriormente (véase la FC). Se encontró que una densidad celular de 200.000 células/pocillo era suficiente para el aislamiento del ARN. El aislamiento del ARN se realizó de acuerdo con el protocolo del fabricante usando ARN NucleoSpin (Macherey Nagel). El ARN total aislado se examinó a la concentración y la calidad del ARN mediante mediciones espectrofotométricas y análisis en gel. Además, se usaron 0,5 μ g del ARN total de cada construcción de UTR y el control para la síntesis de ADNc usando Oligo (dT) del kit de síntesis de ADNc First Strand (Thermo Scientific). Se ensayaron cantidades equivalentes de ADNc (diluido 1:50) con 125 nM de cada cebador de d2EGFP (cebador directo: 5'-CAA CCA CTA CCT GAG CAC CC-3' (**SEQ ID NO: 3**); cebador inverso: 5'-GTC CAT GCC GAG AGT GAT CC-3' (**SEQ ID NO:4**)) usando SsoAdvanced™ Universal SYBR® Green Supermix (BioRad). Como patrón para la cuantificación absoluta, se usó el ARNm de d2EGFP puro producido por IVT para la síntesis de ADNc. La cuantificación absoluta de ARNm se realizó en un dispositivo Lightcycler 96 (Roche).

55 Patrones de superficie y preparación de muestras

Las superficies microestructuradas se produjeron mediante tratamiento selectivo con plasma de oxígeno (Femto Diener, 40 W durante 3 min) en una parte superior como sustrato (ibidi GmbH) con la posterior pasivación. La

selectividad se logró usando un sello de polidimetilsiloxano (PDMS) (emitido desde un maestro producido por fotolitografía) como una máscara. Las partes expuestas al plasma se pasivaron mediante incubación durante 30 minutos con PLL(20k)-g(3.5)-PEG(2k) a una concentración de 1 mg/ml en tampón acuoso (HEPES 10 mM, pH 7,4 y NaCl 150 mM). Posteriormente, se enjuagaron las muestras con PBS y se eliminaron los sellos de PDMS. Las láminas se fijaron luego a portaobjetos adhesivos de seis canales (portaobjetos de micrómetros pegajosos VI). Cada canal se llenó con una solución de 50 µg/ml de fibronectina en PBS durante una hora para convertir los sectores restantes en adhesivos celulares. Las sondas se enjuagaron a fondo con PBS tres veces. Las muestras se almacenaron en medio celular a temperatura ambiente antes de la siembra celular. Para este estudio, se usaron sitios de adhesión cuadrados de 30 µm x 30 µm, porque este tamaño resultó ser razonable para la adhesión monocelular de células A549 y Huh7. Las células se sembraron a una densidad de 10.000 células por canal, de modo que aproximadamente una célula podría adherirse a cada isla adhesiva de células. Para obtener micropatrones fluorescentes como se muestra en la **Figura 3A**, se usó una mezcla de 20 µg/ml de fibronectina y 30 µg/ml de fibrinógeno conjugado con Alexa Fluor 488.

15 **Materiales**

Se adquirieron FBS, medio L-15 de Leibovitz (Gibco), Lipofectamine™ 2000 y OptiMEM (Gibco) en Invitrogen, Alemania. El PBS estéril se preparó internamente. El medio F12K de Ham, DMEM y EDTA de tripsina se adquirieron en c.c.pro GmbH, Alemania. Los portaobjetos de canales se adquirieron en ibidi, Alemania. La fibronectina se adquirió en Yo Proteins, Suecia. El PLL-g-PEG se adquirió en SuSoS AG, Suiza. Alexa Fluor 488 se adquirió en Life Technologies, Alemania. El plásmido pd2EGFP-N1 se adquirió en BD Biosciences Clontech, Alemania.

25 **Cultivo de células**

Se cultivó una estirpe celular de adenocarcinoma alveolar humano (A549, ATCC CCL-185) en medio F12K de Ham complementado con SFB al 10 %. Se cultivó una estirpe de células epiteliales de hepatoma humano (Huh7, JCRB0403, banco de células JCRB, Japón) en medio DMEM, suplementado con suero bovino fetal al 10 %. Todas las estirpes celulares se cultivaron en una atmósfera humidificada al nivel de CO₂ del 5 %.

30 **Transfección *in vitro***

Tres horas antes de la transfección, se sembraron 10.000 células por canal en un portaobjetos de 6 canales. Las células se transfectaron a una dosis de 5 µg de ARNm/célula usando el reactivo de transfección comercial Lipofectamine™ 2000 en una proporción de 2,5 µl de Lipofectamine™ 2000 por 1 µg de ARNm. La formación del complejo se preparó de la siguiente manera: se diluyeron Lipofectamine™ 2000 y ARNm por separado en medio de transfección OptiMEM para sumar hasta un volumen total de 45 µl, cada uno. Estas mezclas se incubaron a temperatura ambiente durante 5 minutos. Luego, se mezcló la solución de Lipofectamine™ 2000 con la solución de ARNm, seguido de otros 20 minutos de incubación a temperatura ambiente. Cabe señalar que los microcanales nunca estuvieron vacíos durante todas las etapas de enjuague subsiguientes: inmediatamente antes de la transfección, las células se lavaron con PBS. Finalmente, las soluciones lipoplex que contienen diferentes construcciones de ARNm se llenaron en los seis canales. Las cinco construcciones diferentes de ARNm más la construcción de referencia podrían medirse en las mismas condiciones experimentales. Las células se incubaron en un volumen de transfección total de 90 µl a 37 °C (nivel de CO₂ del 5 %) durante una hora. El medio de transfección se eliminó después y las células se lavaron con PBS. Posteriormente, las células se incubaron de nuevo con medio L-15 de Leibovitz que contenía SFB al 10 %. Se añadió una gota de aceite anti-evaporación (ibidi GmbH, Alemania) en la parte superior de cada depósito de medio antes de la monitorización microscópica de la expresión de d2EGFP.

50 **Adquisición de datos y análisis cuantitativo de imágenes**

La imagen de células vivas se realizó en un microscopio invertido motorizado (Nikon, Eclipse Ti-E) dotado de una lente objetivo (CFI PlanFluor DL-10x, Phase1, NA 0,30; Nikon) y con un marco de montaje con temperatura controlada para la etapa de microscopio. Se usó un sistema de calentamiento ibidi (ibidi GmbH, Alemania) con un controlador de la temperatura para estabilizar la temperatura de las muestras a 37 °C (± 2 °C) a lo largo de las mediciones. Para adquirir imágenes celulares, se usó una cámara CCD enfriada (CLARA-E, Andor). Se usó una fuente de luz de mercurio (Intensilight C-HGFIE, Nikon) para la iluminación y se usó un cubo de filtro con el conjunto de filtro 41024 (Chroma Technology Corp., BP450-490, FT510, LP510-565) para la detección de d2EGFP. Se usó un control de obturación de iluminación para evitar el blanqueamiento. Las imágenes se tomaron con un aumento de 10 veces con un tiempo de exposición constante de 600 ms a intervalos de 10 minutos durante al menos 25 horas después de la transfección. Las imágenes de fluorescencia se consolidaron en archivos de secuencia de una sola imagen. El análisis cuantitativo de los parámetros característicos de la cinética de expresión de una sola célula permite la comparación de diversos rendimientos de vector en términos de eficacia y estabilidad de la expresión. El análisis de la imagen consistió en varias etapas y se realizó usando un software desarrollado internamente basado en ImageJ. En primer lugar, se superpuso una cuadrícula rectangular con la película original de lapso de tiempo y se ajustó al tamaño y la orientación del patrón de celda subyacente. A continuación, el software detectó automáticamente las células que expresaban d2EGFP leyendo las intensidades de fluorescencia de todos los

cuadrados. Se usaron cuadrados desocupados para la corrección de fondo. El software calcula la fluorescencia de las células en toda la secuencia y conecta las intensidades correspondientes a cursos de tiempo de la fluorescencia por célula. Finalmente, se extrajeron las intensidades de fluorescencia de células individuales por cuadrado.

- 5 Luego se analizaron los datos como se ha descrito recientemente ajustando cada curso de tiempo con la solución analítica para la expresión de proteína inducida por ARNm (véase la Ecuación 1) usando el software IgorPro, que es la solución a las ecuaciones diferenciales para ARNm y d2EGFP,

$$\frac{d}{dt} \text{ARNm} = -\delta \cdot m \quad \text{(Ecuación 4)}$$

$$\frac{d}{dt} \text{d2EGFP} = k_{TL} \cdot m - \beta \cdot \text{d2EGFP} \quad \text{(Ecuación 5)}$$

- 10 En la **Figura 3C**, se representa una representación esquemática del modelo simplista subyacente supuesto para la expresión de proteína inducida por ARNm.

II. Ejemplo 1: Microscopía de fluorescencia y análisis mediante citometría de flujo (FC)

- 15 Para evaluar el efecto de diferentes combinaciones de UTR sobre la cinética de la expresión transgénica, se transfirieron dos estirpes celulares diferentes usando Lipofectamine™ 2000 con diferentes construcciones de ARNm de d2EGFP que contenían una UTR 5' sola, una UTR 3', UTR 5'+3', dos copias de UTR 3' y UTR 5'+2 x UTR 3'. En la **Figura 1A**, se puede ver una representación esquemática de los bloques constituyentes de todas las construcciones.

- 20 En diferentes puntos de tiempo hasta tres días después de la transfección, se cuantificó la expresión de d2EGFP usando FC. Un gráfico de puntos ilustrativo para t = 24 h, que ilustra los niveles de expresión de d2EGFP de células A549 vivas, se muestra en la **Figura 1C** (véase la **Figura 7B** para los datos correspondientes de Huh7). Además, se obtuvieron imágenes de las células usando microscopía de fluorescencia (véase la **Figura 1B y D** y la **Figura 7A y C**). Se confirmaron las eficacias de transfección comparables para todas las construcciones de ARNm 24 horas después de la transfección (**Figura 1B y Figura 8A**). Por lo tanto, se pueden descartar las eficacias de transferencia diferencial por ser un factor causal de las diferencias observadas en la cinética de la expresión. Basándose en las imágenes de microscopía de fluorescencia, se detectó una reducción drástica de la expresión de d2EGFP para todas las construcciones a las 48 h después de la transfección (véanse las **Figuras 1B y D, Figuras 7A y C**). Sin embargo, se encontraron niveles de expresión de EGFP más altos con respecto al control para todos los ARNm estabilizados por UTR. Más concretamente, las construcciones de ARNm que contenían UTR 3' parecían mejorar la expresión más que las construcciones sin UTR 3'. Esto se observó para las células A549 y Huh7 (véase la **Figura 1 y Figura 7**, respectivamente). En puntos de tiempo posteriores a 48 h, este efecto se hizo aún más pronunciado (datos no mostrados). En la **Figura 2 A y B**, se muestran los cursos de tiempo de las intensidades medias de fluorescencia (MFI) determinadas por FC para todas las construcciones en ambos tipos de células.

- 35 También en el presente documento, todas las construcciones de ARNm que contienen UTR mostraron valores de MFI más altos que la construcción de control en ambas estirpes celulares en todos los puntos temporales. Tomados en conjunto, la microscopía de fluorescencia y los datos de FC sugieren que las moléculas de ARNm provistas de UTR de CYBA muestran expresión de d2EGFP persistente durante más de 24 horas.

III. Ejemplo 2: PCR cuantitativa en tiempo real

- 45 Se realizó una medición de qRT-PCR como un enfoque adicional para determinar la semivida del ARNm "física" de las diferentes construcciones. La unión de los presentes cebadores seleccionados a d2EGFP tuvo lugar 600 nt cadena abajo del codón de inicio. Por lo tanto, las mediciones de la semivida física del ARNm comprometen tanto los ARNm intactos como los que han sido desprotegidos, pero aún no degradados o tanto desprotegidos como degradados hasta la base 599. También incluye ARNm que se ha eliminado de la combinación de traducción y se ha almacenado en cuerpos P (29-32). Aunque los ARNm intactos contribuyen a la expresión de d2EGFP, el último grupo de transcripciones desprotegidas y/o parcialmente degradadas y las de los cuerpos P no conducen a ninguna expresión. La determinación de la semivida física del ARNm no reveló ninguna prolongación significativa del tiempo de vida de las UTR en comparación con el control en las células A549 y Huh7 (véanse las **Figuras 8A y B**, respectivamente). Curiosamente, en cambio, se observó una disminución en la semivida física del ARNm para construcciones de UTR 5', 3', 5'+ 2x3' y 2x3' en ambas estirpes celulares.

55 **Determinación de la semivida del ARNm mediante qRT-PCR en células A549 y Huh7**

- 60 En un experimento adicional, se investigó la semivida del ARNm de las diferentes construcciones de ARNm con qRT-PCR, que es un enfoque convencional (véase **Figura 8A y B**). Por lo tanto, las construcciones de ARNm se transfirieron como se describe en el presente documento. Al final, se obtuvo la cantidad absoluta de ARNm en cada punto de tiempo específico y se calculó la semivida del ARNm para cada ARNm provisto de UTR. No se observaron

efectos de estabilización del ARNm significativos para ninguna de las construcciones de ARNm seleccionadas en comparación con el control.

IV. Ejemplo 3: matrices de expresión unicelulares

5 Se fabricaron sustratos microestructurados y adhesivos celulares como se muestra en la **Figura 3A y B** como una plataforma para microscopía de lapso de tiempo unicelular

10 Los cuadrados rectangulares se funcionalizan con la proteína de matriz extracelular fibronectina, mientras que la superficie oscura circundante se pasiva con PLL-g-PEG repelente de células. Las células se sembraron a una densidad celular apropiadamente diluida, de manera que después de aproximadamente tres horas, las células se adhirieron a los cuadrados rectangulares. Este proceso de autoorganización celular se ha estudiado en detalle antes (27). El tamaño de los cuadrados fue de 30 µm para un llenado óptimo con células individuales. La distancia entre los cuadrados era lo suficientemente elevada (60 µm) para reducir al mínimo los efectos de puente de las células que se adhieren a más de un cuadrado al mismo tiempo. La microscopía de fluorescencia de lapso de tiempo y el análisis automatizado de imágenes de la señal de fluorescencia por cuadrado producen cientos de cursos de tiempo individuales. En la **Figura 3D**, se muestra un conjunto típico de datos brutos corregidos para el fondo. Las líneas negras representan ajustes ilustrativos a la expresión matemática para la traducción del ARNm (véase también el apartado de Materiales y métodos). Los datos se analizaron como se ha descrito recientemente (26) ajustando cada curso de tiempo con la solución analítica para la expresión de proteína inducida por ARNm,

$$G_{d2EGFP}(t) = \frac{K}{\delta - \beta} \cdot (1 - e^{-(\delta - \beta)(t - t_0)}) \cdot e^{-\beta(t - t_0)} \tag{Ecuación 1}$$

25 usando el software IgorPro. En este caso, G indica la cantidad de proteína, K es la velocidad de expresión, δ es la velocidad de degradación del ARNm y β es la velocidad de degradación de la proteína indicadora d2EGFP. La velocidad de expresión $K = m_0 \cdot k_{TL}$ es el producto de la cantidad inicial de moléculas de ARNm dentro de la célula (m_0) y la velocidad de traducción k_{TL} . El curso de tiempo que se describe en la Ecuación 1 se analizará en detalle en el apartado "curvas maestras de expresión de proteínas".

