

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 614 402**

51 Int. Cl.:

C12N 15/85 (2006.01)

C07H 21/04 (2006.01)

A01K 67/027 (2006.01)

C12N 15/63 (2006.01)

C12N 15/87 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **26.09.2008** **PCT/US2008/078028**

87 Fecha y número de publicación internacional: **02.04.2009** **WO09042971**

96 Fecha de presentación y número de la solicitud europea: **26.09.2008** **E 08833086 (5)**

97 Fecha y número de publicación de la concesión europea: **09.11.2016** **EP 2205618**

54 Título: **5'UTR sintéticas, vectores de expresión y métodos para aumentar la expresión transgénica**

30 Prioridad:

26.09.2007 US 975407 P

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
31.05.2017

73 Titular/es:

INTREXON CORPORATION (100.0%)
1750 Kraft Drive, Suite 1400
Blacksburg, VA 24060, US

72 Inventor/es:

REED, THOMAS

74 Agente/Representante:

VALLEJO LÓPEZ, Juan Pedro

ES 2 614 402 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

5'UTR sintéticas, vectores de expresión y métodos para aumentar la expresión transgénica

5 Antecedentes de la invención

Campo de la invención

10 La presente invención pertenece al campo de la biotecnología. En particular, se refiere a mejoras en el control de la expresión génica posterior a la transcripción en células eucariotas.

Antecedente de la invención

15 La expresión génica eucariota pasa por varios puntos de control tras la transcripción del ARNm primario a partir del ADN. El transcrito de ARNm primario comprende porciones codificantes (exones) y porciones no codificantes (intrones). Durante el corte y empalme del ARNm, los intrones se cortan y retiran del transcrito y los exones se unen entre sí para generar ARN mensajero maduro (ARNm). El corte y empalme sirve como punto de control para generar múltiples isoformas de proteínas a partir de un único gen mediante la adición y retirada de exones en diversas combinaciones. Este proceso, denominado corte y empalme alternativo, se produce en estructuras multicomponentes muy estrechamente reguladas denominadas ayustosomas, que están bajo el control de rutas de señalización intracelulares y extracelulares.

25 El corte y empalme alternativo en la región de codificación de una proteína puede dar como resultado la generación de múltiples isoformas con diversas funciones. Adicionalmente, el corte y empalme ha mostrado aumentar drásticamente la síntesis de proteínas en células de mamíferos (Huang y Gorman, 1990 Nucleic Acids Research 18(4):937-947). El mecanismo para esto es desconocido. El corte y empalme alternativo puede producirse también en las regiones no traducidas del transcrito, que puede contribuir a dominios potenciadores o de estabilización en el transcrito final, dando como resultado una traducción aumentada de proteínas.

30 La adición de elementos de corte y empalme a la región 5' reguladora en una construcción de gen sintético ha mostrado aumentar la expresión génica, teóricamente como resultado del transporte mejorado del ARNm desde el núcleo al citoplasma (Huang y Gorman, *supra*; Choi et al., 1991 Molecular and Cellular Biology 11(6):3070-3074). Como resultado de este trabajo, se incluyen a menudo intrones entre el promotor y el sitio de clonación múltiple de vectores de expresión de mamíferos comercialmente disponibles. Sin embargo, no se han evaluado las combinaciones de intrones con otras regiones reguladoras para aumentar la expresión génica.

Sumario de la invención

40 La presente invención proporciona secuencias de polinucleótidos 5'UTR sintéticas diseñadas para aumentar la expresión de un componente del transgén de la construcción de un gen sintético en una célula hospedadora como se define en las reivindicaciones. Sin pretender quedar vinculado por teoría alguna, las 5'UTR sintéticas se diseñan con el fin de que pueda aumentarse la expresión de un transgén mediante el transporte y estabilidad aumentados del ARN.

45 La secuencia 5'UTR sintética comprende una secuencia de polinucleótido quimérica que comprende un fragmento de polinucleótido que comprende un sitio de corte y empalme del gen de la calcio ATPasa del retículo sarcoplásmico/endoplasmático y un fragmento de polinucleótido que comprende al menos una porción de una 5'UTR de un gen de la caseína.