V. Ejemplo 4: Transfección *in vitro* en matrices de células

30 En un experimento típico, se permitió que las células se adhirieran a los micropatrones durante tres horas antes de la transfección. Cada uno de los seis microcanales se llenó con una solución lipoplex diferente, que contenía una de las construcciones de interés. En experimentos iniciales, se compararon dos reactivos de transfección diferentes, disponibles en el mercado (en concreto, Lipofectamine™ 2000 y DOGTOR). Se encontraron eficacias de transfección más altas para Lipofectamine™ 2000 que para DOGTOR (véase la **Figura 9**). Debido a que se obtuvieron tasas de viabilidad celular elevadas obtenidas por encima del 80 % con Lipofectamine™ 2000 (datos no mostrados), todos los experimentos de transfección adicionales se realizaron usando Lipofectamine™ 2000. Como la expresión de la proteína mediada por ARNm comienza poco después de la transfección, el tiempo de incubación se mantuvo al mínimo. Por consiguiente, se ajustó la proporción entre la dosis de ARNm y el tiempo de incubación para lograr altas eficacias de transfección (véase también la **Figura 9**) y efectos tóxicos insignificantes provocados por la sobreexpresión de la proteína indicadora. A una dosis de ARNm de 5 pg/célula, se encontró que un tiempo de incubación de una hora era óptimo.

Eficacias de transfección en sustratos microestructurados

45 Se evaluó el porcentaje de células transfectadas con éxito para comparar dos agentes de transfección diferentes y garantizar que las eficacias de transfección no se vieran obstaculizadas por el crecimiento celular microestructurado (véase la **Figura 9**). En el presente documento, todas las células crecieron en matrices de proteínas microestructuradas. Se obtuvieron mayores eficacias de transfección para Lipofectamine™ 2000 en comparación con DOGTOR. Usando un ensayo comercial de viabilidad de células vivas/muertas (Molecular Probes, Alemania), se encontraron tasas de viabilidad celular elevadas por encima del 80 % (datos no mostrados).

VI. Ejemplo 5: Velocidades de expresión

55 Todos los resultados para los dos tipos de células se basan en cuatro mediciones independientes en las mismas condiciones experimentales. Se han analizado datos de lapso de tiempo de aproximadamente mil células A549 y miles de células Huh7. Las distribuciones de las velocidades de expresión K obtenidas se muestran en la **Figura 4A** y los valores medios correspondientes se pueden ver en la **Figura 4D**.

60 Se encontró que tanto las velocidades de expresión medias como la forma de sus distribuciones eran bastante

similares para las diferentes construcciones.

VII. Ejemplo 6: Semividas del ARNm

- 5 Se convirtieron las velocidades de degradación del ARNm ajustadas δ en semividas de ARNm de acuerdo con:

$$\tau = \frac{\ln 2}{\delta}$$

(Ecuación 2)

- 10 La **Figura 4B** muestra las distribuciones de semivida de construcciones de ARNm estabilizadas de forma diferente en células A549 y Huh7, respectivamente. En este caso, se hace evidente que para las construcciones estabilizadas, tanto la semivida como la amplitud de la distribución subyacente aumentan en comparación con la construcción de referencia.

- 15 En la **Figura 4D**, se da una visión general de todas las semividas determinadas. Tanto para A549 como para las células Huh7, se encontraron semividas más largas para los ARNm estabilizados por elementos de UTR en comparación con la construcción de control (5,8 horas para las células A549 y 7,8 horas para las células Huh7) que no contiene ninguna UTR estabilizadora. El efecto de prolongación del tiempo de vida fue más pronunciado en las células A549.

20 VIII. Ejemplo 7: Semividas de proteínas

- Las distribuciones de los tiempos de vida de degradación de proteína (d2EGFP) se presentan en la **Figura 4C**. Como era de esperar, las semividas de la proteína expresada no varían para las diferentes construcciones de ARNm. Los tiempos de vida medios determinados varían de 4,2 a 4,9 horas para las células A549 y de 5,6 a 8,5 horas para las células Huh7, como se muestra en la **Figura 4D**. Los coeficientes de variación son de aproximadamente 0,29 (A549) y 0,45 (Huh7) y, por lo tanto, son significativamente más bajos que el coeficiente de variación de hasta 0,6 que se encontró para la distribución en los tiempos de vida del ARNm. Como control, también se midieron las semividas en un enfoque alternativo, en el que la traducción se inhibió mediante la adición de cicloheximida en un punto de tiempo dado, t_0 , después de la transfección (véase la **Figura 10**). En este caso, la expresión de proteína se induce por un tiempo y luego se detiene. La disminución exponencial de la fluorescencia después de la inhibición produce tiempos de vida de proteína. Se encontró que estas semividas eran aproximadamente dos veces más bajas, en comparación con los experimentos anteriores sin inhibición. Sin embargo, en ambos experimentos, las proporciones relativas de los tiempos de vida de proteína en células Huh7 en comparación con las células A549 son las mismas.

35 Velocidad de degradación de la proteína indicadora

- Para comprobar las velocidades de degradación de d2EGFP ajustadas, se midió independientemente la velocidad de degradación de d2EGFP dentro de las células A549 y Huh7 en portaobjetos de seis canales microestructurados. La síntesis de proteína fue bloqueada por el antibiótico cicloheximida, que interfiere con la actividad peptidil transferasa (42). Los intervalos de tiempo de intensidad de fluorescencia de una sola célula se controlaron durante aproximadamente 20 h (véase la **Figura 10**). Los experimentos de control aseguraron que la disminución de la intensidad de fluorescencia no se debió al fotoblanqueo del cromóforo. Los intervalos de tiempo unicelulares se ajustaron mediante un solo ajuste exponencial, produciendo distribuciones de velocidades de degradación de proteína. Las velocidades de degradación medias fueron de 0,28/h (DT de 0,08/h) en células A549 y de 0,17/h (DT 0,08/h) en células Huh7, lo que corresponde a tiempos de vida de proteína de 2,46 h y 4,04 h, respectivamente. Aunque estos tiempos de vida son significativamente más cortos que los tiempos de vida determinados mediante el análisis del curso temporal de células individuales de la expresión de proteína mediada por el ARNm, la proporción entre los tiempos de vida medios de d2EGFP en células Huh7 y A549 es la misma (4,04 h/2,46 h = 1,64 medida mediante el bloqueo de la traducción en comparación con 7,4 h/4,5 h = 1,64 según lo determinado ajustando la solución analítica para la expresión de ARNm).

IX. Ejemplo 8: Curvas maestras de expresión de proteínas

- 55 Las características de la expresión de proteína inducida por ARNm se hacen evidentes en la denominada curva maestra de expresión de proteína como se representa en la **Figura 5A** (A549) y B (Huh7).

- La curva maestra es la media de la población de los rastros de células individuales corregidos en el tiempo de inicio, es decir, todos los tiempos de inicio fueron desplazados al punto cero. Las intensidades de fluorescencia se convirtieron en números reales de d2EGFP como se ha descrito anteriormente en la referencia (26). Las propiedades superiores de las construcciones de ARNm estabilizadas en 3' y 5' + 3' se ilustran en la gráfica de la curva maestra. Estas construcciones mostraron la disminución más superficial en la expresión de proteínas con el

tiempo y, por lo tanto, las semividas más largas además de valores de expresión de proteínas más elevados en comparación con el resto de construcciones.

X. Ejemplo 9: Área bajo la curva (AUC)

5 En farmacocinética, la exposición total de un fármaco se conoce como el "área bajo la curva". La expresión análoga en terapia génica es la integral de la cantidad de proteína expresada artificialmente a lo largo del tiempo, es decir, el área bajo la curva (expresión frente al tiempo) (AUC). La AUC es un medio de cuantificar simultáneamente la eficacia de la traducción y la estabilidad de una construcción de ARNm. Se puede interpretar como la dosis acumulativa en el tiempo de la proteína que está codificada en el ARNm y, por lo tanto, describe la eficacia de una construcción de ARNm seleccionada. Dado el modelo de velocidad bioquímica (véase **Figura 3A**), la AUC se puede calcular explícitamente:

$$AUC = 0,48 \cdot m_0 \cdot k_{TL} \cdot \tau_{ARNm} \cdot \tau_{d2EGFP} \quad \text{(Ecuación 3)}$$

15 Por lo tanto, una construcción de ARNm terapéutica óptima debería tener, deseablemente, tanto ARNm largo, τ_{ARNm} , como semivida de proteína, τ_{d2EGFP} y alta eficacia de traducción, k_{TL} . Además, la eficacia de transferencia que determina la cantidad inicial de ARNm terapéutico, m_0 , es directamente proporcional al AUC. En la **Figura 6^a**, se puede ver una explicación ilustrativa del curso temporal teórico de la expresión de proteína y la AUC calculado.

20 Si no hubiera degradación de proteína ($\beta = 0$), la cantidad de proteína dentro de una célula se encontraría en un nivel estable como consecuencia de un flujo equilibrado de traducción de ARNm y degradación de ARNm. En este

$$\frac{K}{\delta}(1 - e^{-\delta t})$$

caso, la dinámica de la expresión sigue $\frac{K}{\delta}(1 - e^{-\delta t})$. Lo mismo se cumpliría de manera análoga para el caso en el que δ fuera igual a cero. La superposición de esto con el decaimiento permanente y exponencial de la proteína $d2EGFP$ (después de $e^{-\beta t}$) da lugar a la forma característica de la AUC como se muestra en la **Figura 6A**. Las

25 **Figuras 6B y C** muestran las AUC medias relativas totales, así como las AUC relativas "por experimento" normalizadas con respecto a la AUC media del control, siendo esta última la AUC de expresión de proteína después de la transfección con la construcción de control. En ambos tipos de células, se encontraron las AUC relativas más altas para la construcción estabilizada en UTR 3' y UTR 5'+ 3'. Esto coincide con las semividas largas observadas para estas construcciones, porque contribuyen a la AUC como se ve en la **Ecuación 3**. En la **Figura 11**, se pueden encontrar las distribuciones detalladas de AUC de células individuales.

Más concretamente, asumiendo las ecuaciones de velocidad bioquímica (4) y (5) para la traducción y degradación de acuerdo con la **Figura 3C**, la cantidad de proteína expresada después de la transfección del ARNm viene dada por:

$$G_{d2EGFP}(t) = \frac{K}{\delta - \beta} \cdot (1 - e^{-(\delta - \beta)(t - t_0)}) \cdot e^{-\beta(t - t_0)} \quad \text{(Ecuación 1)}.$$

El área bajo la curva (AUC) se calcula integrando el nivel de expresión $G_{d2EGFP}(t)$ desde t_0 , cuando la expresión se establece en tiempos largos ($t \rightarrow \infty$):

$$AUC = \int_{t=t_0}^{t=\infty} G(t) = \frac{K}{\delta - \beta} \int_0^{\infty} [e^{-\beta\tau} - e^{-\delta\tau}] d\tau = \frac{K}{\delta - \beta} \cdot \left[\frac{1}{\beta} - \frac{1}{\delta} \right] = \frac{K}{\delta \cdot \beta}$$

con $\tau = t - t_0$.

$$\tau_{ARNm} = \ln 2 / \delta, \tau_{d2EGFP} = \ln 2 / \beta \quad \text{y} \quad K = m_0 \cdot k_{TL}$$

Usando Ecuación 3: , se obtiene la

$$AUC = 0,48 \cdot m_0 \cdot k_{TL} \cdot \tau_{ARNm} \cdot \tau_{d2EGFP}$$

El transcurso del tiempo de $G_{d2EGFP}(t)$ y la AUC se representa esquemáticamente en la **Figura 6A**. Las distribuciones experimentales de AUC de células individuales se pueden ver en la **Figura 11**. Debido a que la AUC depende linealmente de los tiempos de vida del ARNm y de la proteína, las distribuciones de AUC unicelulares

están estrechamente relacionadas con las distribuciones de las semividas del ARNm y de la proteína que se muestran en la **Figura 4B y 4C** del texto principal.

XI. Ejemplo 10: Factor de prolongación del tiempo de vida

Los factores de prolongación del tiempo de vida para las células A549 y Huh7 se muestran en la **Figura 6D y E**, respectivamente. Como era de esperar, todas las construcciones estabilizadas producen factores de prolongación del tiempo de vida superiores a uno, lo que significa que la inserción de UTR en cualquier extremo provoca la estabilización del ARNm. Sin embargo, la construcción de ARNm de UTR 3' muestra tiempos de vida del ARNm más largos que la construcción 2 x UTR 3'. De forma similar, la construcción UTR 5'+ 3' es más estable que la construcción 5' + 2x3'. Estos resultados son válidos para ambos tipos de células. Curiosamente, los efectos estabilizadores son significativamente más destacados en las células A549 que en las células Huh7 en todos los casos.

XI. Ejemplo 11: Comparación de construcciones que tienen UTR de diferentes genes en comparación con la construcción de UTR de CYBA

Se han comparado las construcciones n.º 2 a n.º 5 que tienen UTR de diferentes genes como se indica en la **Tabla 4** que se presenta a continuación con la construcción n.º 1 de UTR de CYBA para optimizar la estructura del ARNm en términos de estabilidad y productividad. Se seleccionaron cinco UTR celulares diferentes de un gen basándose en los datos de publicación (Hoen *et al.*, 2010) que presentan semividas del ARNm largas. Estas UTR celulares son CYBA, DECR1, GMFG, MAPBPIP y MYL6B. Las secuencias de las regiones no traducidas 5' y 3' de cada gen celular se obtuvieron de la base de datos de UTR (<http://utrdb.ba.itb.cnr.it/search>) y se clonaron en cinco combinaciones diferentes, que fueron UTR 5' sola, UTR 3' sola, UTR 5'+ 3', UTR 5' + 2 x UTR 3' y 2 x UTR 3'. En primer lugar, las secuencias de la región no traducida se clonaron en la cadena principal pVAX1-A120. En el caso de las UTR 5', la clonación se produjo a través del sitio de restricción de HindIII en el extremo 5' y el sitio de restricción de BamHI en el extremo 3', y se insertó cadena arriba del gen informador que codifica la luciferasa de *Metridia* (MetLuc). Los sitios de restricción para las UTR 3' fueron EcoRI (extremo 5') y PstI (extremo 3'), y se clonaron cadena abajo de MetLuc. Los plásmidos que contenían UTR 5' sola y UTR 5' + 3' para cada UTR celular fueron producidos por Eurofins MWG Operon. Estos plásmidos se transformaron en bacterias de *E. coli* (DH10B) mediante electroporación. Las otras combinaciones, incluyendo UTR 3' sola, UTR 5'+ 2 x UTR 3' y 2 x UTR 3' se clonaron internamente. La clonación de los plásmidos con UTR 3' se realizó simplemente cortando la UTR 5' de la cadena principal mediante digestión con HindIII (roma) y BamHI (roma). Las construcciones que contenían UTR 5' + 2 x UTR 3' se clonaron insertando MetLuc que contenía UTR 3' (BamHI/PstI romas) en la cadena principal de pVAX1-A120 MetLuc que comprendía UTR 5' + 3', reemplazando así MetLuc e insertando una segunda UTR 3' en frente de la respectiva UTR 3' de la cadena principal. Finalmente, las construcciones que contenían 2 x UTR 3' se generaron mediante la eliminación de la UTR 5' (HindIII y BamHI, ambas romas) del plásmido que contenía UTR 5'+ 2 x UTR 3'. Después de la clonación, todos los plásmidos se amplificaron en bacterias de *E. coli* (DH10B) tras la electroporación. En segundo lugar, el ARNm modificado químicamente se produjo por transcripción *in vitro*. Con este fin, se linealizaron los plásmidos mediante digestión con XbaI y se purificaron mediante precipitación con cloroformo/etanol. El kit de transcripción *in vitro* (Promega) incluía la mezcla de enzima polimerasa T7 requerida, así como los tampones adecuados. La mezcla de transcripción también contenía los nucleótidos no modificados adenosina-trifosfato (ATP), guanosina-trifosfato (GTP), uridina-trifosfato (UTP) y citosina-trifosfato (CTP), así como los nucleótidos químicamente modificados metil-CTP y tio-UTP (Jena Bioscience, GmbH, Jena, Alemania) con una concentración final de ATP: GTP: UTP: CTP: metil-CTP:tio-UTP de 7,13 mM: 1,14 mM: 5,36 mM: 5,36 mM: 0,536 mM: 0,536 mM. Además, se añadió el análogo de estructura de capuchón ARCA (análogo de capuchón anti-inverso) a la mezcla para asegurar la incorporación del capuchón de 5' en el sentido correcto. Finalmente, el ADN linealizado se añadió a la mezcla de reacción. La mezcla de IVT se incubó a 37 °C durante 2 h. Se permitió la digestión del ADN restante mediante la adición de ADNasa I y una incubación adicional a 37 °C durante otros 20 minutos. La precipitación del ARN se realizó mediante la adición de acetato de amonio previamente enfriado hasta una concentración final de 2,5 M. El sedimento de ARN se lavó con etanol al 70 %. La etapa de lavado se realizó dos veces. Al final, se volvió a suspender el ARN en agua exenta de RNasa. La concentración de ARN se determinó con un dispositivo espectrofotométrico y la pureza se ensayó en un gel de agarosa. Después de IVT, se ensayaron los diferentes ARNm en dos estirpes celulares diferentes, es decir, en NIH3T3 y A549. Para los experimentos de selección, se usó un sistema de administración de ácido nucleico no vírico, como lipofección. En un primer experimento de transfección, se ensayaron diferentes agentes de transfección para comparar la expresión de la proteína y la viabilidad celular (datos no mostrados). A continuación, se llevaron a cabo los experimentos de selección, incluida la valoración de la dosis, para evaluar los efectos dependientes de la dosis. La configuración experimental es la siguiente: se sembraron 5.000 células (NIH3T3) en 150 µl de medio completo de DMEM por pocillo en placas de 96 pocillos y se transfectaron 24 horas después de la siembra. Las células se transfectaron a una dosis inicial de 500 ng/pocillo (100 ng de ARNm /célula) usando el reactivo de transfección comercial Dreamfect Gold (DFG). Los complejos se prepararon a una proporción de 4 µl de Dreamfect Gold por 1 µg de ARNm. Para la formación de lipoplexos, se diluyó el ARNm (3,6 µg) por separado en DMEM sin complementos en un volumen total de 340 µl para cada ARNm. En una placa de 96 pocillos, se mezclaron 14,4 µl de DFG con 5,6 µl de agua en un pocillo preparado para cada dilución de ARNm. La formación del complejo tuvo lugar cuando se añadió la dilución de ARNm a la DFG y se mezcló mediante pipeteo ascendente y descendente. Las mezclas se incubaron a temperatura

ambiente durante 20 minutos. Mientras tanto, se prepararon las series de dilución. En los siete pocillos restantes subyacentes a la mezcla compleja, se añadieron 180 μ l de DMEM sin suplementos por pocillo. Tras el tiempo de incubación, se eliminaron 180 μ l de la solución compleja y se añadieron al primer pocillo de la serie de dilución. Este procedimiento se realizó hasta la última etapa de dilución. Finalmente, se añadieron 50 μ l de la solución compleja a las células y se incubaron durante 4 horas. Para cada construcción de ARNm, se prepararon triplicados biológicos. Después de 4 horas, se retiró el sobrenadante completo de la placa de cultivo celular para la medición y se añadieron 150 μ l de medio nuevo a cada pocillo. La bioluminiscencia se midió después de 4, 24, 48, 72, 96, 120 y 144 horas usando un lector de placas multilabel. A esto, se mezclaron 50 μ l de sobrenadante con 20 μ l de coelenterazina, y se midió la luz generada. Finalmente, se observó la cantidad de proteína a lo largo del tiempo, y se representa como el área bajo la curva (AUC). Los resultados se muestran en la **Figura 12**.

n.º	Nombre	UTR 5'	UTR 3'
1	CYBA	-	Elemento de estabilidad de UTR 3' de la insulina (INS_SCE), señal de poliadenilación (PAS)
2	DECR1 (2,4-dienoil CoA reductasa 1, mitocondrial)	SNP	PAS, PhastConsElements17way (bloque conservado)
3	GMFG (factor de maduración glial, gamma)	Marco de lectura abierta cadena arriba (uORF), PhastConsElements17way (2x)	PAS
4	MAPBPIP (adaptador endosomal/lisosomal, activador 2 MAPK y MTOR)	-	PAS
5	MYL6B (cadena ligera de la miosina 6B)	-	PAS, PhastConsElements17way

Tabla 4: Sumario de las construcciones n.º 1 a n.º 5.

15 XIII. Discusión

La determinación de la estabilidad del ARNm y su expresión son dos factores principales que se han de considerar cuando se trata de desarrollar nuevas terapias de ARNm. En el presente documento, se usaron diferentes combinaciones de UTR, una UTR 5', UTR 3', una UTR 5' + UTR 3', UTR 5' + 2 x UTR 3' y dos copias DE UTR 3' para mejorar el ARNm en términos de estabilidad y de su expresión. La AUC del curso temporal de d2EGFP también se evalúa, porque la expresión de proteína total es relevante para un efecto terapéutico sostenido. Con el fin de obtener datos detallados resueltos en el tiempo y controlar la dinámica de expresión de proteína a nivel unicelular, se usaron matrices microcelulares unicelulares para mediciones cuantitativas paralelas de la estabilidad del ARNm y la eficacia de traducción. La disposición regular de las células garantizó microambientes reproducibles y permitió un análisis de imágenes rápido y automatizado, que son requisitos previos para estudios comparativos de alto rendimiento y de una sola célula. El enfoque permite la determinación de las funciones de distribución para (i) la semivida proteica, (ii) las velocidades de expresión y (iii) la semivida del ARNm funcional. Tanto en las células Huh7 como en las células d2EGFP, las semividas proteicas medias de d2EGFP se distribuyeron estrechamente e independientemente de la secuencia de UTR. Los valores de semivida calculados de 4,5 horas para las células A549 y de 7,4 horas para las células Huh7 podrían atribuirse a las diferencias específicas del tipo de célula entre las estirpes celulares comparadas. Dichas diferencias específicas de las células en la semivida de d2EGFP se han publicado previamente. Un estudio en células NIH3T3 usando un enfoque de citometría de imagen similar, registró una semivida de 2,8 h dentro de una ventana de medición de 12 horas (33). Se ha informado de una semivida incluso más corta, de menos de dos horas, para las células CHO por Li *et al.* (34). En el presente documento, la degradación de la proteína se midió mediante transferencia Western y citometría de flujo solo durante tres horas. Para validar los presentes hallazgos del análisis de datos de una sola célula, se determinaron además los tiempos de vida de d2EGFP en mediciones directas usando cicloheximida (véase la **Figura 10**). Se encontraron tiempos de vida más cortos en comparación con los valores observados del análisis de datos de una sola célula. Esto podría deberse al hecho de que, en el análisis de datos de una sola célula, se supuso un número inicial constante de moléculas de ARNm como parte de la velocidad de expresión combinada $K = m_0 \cdot k_{TL}$ (véase la **Ecuación 1**). Sin embargo, independientemente del hecho de que las células se hubieran lavado después de un tiempo de incubación de una hora, todavía era probable que el número de moléculas de ARNm no fuera constante desde el comienzo de la observación. Por consiguiente, las moléculas de ARNm que están disponibles para su posterior traducción, que conducen a la expresión de proteínas, podrían dar lugar a valores de semivida más largos obtenidos a partir del

ajuste del tiempo de expresión de una sola célula. Cuando se compara la semivida determinada para las células A549 con la semivida determinada para las células Huh7, se encuentra la misma relación de aproximadamente 1,64 para ambos métodos de medición. Además, incluso una posible sobreestimación sistemática de las semividas de ARNm y proteína no cambia el orden cualitativo del rendimiento del ARNm. La velocidad de expresión depende del número inicial de moléculas de ARNm, m_0 , así como de la velocidad de traducción K_{TL} . Cabe señalar que el número de moléculas de ARNm entregadas con éxito varía debido a la estocasticidad intrínseca del proceso de administración. Sin embargo, se espera que el número medio de moléculas de ARNm sea el mismo, ya que el protocolo de transfección se ha mantenido escrupulosamente en todos los experimentos. Por el contrario, la actividad de traducción (K_{TL}) de las diversas construcciones de UTR podría variar. Aún así, el hecho de que las distribuciones, así como los valores medios de la velocidad de expresión K sean bastante similares para todas las construcciones (véanse las **Figuras 3A y D**) indica que la velocidad de traducción está simplemente influenciada por las UTR insertadas. El parámetro de mayor interés es la semivida del ARNm. En el presente documento, se comparó la semivida del ARNm funcional con la semivida física del ARNm. Los resultados con estudios de transfección de células individuales sugieren que cualquier inserción de UTR 5' o 3' en la secuencia de ARNm aumenta su semivida del ARNm funcional. Todas las modificaciones ensayadas en este estudio condujeron a semividas de ARNm prolongadas (véanse las **Figuras 2 y 3**), lo que dio lugar a una expresión prolongada medida mediante microscopía de fluorescencia y FC (véase la **Figura 1**). En contraste con la semivida del ARNm funcional, la semivida física del ARNm determinada mediante qRT-PCR mostró una disminución en la estabilidad del ARNm para UTR 5', 3', 5'+ 2 x 3' y 2 x 3' en ambas estirpes celulares (véase **Figura 8A y B**). Una diferencia importante es la capacidad de traducción para cada ARNm medido en ambos métodos. En el caso de medir semividas funcionales de ARNm, el ARNm participa en la traducción activa, mientras que la semivida de ARNm física se controla independientemente del estado de la traducción del ARNm detectado. Se ha informado de hallazgos similares por parte de Gallie *et al.* (35). Se cree que la semivida física del ARNm no es un indicador apropiado de la capacidad de traducción del ARNm. La capacidad de traducción para un ARNm podría evaluarse a partir de su semivida funcional (longevidad de la expresión) y la cantidad de proteína total producida (área bajo la curva). Para un ARNm terapéutico, es imperativo que la molécula sea funcional durante el mayor tiempo posible y produzca la máxima proteína posible. Esto lleva a la conclusión de que tanto la semivida funcional del ARNm como las cantidades totales de proteína producidas son las mejores medidas para identificar, comparar y ensayar la administración del ARNm. Además, la distribución heterogénea de las semividas señala la importancia de las técnicas de medición de las células individuales, porque estos efectos se oscurecen en las mediciones de conjunto (véase **Figuras 2, 4 y 8A y B**). Curiosamente, se observó un efecto positivo sobre la expresión de proteína para UTR 5' sola, aunque hasta ahora, no se ha descubierto ningún motivo conocido en la UTR 5' de CYBA. Por primera vez, se ha demostrado que las UTR de CYBA en cualquier extremo son suficientes para aumentar tanto el máximo como la persistencia de la expresión de la proteína en ambas estirpes celulares. Estos hallazgos coinciden con las publicaciones que afirman el comportamiento individual o sinérgico de las UTR 5' y UTR 3' (14). En contraste con Holtkamp *et al.* (16), no se pudo observar un aumento adicional en la expresión de la proteína o la estabilidad del ARNm con dos copias secuenciales de la UTR 3' en comparación con una sola UTR 3' (véase la **Figura 4**). Por el contrario, incluso se produjeron tiempos de vida más cortos tanto para la inserción UTR 5'+ 3' frente a 5'+ 2 x 3' como para la inserción de UTR 3' frente a 2 x 3'. Esto podría deberse al hecho de que se usó un tipo diferente de células (en concreto, las células dendríticas) en el estudio de Holtkamp *et al.* (16). También se han informado efectos específicos similares de tipo celular para los hepatocitos (39). Otro factor contribuyente que afecta tanto a la estabilidad del ARNm como a su eficacia de traducción podría ser la estructura secundaria de los diferentes ARNm. Dichos efectos de la estructura secundaria de ARNm en la regulación de la expresión génica se han descrito anteriormente (36,37). Las características estructurales importantes junto con su energía libre mínima para las construcciones de ARNm usadas en el presente estudio se resumen en la **Tabla 3**.