50 Las secuencias del polinucleótido 5'UTR sintéticas de la invención tienen utilidad para aumentar la expresión de una secuencia de interés o región de codificación de interés en una construcción de gen sintético. La secuencia 5'UTR sintética puede insertarse en vectores víricos o no víricos entre un promotor y una secuencia del nucleótido de interés utilizando técnicas de ADN recombinante. Las secuencias 5'UTR sintéticas están flanqueadas opcionalmente por secuencias de nucleótidos que comprenden sitios de la endonucleasa de restricción y otros nucleótidos necesarios para la actividad de la endonucleasa de restricción. Las secuencias flanqueantes proporcionan opcionalmente sitios de clonación en un vector.

La presente invención proporciona también vectores que comprenden 5' UTR sintéticas. En una realización de la invención, el vector es un vector de expresión eucariota.

60 La presente divulgación proporciona también métodos para aumentar la expresión de un transgén en una célula eucariota. Los métodos comprenden las etapas de crear una secuencia 5'UTR sintética fusionando un fragmento de polinucleótido o un primer gen eucariota que comprende un sitio de corte y empalme y un fragmento de polinucleótido de un segundo gen eucariota que comprende al menos una porción de una 5'UTR para crear una secuencia de polinucleótido quimérica, e insertar la secuencia de polinucleótido quimérica en un vector de expresión entre un promotor y una secuencia de interés.

- El solicitante ha realizado el sorprendente descubrimiento de que una secuencia 5'UTR sintética creada mediante fusión de un fragmento de nucleótido que comprende un intrón de un gen de la calcio ATPasa del retículo sarcoplásmico/endoplasmático con un fragmento que comprende al menos una porción de un gen de la caseína da como resultado una expresión génica aumentada. Tal como se describe en detalle a en el presente documento, dos realizaciones diferentes de una 5'UTR sintética aumentaron la expresión de un gen indicador, en comparación con la del control, en dos tipos de células diferentes transfectadas con un vector de expresión que comprende la 5'UTR sintética.
- 10 Por lo tanto, es un objetivo de la divulgación proporcionar una secuencia 5'UTR sintética que comprende un fragmento de polinucleótido que comprende un sitio de corte y empalme fusionado a un fragmento de polinucleótido que comprende al menos una porción de una región 5' no traducida heteróloga para aumentar la expresión de un transgén en una célula eucariota.
- 15 Es otro objetivo de la divulgación proporcionar una secuencia 5'UTR sintética que comprende un fragmento de polinucleótido que comprende un intrón fusionado a un fragmento de polinucleótido que comprende al menos una porción de una región 5' no traducida heteróloga para aumentar la expresión de un transgén en una célula eucariota.
- 20 Es otro objetivo de la divulgación proporcionar una secuencia 5'UTR sintética que comprende un fragmento de polinucleótido que comprende un intrón que incluye porciones 5' y 3' flanqueantes de exones adyacentes fusionados a un fragmento de polinucleótido que comprende al menos una porción de una región 5' no traducida heteróloga para aumentar la expresión de un transgén en una célula eucariota.
- 25 Es otro objetivo de la divulgación proporcionar una secuencia 5'UTR sintética que sea compatible para su inserción en un vector.
- Es otro objetivo de la invención proporcionar vectores que comprendan 5'UTR sintéticas, tal como se define en las reivindicaciones.
- 30 Es otro objetivo de la invención proporcionar células hospedadoras que comprendan 5'UTR sintéticas, tal como se define en las reivindicaciones.