La expresión de proteína persistente de la construcción estabilizada con UTR 5'+ 3' podría deberse a la unión del extremo 5' al 3', lo que facilita la circularización del ARNm (19). Debido a que no se pudieron encontrar estructuras secundarias estables dentro de la UTR 5', se supone que esta característica permite un inicio temprano de la expresión (38). Por el contrario, se identificaron estructuras secundarias dentro de las UTR 3'. Estas podrían proteger al ARNm de la ruta de degradación 3'-5'. Dos UTR de 3' mostraron aún más estructuras secundarias (dos horquillas) con la mejor energía libre mínima, lo que indica una expresión más persistente. Tomados en conjunto, estos hallazgos podrían ser la explicación de la expresión de inicio inferior de la 2 x UTR 3' en comparación con la UTR 5' y la expresión persistente en los puntos de tiempo posteriores de las construcciones de ARNm que contienen UTR 3'. De acuerdo con las semividas de proteínas, se obtuvieron valores de semivida más largos para los ARNm estabilizados con UTR. Esto se observó en ambas estirpes celulares con diferencias específicas de células que muy probablemente afectan a los valores absolutos. En las células A549, las semividas de ARNm para las construcciones con UTR variaron de 13,0 a 23,0 h en comparación con 5,8 h para el control. En las células Huh7, se midieron las semividas de 9,9 a 13,6 h para las construcciones que contenían UTR, en oposición a una semivida de 7,8 h para el ARNm de control. La semivida del ARNm estabilizado con UTR 3' en las células A549 concuerda bien con los tiempos de vida del ARNm de los ARNm estabilizados de manera similar que se informaron anteriormente (16,26). El hecho de que la estabilidad y la cinética de degradación del ARNm y proteína difieren en diferentes tipos de células es muy probable debido a las diferencias en las complejas redes de interacciones entre el ARNm y las proteínas que, muy probablemente, dependen del tipo de célula. Tomados en conjunto, los presentes resultados tanto en células A549 como en células Huh7, independientemente del método de análisis (FC o análisis de células individuales), sugieren que se pueden inducir altos niveles de expresión proteica mediante el ARNm estabilizado con

UTR de CYBA. La elección de la combinación de UTR depende de la necesidad del experimento de aplicación. Cuando se desea una expresión de proteína persistente con disminución de la degradación del ARNm, el ARNm estabilizado con una UTR 3' sola podría servir para el fin. Sin embargo, la combinación de UTR 5'+ 3' da lugar a características deseables adicionales de inicio temprano, alto máximo y expresión de proteína acumulada. En el presente documento, se demuestra que el análisis de células individuales de la expresión de proteína inducida por ARNm es un medio para caracterizar y mejorar las propiedades farmacocinéticas de las construcciones de ARNm. Usando este enfoque, es posible evaluar sistemáticamente la biodisponibilidad intracelular de diferentes construcciones de ARNm para identificar secuencias que produzcan una expresión de proteína sostenida. Se encontró una persistencia prolongada de la expresión de proteínas para construcciones estabilizadas mediante inserciones de UTR usando un modelo unicelular y análisis FC en dos tipos de células. Este hallazgo es deseable en el caso de desarrollar terapias de ARNm. Las construcciones de ARN mensajero con expresión de proteína persistente durante un período de tiempo (AUC) son deseables y permiten la dosis reducida apropiada a un paciente con un resultado terapéutico final.

15 Referencias

1. G. Tavernier, O. Andries, J. Demeester, N. N. Sanders, S. C. De Smedt y J. Rejman, *J Control Release*, 2011, **150**, 238-247.
2. A. Yamamoto, M. Kormann, J. Rosenecker y C. Rudolph, *Eur J Pharm Biopharm*, 2009, **71**, 484-489.
3. M. S. D. Kormann, G. Hasenpusch, M. K. Aneja, G. Nica, A. W. Flemmer, S. Herber-Jonat, M. Huppmann, L. E. Mays, M. Illenyi, A. Schams, M. Griese, I. Bittmann, R. Handgretinger, D. Hartl, J. Rosenecker y C. Rudolph, *Nat Biotech*, 2011, **29**, 154-157.
4. M. Esteller, *Nat Rev Genet*, 2011, **12**, 861-874.
5. K. Kariko y D. Weissman, "Current Opinion in Drug Discovery and Development", 2007, **10**, 523
6. G. Pesole, G. Grillo, A. Larizza y S. Liuni, *Briefings in Bioinformatics*, 2000, **1**, 236-249.
7. T. V. Pestova, J. R. Lorsch y C. U. Hellen, "Cold Spring Harbor Monograph Archive", 2007, **48**, 87-128.
8. L. Barrett, S. Fletcher y S. Wilton, *Cell. Mol. Life Sci.*, 2012, **69**, 3613-3634.
9. F. Mignone, G. Grillo, F. Licciulli, M. Iacono, S. Liuni, P. J. Kersey, J. Duarte, C. Saccone y G. Pesole, *Nucleic Acids Research*, 2005, **33**, D141-D146.
10. M. J. Moore, *Science*, 2005, **309**, 1514-1518.
11. X. Pichon, L. A. Wilson, M. Stoneley, A. Bastide, H. A. King, J. Somers y A. E. Willis, *Current Protein & Peptide Science*, 2012, **13**, 294-304.
12. P. A. C. 't Hoen, M. Hirsch, E. J. d. Meijer, R. X. d. Menezes, G. J. van Ommen y J. T. d. Dunnen, *Nucleic Acids Research*, 2011, **39**, 556-566.
13. G. Pesole, F. Mignone, C. Gissi, G. Grillo, F. Licciulli y S. Liuni, *Gene*, 2001, **276**, 73-81.
14. F. Gebauer y M. W. Hentze, *Nat Rev Mol Cell Biol*, 2004, **5**, 827-835.
15. L. Tillmar, C. Carlsson y N. Welsh, *Journal of Biological Chemistry*, 2002, **277**, 1099-1106.
16. S. Holtkamp, S. Kreiter, A. Selmi, P. Simon, M. Koslowski, C. Huber, O. Tureci y U. Sahin, *Blood*, 2006, **108**, 4009-4017.
17. C. H. d. Moor, H. Meijer y S. Lissenden, *Seminars in Cell & Developmental Biology*, 2005, **16**, 49-58.
18. N. L. Garneau, J. Wilusz y C. J. Wilusz, *Nat Rev Mol Cell Biol*, 2007, **8**, 113-126.
19. E. Szostak y F. Gebauer, *Briefings in Functional Genomics*, 2012.
20. S. Tyagi, *Nat Meth*, 2009, **6**, 331-338.
21. H. Y. Park, H. Lim, Y. J. Yoon, A. Follenzi, C. Nwokafor, M. Lopez-Jones, X. Meng y R. H. Singer, *Science*, 2014, **343**, 422-424.
22. C. Miller, B. Schwalb, K. Maier, D. Schulz, S. Dümcke, B. Zacher, A. Mayer, J. Sydow, L. Marcinowski, L. Dölken, D. E. Martin, A. Tresch y P. Cramer, *Molecular Systems Biology*, 2011, **7**.
23. T. Nolan, R. E. Hands y S. A. Bustin, *Nat. Protocols*, 2006, **1**, 1559-1582.
24. M. Rabani, J. Z. Levin, L. Fan, X. Adiconis, R. Raychowdhury, M. Garber, A. Gnirke, C. Nusbaum, N. Hacohen, N. Friedman, I. Amit y A. Regev, *Nat Biotech*, 2011, **29**, 436-442.
25. B. Schwanhausser, D. Busse, N. Li, G. Dittmar, J. Schuchhardt, J. Wolf, W. Chen y M. Selbach, *Nature*, 2011, **473**, 337-342.
26. C. Leonhardt, G. Schwake, T. R. Stögbauer, S. Rappl, J.-T. Kuhr, T. S. Ligon y J. O. Rädler, *Nanomedicine: Nanotechnology, Biology and Medicine*, 2014, **10**, 679-688.
27. P. J. F. Rottgermann, A. P. Alberola y J. O. Radler, *Soft Matter*, 2014.
28. P. Corish y C. Tyler-Smith, *Protein Engineering*, 1999, **12**, 1035-1040.
29. J. J. Rossi, *Nat Cell Biol*, 2005, **7**, 643-644.
30. U. Sheth y R. Parker, *Cell*, 2006, **125**, 1095-1109.
31. A. Jakymiw, K. M. Pauley, S. Li, K. Ikeda, S. Lian, T. Eystatoy, M. Satoh, M. J. Fritzler y E. K. L. Chan, *Journal of Cell Science*, 2007, **120**, 1317-1323.
32. R. Parker y U. Sheth, *Molecular Cell*, 2007, **25**, 635-646.
33. M. Halter, A. Tona, K. Bhadriraju, A. L. Plant y J. T. Elliott, "Cytometry Part A", 2007, **71A**, 827-834.
34. X. Li, X. Zhao, Y. Fang, X. Jiang, T. Duong, C. Fan, C.-C. Huang y S. R. Kain, *Journal of Biological Chemistry*, 1998, **273**, 34970-34975.
35. D. R. Gallie, *Genes & Development*, 1991, **5**, 2108-2116.
36. J.-M. Chen, C. Férec y D. Cooper, *Hum Genet*, 2006, **120**, 301-333.

37. P. Gaspar, G. Moura, M. A. Santos y J. L. Oliveira, *Nucleic acids research*, 2013, **41**, e73-e73.
 38. J. R. Babendure, J. L. Babendure, J.-H. Ding y R. Y. Tsien, *RNA* (Nueva York, N.Y, 2006, **12**, 851-861.
 39. B. T. Kren y C. J. Steer, *FASEB J.*, 1996, **10**, 559-573.
 5 40. Zuker, M. (2003) "Mfold web server for nucleic acid folding and hybridization prediction". *Nucleic acids research*, **31**, 3406-3415.
 41. Babendure, J. R., Babendure, J. L., Ding, J.-H. y Tsien, R. Y. (2006) "Control of mammalian translation by mRNA structure near caps". *RNA* (Nueva York, N.Y, **12**, 851-861.
 42. Siegel, M. R. y Sisler, H. D. (1963) "Inhibition of Protein Synthesis in vitro by Cycloheximide". *Nature*, **200**, 675-676.
 10 43. Friedel CC, Dölken L, Ruzsics Z, Koszinowski UH, Zimmer R. (2009) "Conserved principles of mammalian transcriptional regulation revealed by RNA half-life". *Nucleic acids research*, **37**, e115.

LISTADO DE SECUENCIAS

- 15 <110> ethris GmbH
 <120> UTR que aumentan la eficacia de la traducción de las moléculas de ARN
 <130> X2090 EP S3
 20 <150> EP 15 17 4683.1
 <151> 30-06-2015
 <160> 11
 25 <170> BiSSAP 1.2
 <210> 1
 <211> 36
 30 <212> ARN
 <213> Secuencia artificial
 <220>
 <221> fuente
 35 <222> 1..36
 <223> /tipo_mol = "ARN sin asignar"
 /nota = "UTR 5\222 del gen del polipéptido alfa del citocromo b-245 (CYBA) humano"
 /organismo = "Secuencia artificial"
 40 <400> 1
cgcgccuagc agugucccag ccggguucgu gucgcc **36**
 <210> 2
 45 <211> 64
 <212> ARN
 <213> Secuencia artificial
 <220>
 50 <221> fuente
 <222>1..64
 <223> /tipo_mol = "ARN sin asignar"
 /nota = "UTR 3\222 del gen del polipéptido alfa del citocromo b-245 (CYBA) humano"
 /organismo ="Secuencia artificial"
 55 <400> 2
ccucgccccg gaccugcccu cccgccaggu gcaccaccu gcauuaaag cagcgaagcc **60**
ggga **64**
 60 <210> 3
 <211> 20
 <212> ADN
 <213> Secuencia artificial

ES 2 667 644 T3

<211> 778
 <212> ARN
 <213> *Homo sapiens*

5 <220>
 <221> fuente
 <222>1..778
 <223> /tipo_mol = "ARN sin asignar"
 /nota = "polipéptido alfa del citocromo b-245 (CYBA) humano"
 10 /organismo = "*Homo sapiens*"

<400> 7

ggcgggguuc	ggccgggagc	gcaggggcgg	cagugcgcgc	cuagcagugu	cccagccggg	60
uucgugucgc	cauggggcag	aucgaguggg	ccaugugggc	caacgagcag	gcgucggcgu	120
ccggccugau	ccucaucacc	gggggcaucg	uggccacagc	ugggcgcuuc	accaguggu	180
acuuuggugc	cuacuccauu	guggcgggcg	uguuugugug	ccugcuggag	uacccccggg	240
ggaagaggaa	gaagggcucc	accauggagc	gcuggggaca	gaaguacaug	accgccgugg	300
ugaagcuguu	cgggcccuuu	accaggaauu	acuauguucg	ggccguccug	caucuccugc	360
ucucggugcc	cgccggcuuc	cugcuggcca	ccauccuugg	gaccgccugc	cuggccaauug	420
cgagcggcau	cuaccuacug	gcggcugugc	guggcgagca	guggacgcc	aucgagccca	480
agccccggga	gcggccgcag	aucggaggca	ccaucaagca	gccgcccagc	aacccccgcg	540
cgcgcccccc	ggccgaggcc	cgcaagaagc	ccagcgagga	ggaggcugcg	guggcggcgg	600
ggggaccccc	gggagguccc	caggucaacc	ccaucccggu	gaccgacgag	gucgugugac	660
cucgcccccg	accugcccuc	ccgccaggug	caccaccug	caauaaaugc	agcgaagccg	720
ggaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaaaa	778

15 <210> 8
 <211> 588
 <212> ADN
 <213> *Homo sapiens*

20 <220>
 <221> fuente
 <222>1..588
 <223> /tipo_mol = "ADN sin asignar"
 25 /nota = "Secuencia de codificación del polipéptido alfa del citocromo b-245 humano"
 /organismo = "*Homo sapiens*"

<400> 8

ES 2 667 644 T3

atggggcaga tcgagtgggc catgtgggccc aacgagcagg cgctggcgtc cggcctgatc 60
ctcatcaccg ggggcatcgt ggccacagct gggcgcttca cccagtggta ctttggtgcc 120
tactccattg tggcgggctgt gtttgtgtgc ctgctggagt acccccgggg gaagaggaag 180
aagggctcca ccatggagcg ctggggacag aagtacatga ccgccgtggt gaagctgttc 240
gggcccttta ccaggaatta ctatgttcgg gccgtcctgc atctcctgct ctcggtgccc 300
gccggcttcc tgctggccac catccttggg accgcctgcc tggccattgc gagcggcatc 360
tacctactgg cggctgtgcg tggcgagcag tggacgcca tcgagcccaa gccccgggag 420
cggccgcaga tcggaggcac catcaagcag ccgcccagca acccccgcc ggggcccccg 480
gccgaggccc gcaagaagcc cagcagggag gaggctgagg tggcggcggg gggacccccg 540
ggaggtcccc aggtcaacc catcccgggtg accgacgagg tcgtgtga 588

5 <210> 9
<211> 195
<212> PRT
<213> *Homo sapiens*

10 <220>
<223> Secuencia de codificación del polipéptido alfa del citocromo b-245 humano

<400> 9

Met	Gly	Gln	Ile	Glu	Trp	Ala	Met	Trp	Ala	Asn	Glu	Gln	Ala	Leu	Ala
1				5					10					15	
Ser	Gly	Leu	Ile	Leu	Ile	Thr	Gly	Gly	Ile	Val	Ala	Thr	Ala	Gly	Arg
			20					25					30		
Phe	Thr	Gln	Trp	Tyr	Phe	Gly	Ala	Tyr	Ser	Ile	Val	Ala	Gly	Val	Phe
		35				40						45			
Val	Cys	Leu	Leu	Glu	Tyr	Pro	Arg	Gly	Lys	Arg	Lys	Lys	Gly	Ser	Thr
	50					55					60				
Met	Glu	Arg	Trp	Gly	Gln	Lys	Tyr	Met	Thr	Ala	Val	Val	Lys	Leu	Phe
65					70						75				80
Gly	Pro	Phe	Thr	Arg	Asn	Tyr	Tyr	Val	Arg	Ala	Val	Leu	His	Leu	Leu
				85					90					95	
Leu	Ser	Val	Pro	Ala	Gly	Phe	Leu	Leu	Ala	Thr	Ile	Leu	Gly	Thr	Ala
			100					105					110		
Cys	Leu	Ala	Ile	Ala	Ser	Gly	Ile	Tyr	Leu	Leu	Ala	Ala	Val	Arg	Gly
		115				120						125			
Glu	Gln	Trp	Thr	Pro	Ile	Glu	Pro	Lys	Pro	Arg	Glu	Arg	Pro	Gln	Ile
	130					135					140				
Gly	Gly	Thr	Ile	Lys	Gln	Pro	Pro	Ser	Asn	Pro	Pro	Pro	Arg	Pro	Pro
145					150					155					160
Ala	Glu	Ala	Arg	Lys	Lys	Pro	Ser	Glu	Glu	Glu	Ala	Ala	Val	Ala	Ala
				165					170					175	
Gly	Gly	Pro	Pro	Gly	Gly	Pro	Gln	Val	Asn	Pro	Ile	Pro	Val	Thr	Asp
			180					185					190		
Glu	Val	Val													
		195													

15 <210> 10
<211> 56
<212> ARN
20 <213> Secuencia artificial

<220>

ES 2 667 644 T3

<221> fuente
<222>1..56
<223> /tipo_mol = "ARN sin asignar"
/nota = "UTR 5' extendida del gen del polipéptido alfa del citocromo b-245 (CYBA) humano"
/organismo = "Secuencia artificial"

5

<400> 10

ggagcgcagg ggcggcagtg cgcgcctagc agtgtcccag ccgggttcgt gtcgcc 56

10

<210> 11
<211> 56
<212> ARN
<213> Secuencia artificial

15

<220>
<221> fuente
<222>1..56
<223> /tipo_mol = "ARN sin asignar"
/nota = "UTR 5' extendida del gen del polipéptido alfa del citocromo b-245 (CYBA) humano"
/organismo = "Secuencia artificial"

20

<400> 11

25 ggagcgcagg ggcggcagug cgcgccuagc agugucccag ccggguucgu gtcgcc 56

REIVINDICACIONES

1. Una molécula de ARN que comprende:

- 5 (a) una región de codificación que codifica un polipéptido; y
(b) cadena arriba de dicha región de codificación, una o más UTR que comprenden la secuencia mostrada en SEQ ID NO: 1 o una secuencia que muestra de 1 a 4 sustituciones en comparación con SEQ ID NO: 1 y que da lugar a una molécula de ARN que tiene la misma o mayor eficacia de traducción que una molécula de ARN que comprende una UTR que comprende la SEQ ID NO: 1; y/o
10 (c) cadena abajo de dicha región de codificación, una o más UTR que comprenden la secuencia mostrada en SEQ ID NO: 2 o una secuencia que muestra de 1 a 7 sustituciones en comparación con SEQ ID NO: 2 y que da lugar a una molécula de ARN que tiene la misma o mayor eficacia de traducción que una molécula de ARN que comprende una UTR que comprende la SEQ ID NO: 2,

15 en la que dicho polipéptido codificado por dicha región de codificación no es un polipéptido alfa del citocromo b-245 (CYBA).

2. La molécula de ARN de acuerdo con la reivindicación 1, en la que dicha/s UTR según lo definido en la reivindicación 1(b) está/n ubicada/s en el extremo 5' de la región de codificación según lo definido en la
20 reivindicación 1(a).

3. La molécula de ARN de acuerdo con la reivindicación 1, en la que dicha/s UTR según lo definido en la reivindicación 1(c) está/n ubicada/s en el extremo 3' de la región de codificación según lo definido en la
25 reivindicación 1(a).

4. La molécula de ARN de acuerdo con una cualquiera de las reivindicaciones 1 a 3, en la que dicha/s UTR según lo definido en la reivindicación 1(b) está/n ubicada/s en el extremo 5' de la región de codificación según lo definido en la reivindicación 1(a), y en la que dicha/s UTR según lo definido en la reivindicación 1(c) está/n ubicada/s en el
30 extremo 3' de la región de codificación según lo definido en la reivindicación 1(a).

5. La molécula de ARN de acuerdo con una cualquiera de las reivindicaciones 1 a 4, que comprende una UTR según lo definido en la reivindicación 1(b) en el extremo 5' de la región de codificación según lo definido en la reivindicación 1(a), y que comprende dos UTR según lo definido en la reivindicación 1(c) en el extremo 3' de la región de
35 codificación según lo definido en la reivindicación 1(a).

6. La molécula de ARN de acuerdo con la reivindicación 1 o 3, que comprende dos UTR según lo definido en la reivindicación 1(c) en el extremo 3' de la región de codificación según lo definido en la reivindicación 1(a).

7. La molécula de ARN de acuerdo con una cualquiera de las reivindicaciones 1 a 6, en la que la molécula de ARN
40 comprende una cola de poli-A en el extremo 3'.

8. La molécula de ARN de acuerdo con una cualquiera de las reivindicaciones 1 a 7, en la que la cola de poli-A tiene una longitud de al menos 120 nucleótidos.

45 9. Una molécula de ácido nucleico que codifica la molécula de ARN de una cualquiera de las reivindicaciones 1 a 8.

10. Un vector que comprende la molécula de ácido nucleico de la reivindicación 9.

50 11. Una célula hospedadora que comprende el vector de la reivindicación 10.

12. Una composición farmacéutica que comprende la molécula de ARN de acuerdo con una cualquiera de las reivindicaciones 1 a 8, la molécula de ácido nucleico de acuerdo con la reivindicación 9, el vector de acuerdo con la reivindicación 10 o la célula hospedadora de acuerdo con la reivindicación 11 y, opcionalmente, un vehículo farmacéuticamente aceptable.
55

13. La composición farmacéutica de la reivindicación 12 para su uso en terapias basadas en ARN.

14. Un kit que comprende la molécula de ARN de acuerdo con una cualquiera de las reivindicaciones 1 a 8, la molécula de ácido nucleico de acuerdo con la reivindicación 9, el vector de acuerdo con la reivindicación 10 o la
60 célula hospedadora de acuerdo con la reivindicación 11.

15. Uso *in vitro* de una o más UTR según lo definido en la reivindicación 1(b) y/o una o más UTR según lo definido en la reivindicación 1(c) para aumentar la eficacia de traducción de una región de codificación de una molécula de ARN a un polipéptido o una proteína codificados por dicha región de codificación.
65