Descripción de las secuencias

- 35 SEQ ID NO: 1 representa una realización de una secuencia 5'UTR sintética que comprende: un sitio de restricción MluI, SEQ ID NO: 2, un sitio de restricción KpnI, SEQ ID NO: 3, un sitio de restricción MfeI. SEQ ID NO: 1 se conoce también en el presente documento como 5U2.
- 40 SEQ ID NO: 2 representa una realización de una secuencia del intrón 2 de SERCA2 de canino con un presunto sitio poli-A de consenso mutado, con una porción de exón 2 flanqueando en el extremo 5' y una porción del exón 3 flanqueando en el extremo 3'. SEQ ID NO: 2 es una secuencia parcial mutada del cromosoma 26 de Canis familiaris, la secuencia shotgun del genoma completo (número de acceso público NC_006608.2).
- 45 SEQ ID NO: 3 representa una realización de una secuencia 5'UTR de caseína de bovino. SEQ ID NO: 3 es una secuencia parcial del ARNm beta de la caseína de Bos taurus de longitud completa (número de acceso público NM_181008).
- 50 SEQ ID NO: 4 representa una realización de la secuencia del intrón 2 de SERCA2 natural de canino, con una porción de exón 2 flanqueando en el extremo 5' y una porción del exón 3 flanqueando en el extremo 3'. SEQ ID NO: 4 es una secuencia parcial del cromosoma 26 de Canis familiaris, la secuencia shotgun del genoma completo (número de acceso público NC_006608.2).
- 55 SEQ ID NO: 5 representa una realización de la secuencia del intrón 2 de SERCA2 natural de ser humano, con el exón 2 flanqueando en el extremo 5' y el exón 3 flanqueando en el extremo 3'. SEQ ID NO: 5 es una secuencia parcial del cromosoma 12 de Homo sapiens, ensamblaje de referencia, secuencia completa (número de acceso público NC_000012).
- 60 SEQ ID NO: 6 representa una realización de la secuencia del intrón 2 de SERCA2 natural de ratón, con el exón 2 flanqueando en el extremo 5' y el exón 3 flanqueando en el extremo 3'. SEQ ID NO: 6 es una secuencia parcial del cromosoma 5 de Mus musculus, ensamblaje de referencia (número de acceso público NC_000071).
- 65 SEQ ID NO: 7 representa una realización de una secuencia 5'UTR sintética que comprende un sitio de restricción AscI, un sitio de restricción MluI, SEQ ID NO: 4, un sitio de restricción KpnI, SEQ ID NO: 3, un sitio de restricción MfeI. SEQ ID NO: 7 se conoce también en el presente documento como INXN-1.
- SEQ ID NO: 8 representa una realización de una secuencia 5'UTR de caseína de ratón. SEQ ID NO: 8 es una

secuencia parcial de la beta caseína de *Mus musculus*, ARNm (ADNc clon MGC:91065) (número de acceso público BC080709).

SEQ ID NO: 9 representa una realización de una secuencia 5'UTR de caseína de rata. SEQ ID NO: 9 es una secuencia parcial de la beta caseína de *Rattus norvegicus* (Csn2), ARNm (número de acceso público NM_017120).

SEQ ID NO: 10 representa una realización de una secuencia 5'UTR de caseína de oveja. SEQ ID NO: 10 es una secuencia parcial de la beta caseína de *Ovis aries* (CSN2), ARNm, (número de acceso público NM_001009373).

SEQ ID NO: 11 representa el exón 3 de SERCA2 de canino. SEQ ID NO: 11 es una secuencia parcial del cromosoma 26 de *Canis familiaris*, la secuencia shotgun del genoma completo (número de acceso público NC_006608.2).

SEQ ID NO: 12 representa una realización de una secuencia de vector que comprende la 5'UTR sintética. El vector representado por la SEQ ID NO: 12 comprende la SEQ ID NO: 1 y se representa esquemáticamente en la Figura 10.

SEQ ID NO: 13 representa otra realización de una secuencia de vector que comprende una 5'UTR sintética. El vector representado por la SEQ ID NO: 13 comprende la SEQ ID NO: 7 y se representa esquemáticamente en la Figura 11.

SEQ ID NO: 14 representa un vector que comprende una 5'UTR sintética del control (poliG) y se representa esquemáticamente en la Figura 9.

En cualquiera de estas secuencias, T (timidina) puede sustituirse con U (Uracilo).

Descripción de los dibujos

La Figura 1A muestra una representación esquemática del polinucleótido de la SEQ ID NO: 4.

La Figura 1B muestra una representación esquemática del polinucleótido de la SEQ ID NO: 5 y el polinucleótido de la SEQ ID NO: 6.

La Figura 1C muestra los polinucleótidos de la SEQ ID NO: 2 y las SEQ ID NO: 4-6. El segundo intrón de SERCA2 está resaltado en negro. Los exones adyacentes o sus porciones están sin resaltar.

La Figura 2A muestra una representación esquemática del polinucleótido de la SEQ ID NO: 1.

La Figura 2B muestra una representación esquemática del polinucleótido de la SEQ ID NO: 7.

La Figura 3A muestra una representación esquemática de una 5'UTR sintética en un vector de expresión entre un promotor y una secuencia de interés.

La Figura 3B muestra una representación esquemática de una 5'UTR sintética en un vector de expresión entre un promotor y un sitio de clonación.