Nueva solicitud de patente EP
 Ethris GmbH
 Nuestra referencia: X2090 EP S3

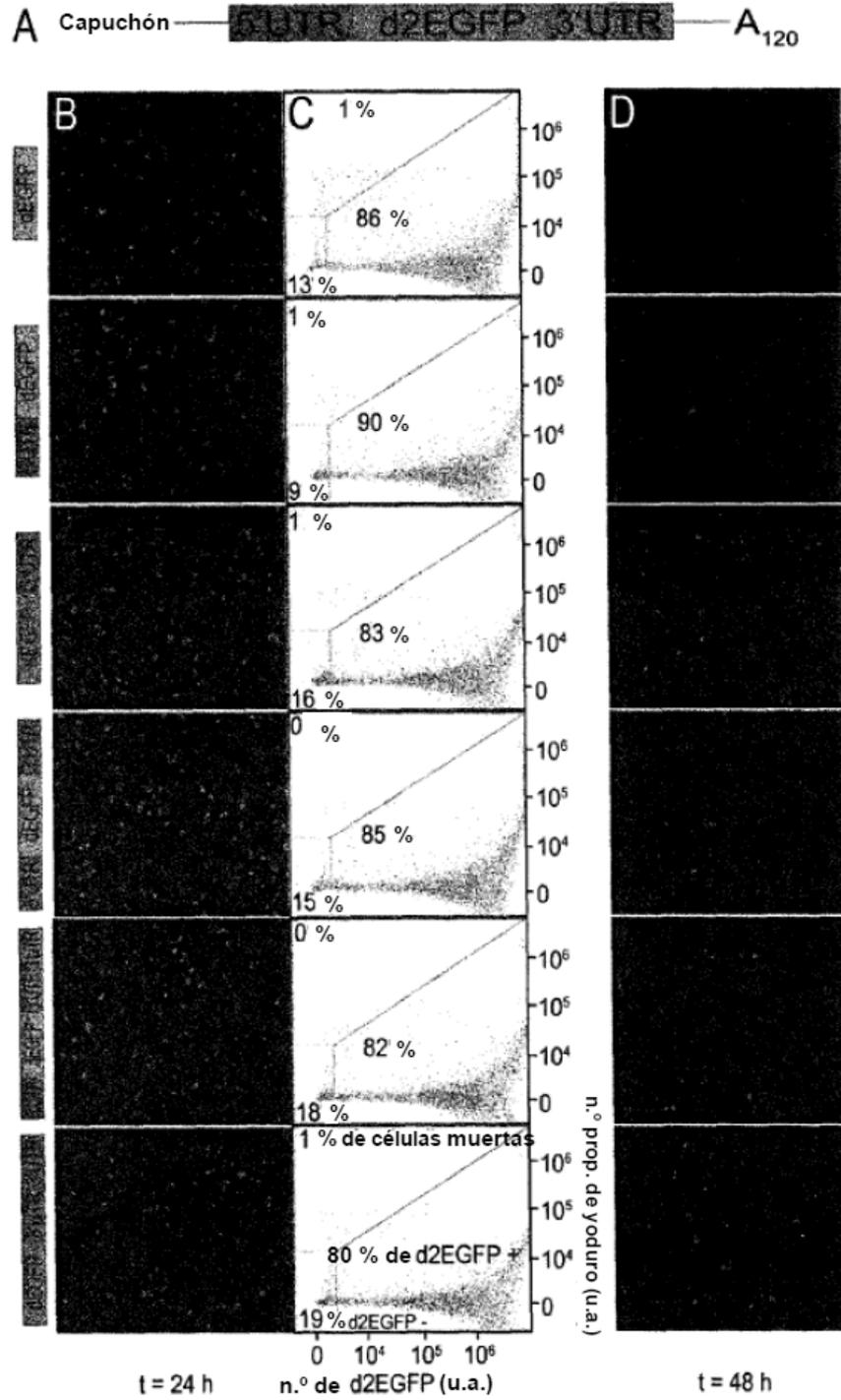


Figura 1

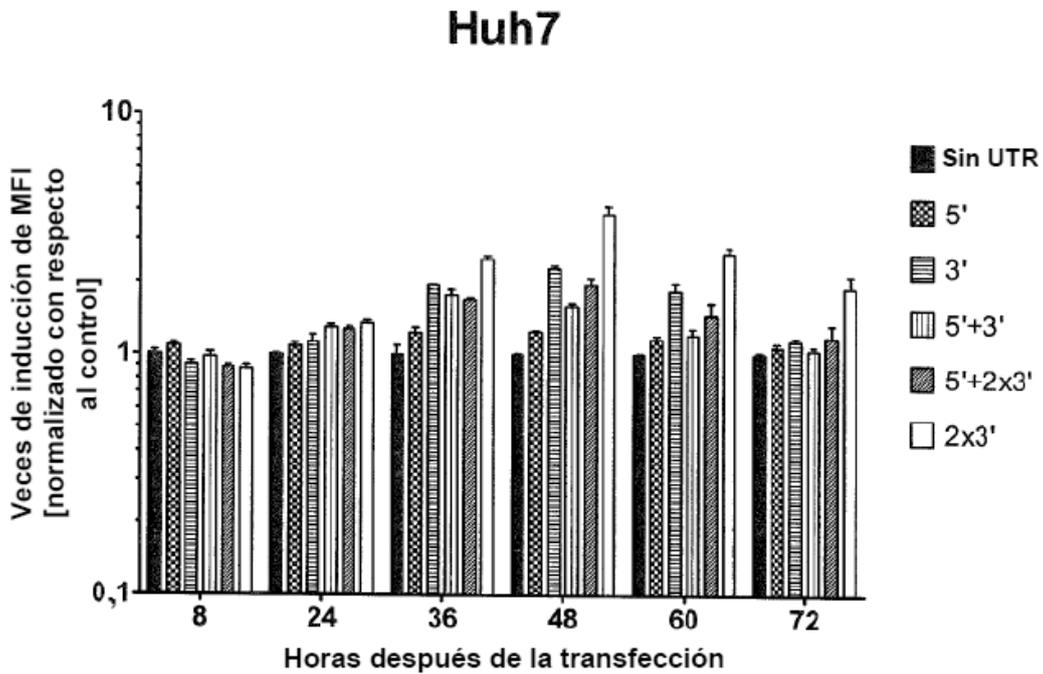
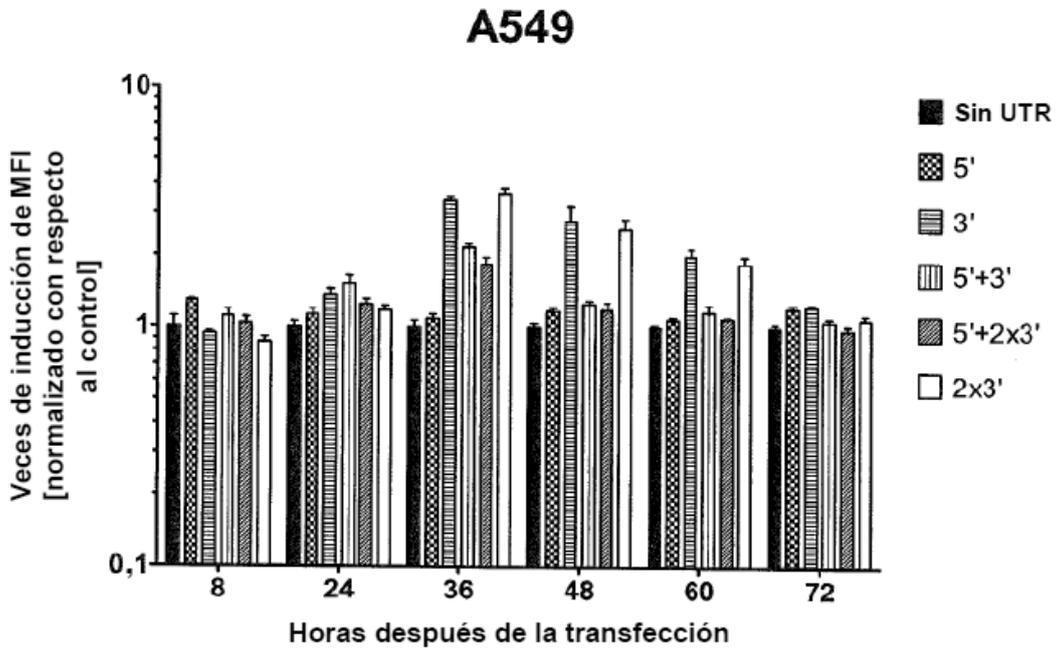


Figura 2: MFI

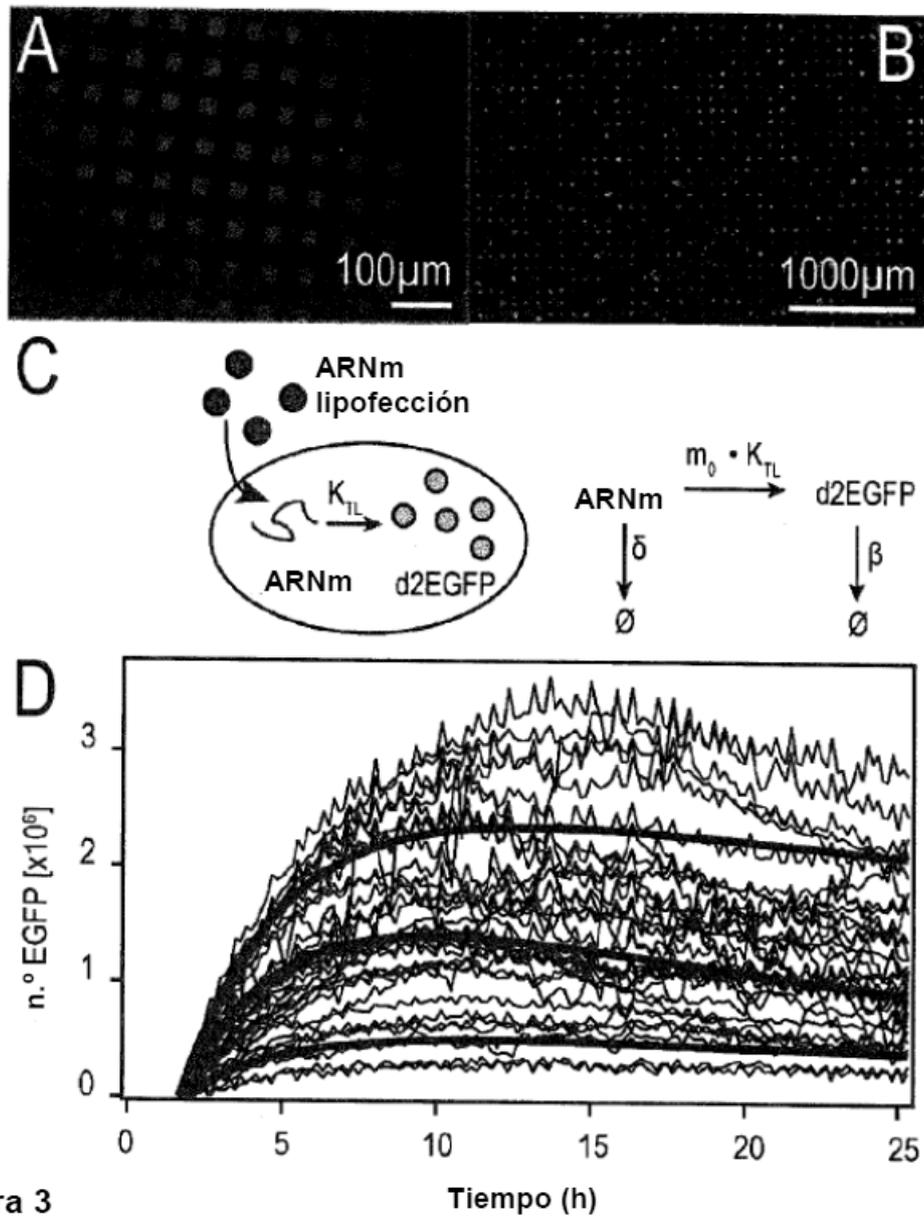


Figura 3

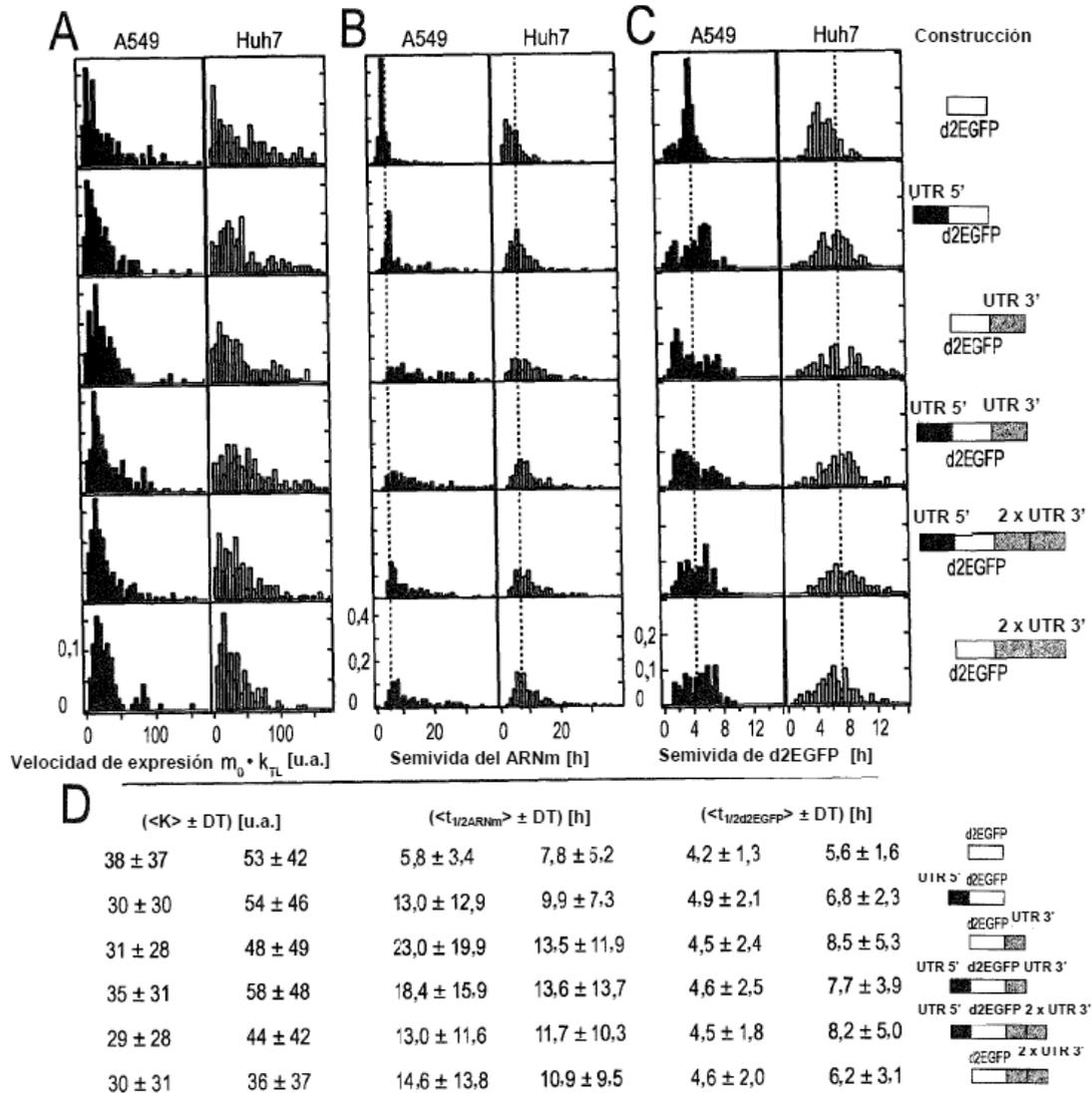


Figura 4

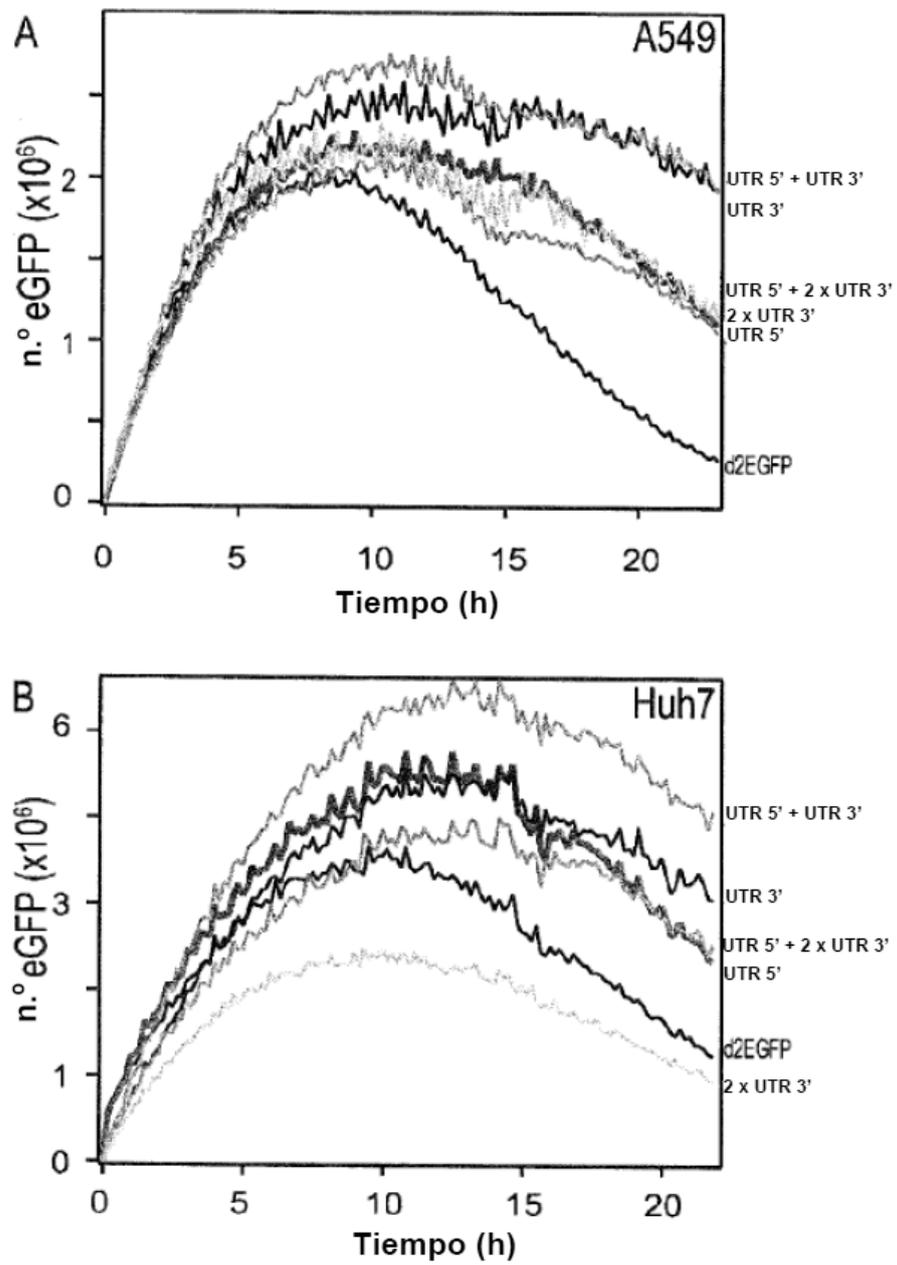


Figura 5

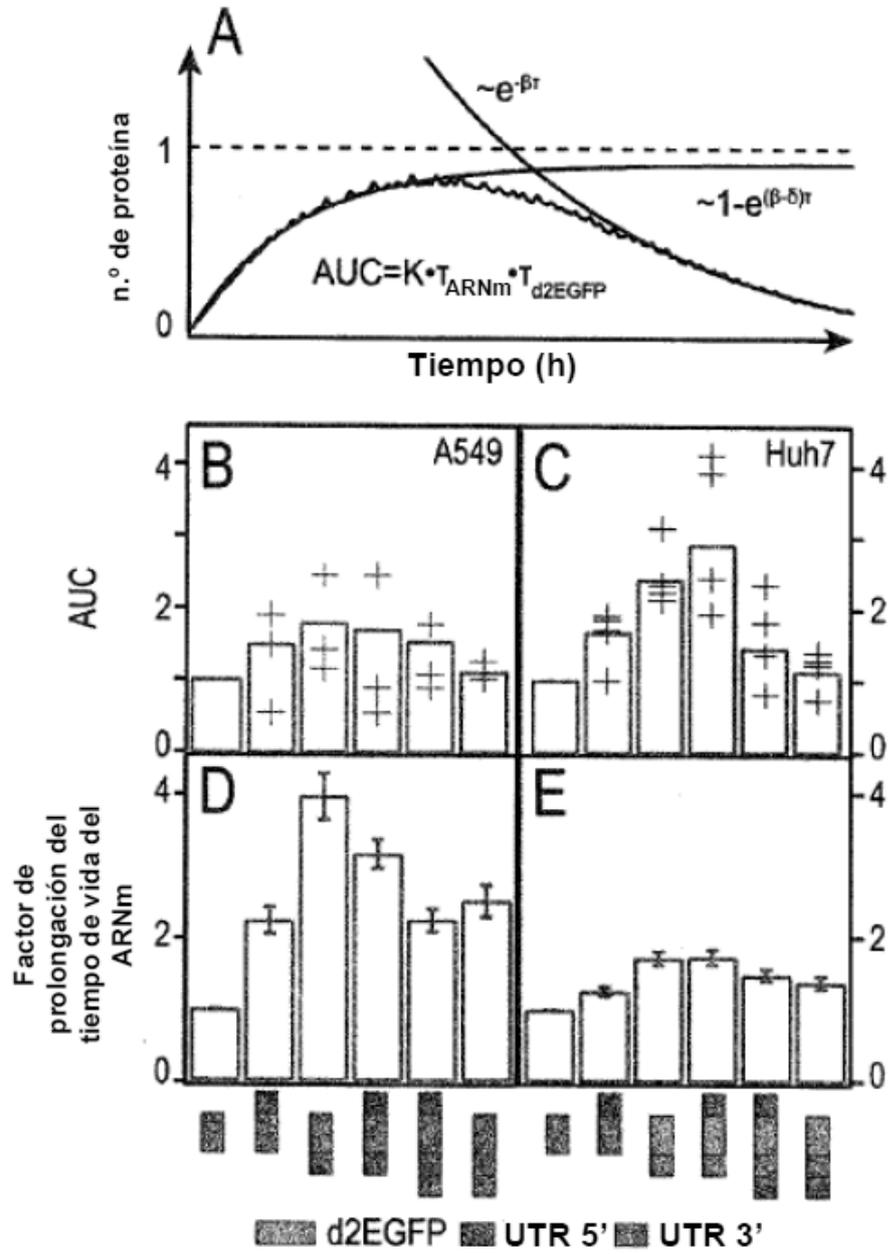


Figura 6

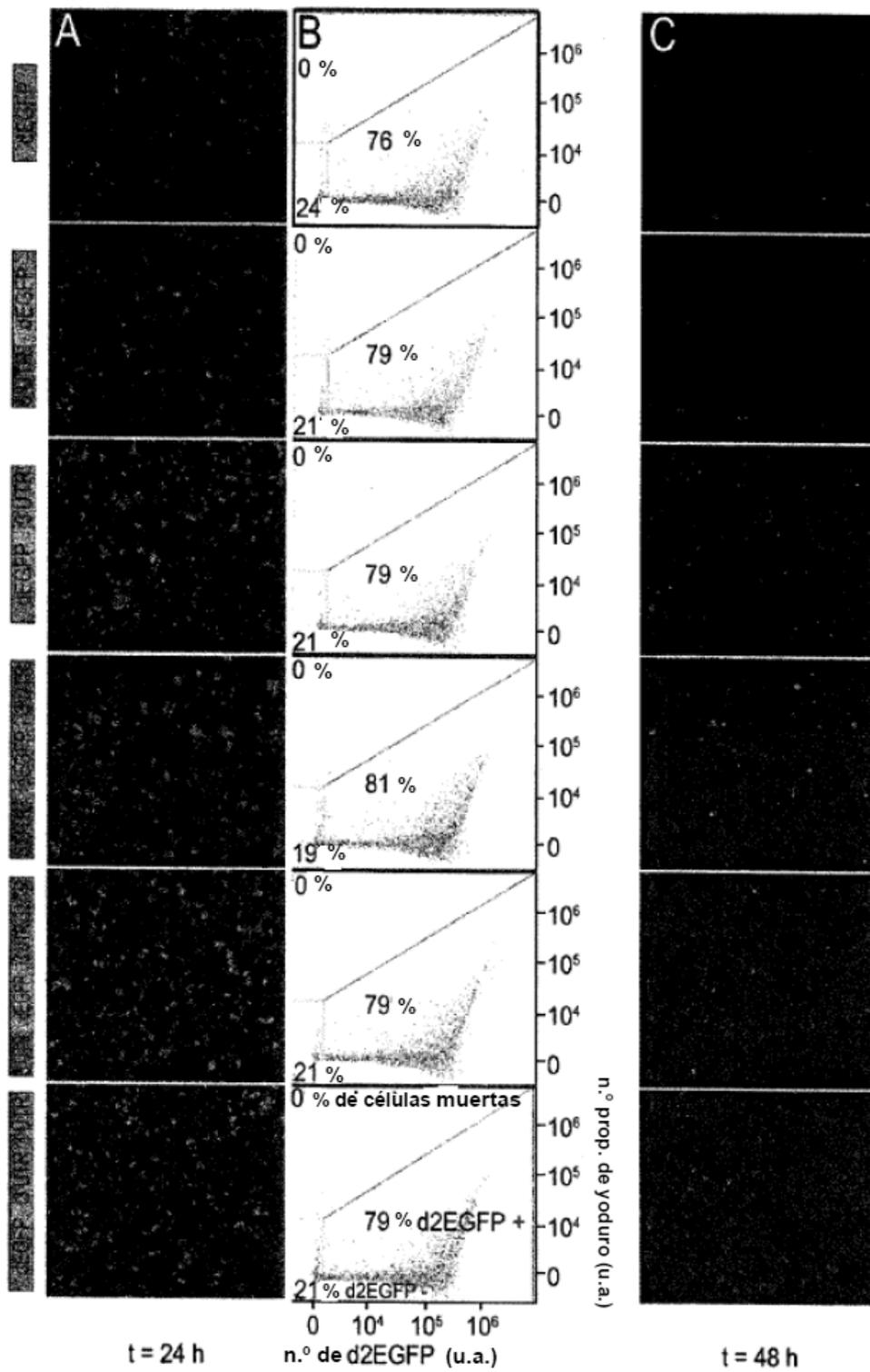


Figura 7

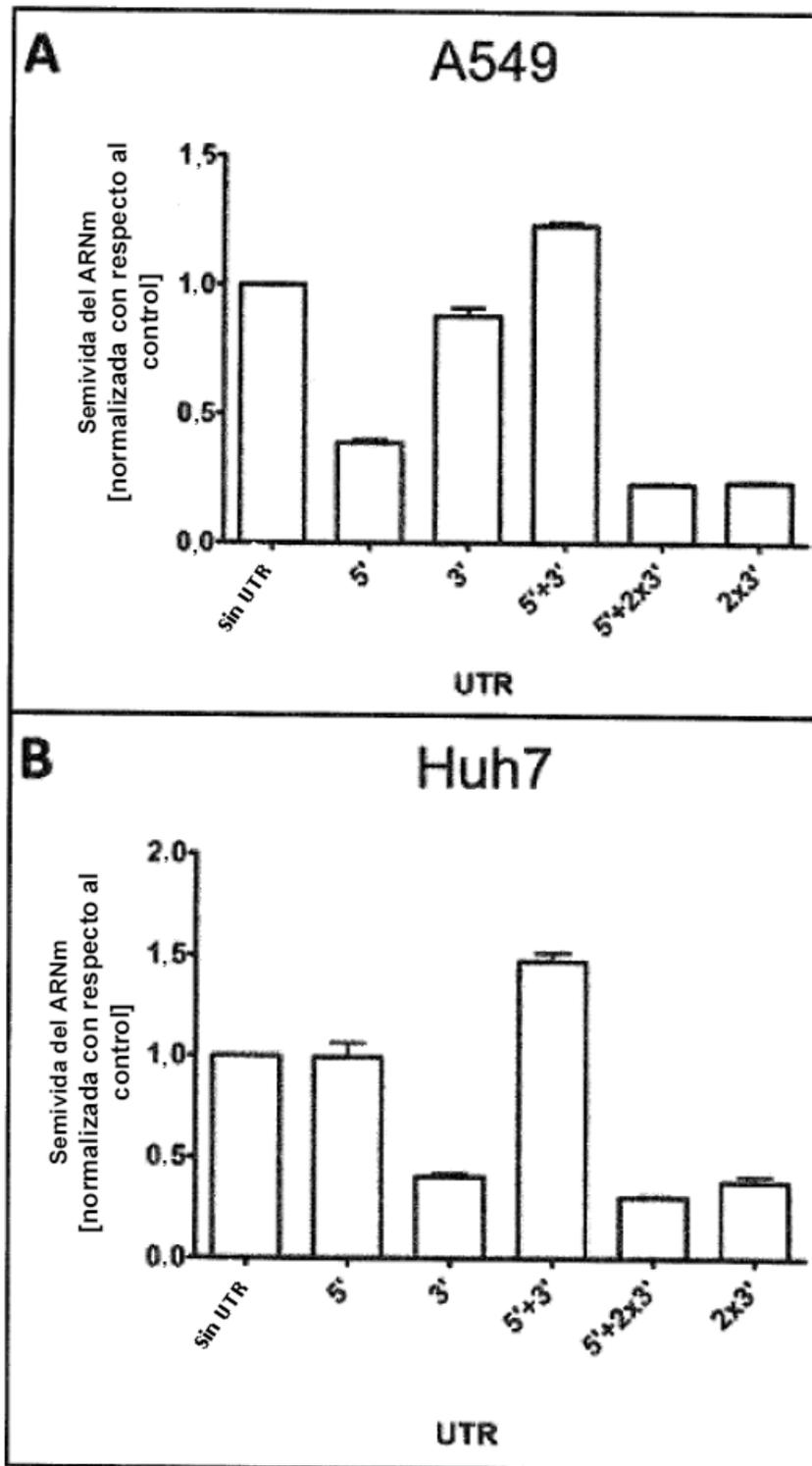


Figura 8

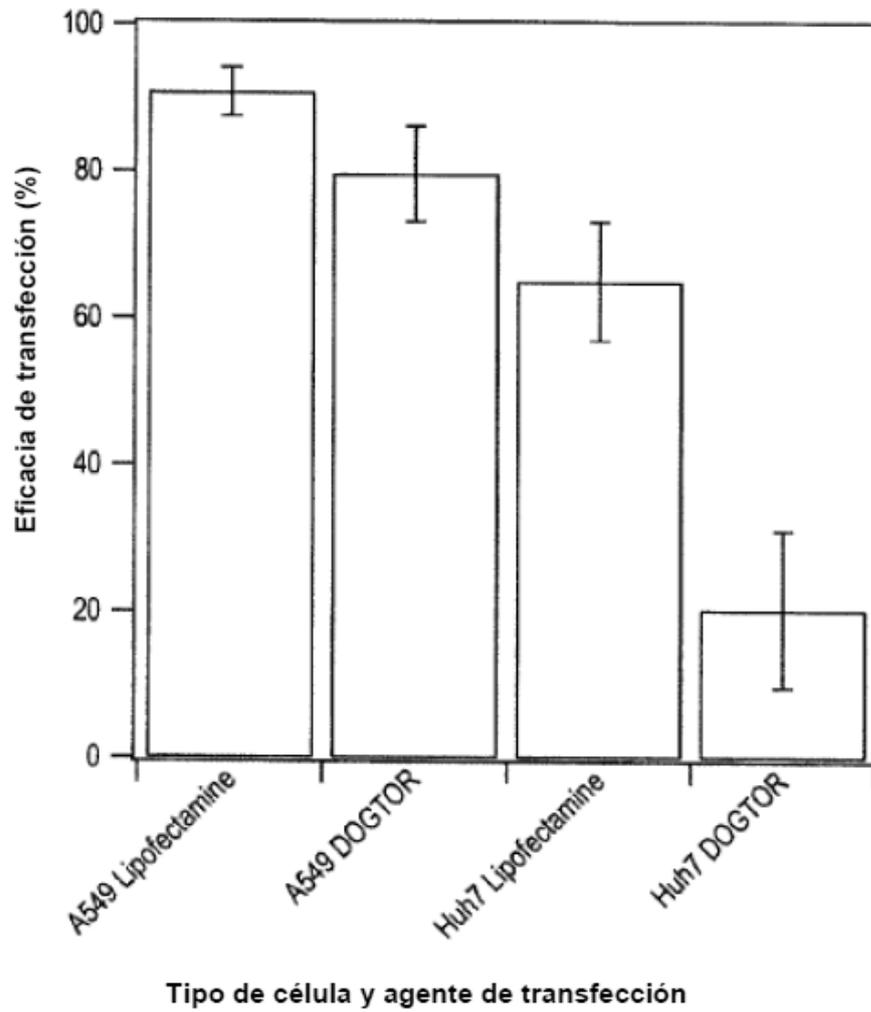


Figura 9

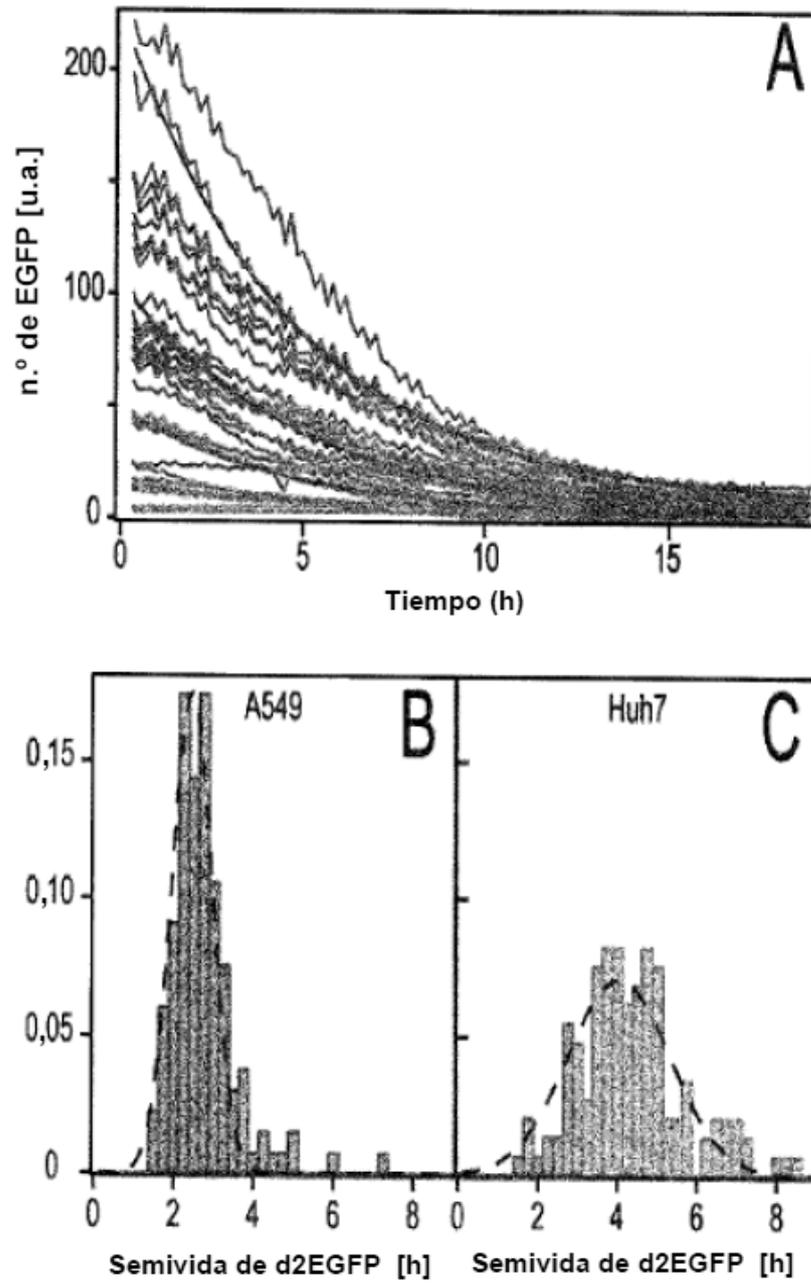


Figura 10

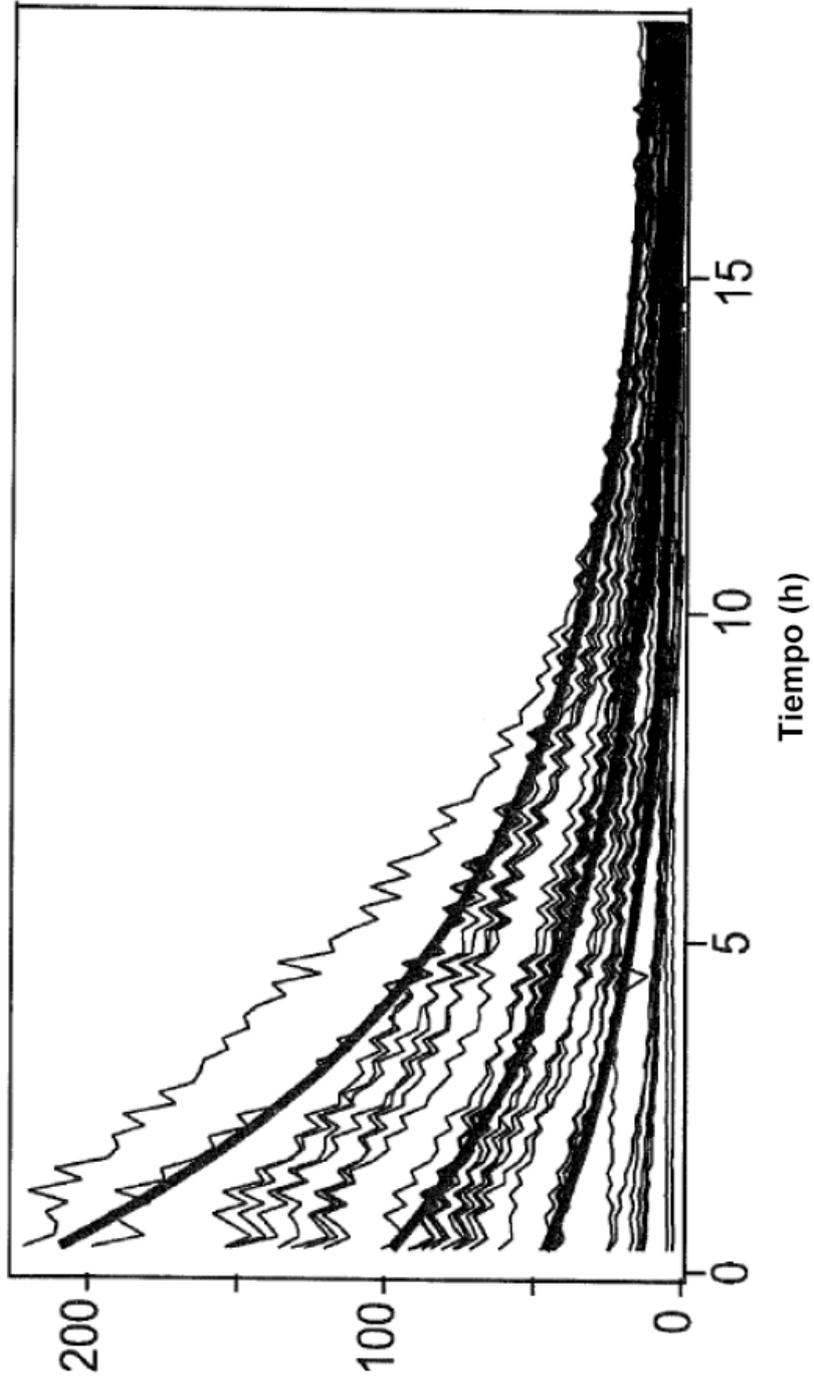


Figura 10A

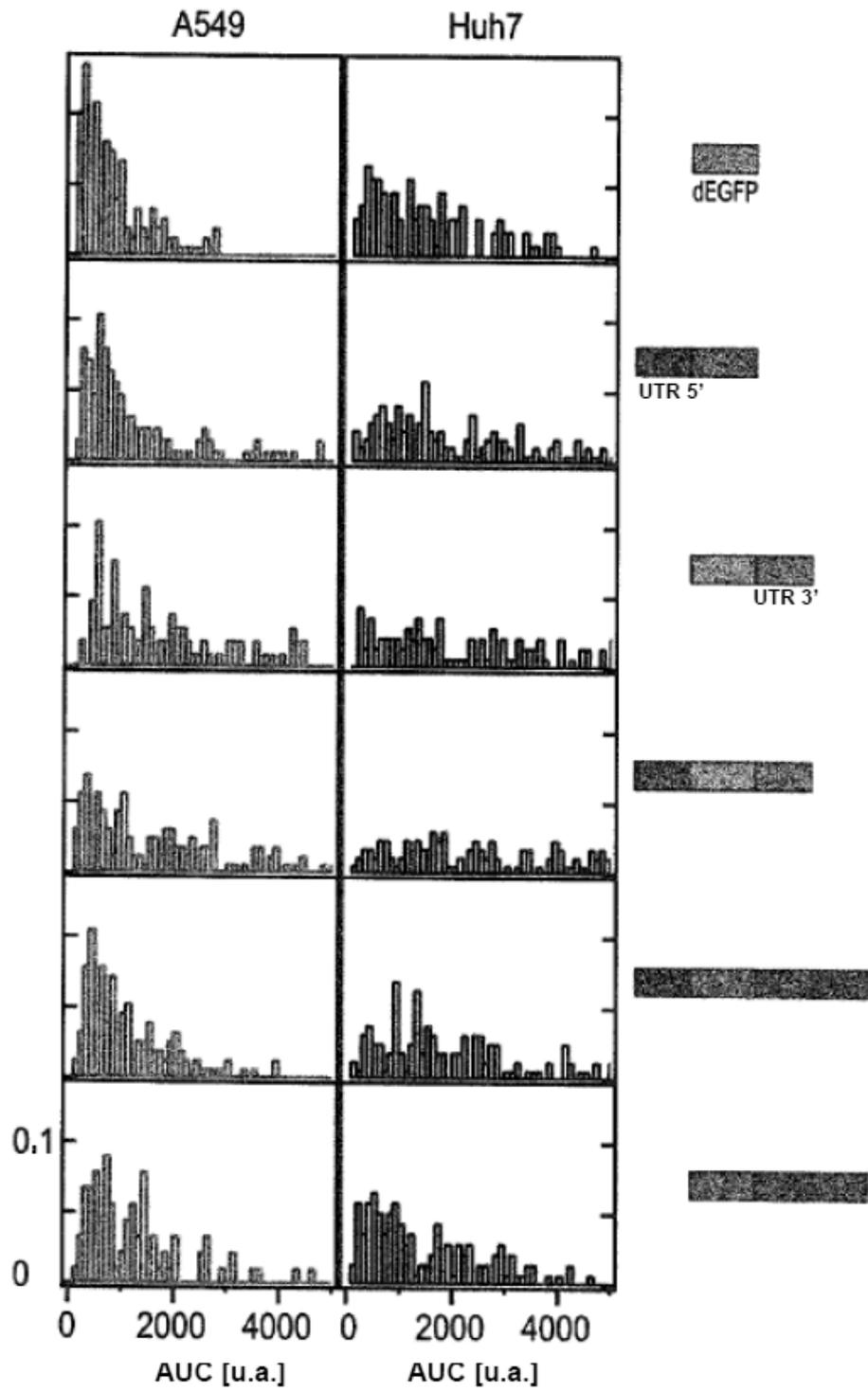


Figura 11

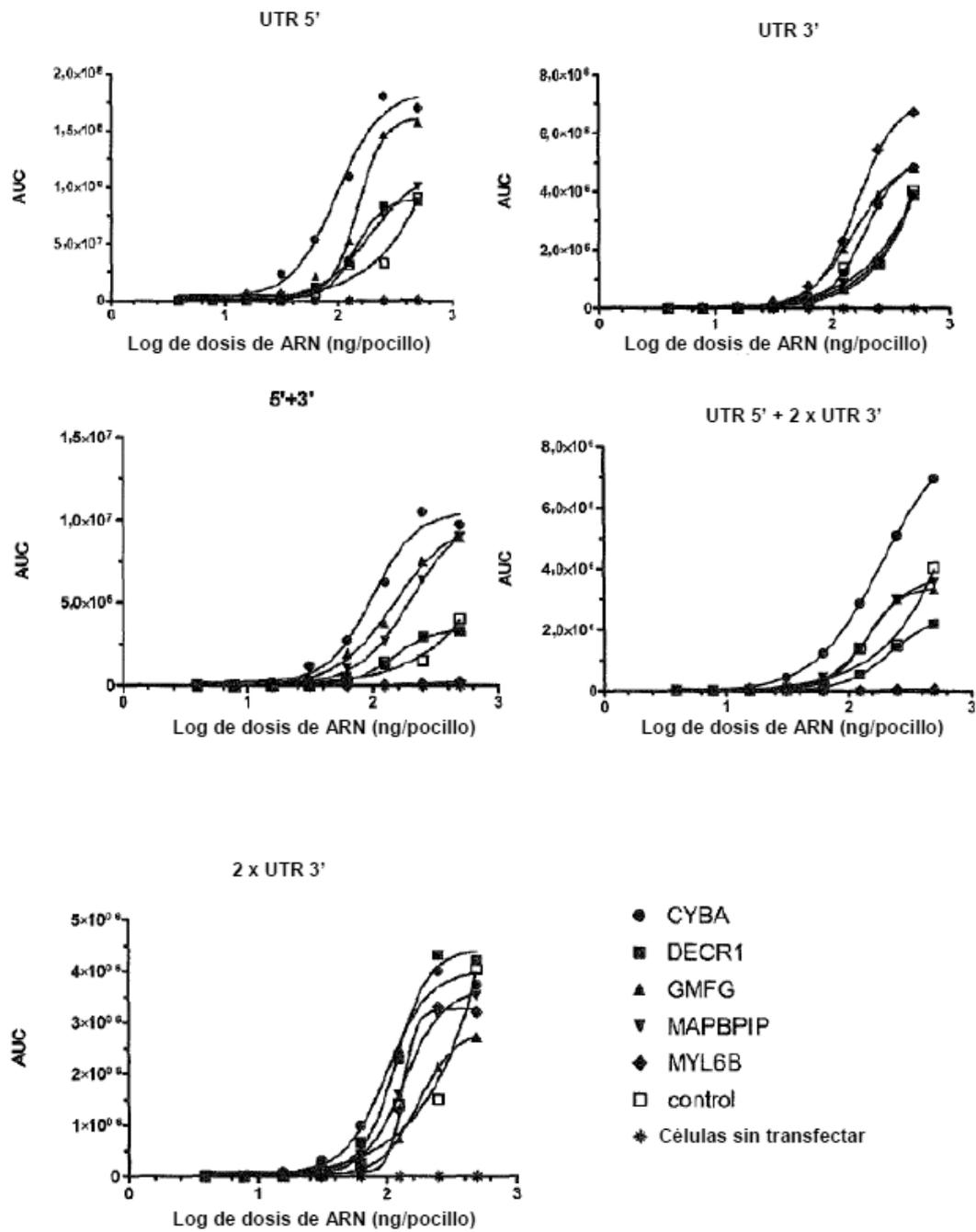


Figura 12: Expresión de MetLuc en AUC – Comparación de UTR celulares en NIH3T3