La Figura 4 muestra los resultados del ensayo de las realizaciones de 5'UTR sintéticas de la invención en células HEK-293.

La Figura 5 muestra los resultados del ensayo de las realizaciones de 5'UTR sintéticas de la invención en células 1080.

La Figura 6 muestra los resultados del ensayo de las realizaciones de 5'UTR sintéticas de la invención a medida que aumenta el plegamiento con respecto al control en células HEK-293 y células 1080, en el que los valores del control se normalizaron a 1.

La Figura 7 muestra un vector del control usado en el Ejemplo 1 (VVN-2712), en el que la secuencia de codificación de la beta-galactosidasa (LacZ) carece de 5'UTR y se une de manera operativa al promotor de CMV.

La Figura 8 muestra un vector del control usado en el Ejemplo 1 (VVN-2713), en el que el vector carece de una 5'UTR y de LacZ.

La Figura 9 muestra un vector utilizado en el Ejemplo 1 (VVN-8318), en el que la secuencia de codificación de la beta-galactosidasa (LacZ) se une de manera operativa a poliG 5'UTR y al promotor de CMV.

La Figura 10 muestra un vector utilizado en el Ejemplo 1 (VVN-8277), en el que la secuencia de codificación de la

beta-galactosidasa (LacZ) se une de manera operativa a 5'UTR de la invención (5U2) y al promotor de CMV.

La Figura 11 muestra un vector utilizado en el Ejemplo 1 (VVN-8276), en el que la secuencia de codificación de la beta-galactosidasa (LacZ) se une de manera operativa a 5'UTR de la invención (INXN-1) y al promotor de CMV.

5 La Figura 12 es una tabla que contiene los datos del Ejemplo 1 representado en las Figuras 4-6.

La Figura 13 muestra una porción de una alineación de las secuencias genómicas y de ARNm de SERCA2 de Equus caballus que incluyen el segundo intrón y el exón 2 y el exón 3. Los extremos 5' y 3' del segundo intrón se indican mediante flechas.

10 Las Figuras 7-11 utilizan las siguientes abreviaturas: CMV pro = promotor del Citomegalovirus, LacZ = secuencia de codificación de LacZ, SV40pA = SV40 poliA, Amp= Gen de resistencia a la ampicilina, Neo= Gen de resistencia a la neomicina, MCS = Sitio de Clonación Múltiple, SPL-1 = porción de exón 2 SERCA2+ intrón 2 SERCA2 + porción de exón 3 SERCA2, UTR-1 = porción de 5'UTR de la caseína.

15 Descripción detallada de la invención

Se aplicarán las siguientes definiciones a toda la descripción, los dibujos, y las reivindicaciones que siguen. Sin embargo, los términos utilizados en la memoria descriptiva y en las reivindicaciones no definidas en el presente documento tienen los significados habituales que se entienden en la técnica.

20 Cuando se utilizan los términos "uno", "una" o "unos" en esta divulgación, significan "al menos uno" o "uno o más" a no ser que se indique otra cosa.

25 "Ácido nucleico", "molécula de ácido nucleico", "secuencia de ácido nucleico", "oligonucleótido", "secuencia de oligonucleótidos", "secuencia de nucleótidos", "polinucleótido", y "secuencia de polinucleótidos" se utilizan de forma indistinta y se refieren a la forma polimérica de éster fosfato de ribonucleósidos (adenosina, guanosina, uridina o citidina; "moléculas de ARN") o desoxirribonucleósidos (desoxiadenosina, desoxiguanosina, desoxitimidina, o desoxicitidina; "moléculas de ADN"), o cualesquiera de sus análogos de fosfoéster, tal como fosforotioatos y tioésteres, en cualquier forma monocatenaria, o una hélice bicatenaria. ADN-ADN bicatenario, son posibles hélices de ADN-ARN y ARN-ARN. El término molécula de ácido nucleico, y en particular molécula de ADN o ARN, se refiere solamente a la estructura primaria y secundaria de la molécula, y no la limita a cualesquiera formas terciarias concretas. Por lo tanto, este término incluye ADN bicatenario que se encuentra, entre otras, en moléculas de ADN lineales o circulares (por ejemplo, fragmentos de restricción), plásmidos, ADN y cromosomas superenrollados. En la descripción de la estructura de moléculas de ADN bicatenarias, en el presente documento se pueden describir secuencias de acuerdo con la convención normal de proporcionar solo la secuencia en la dirección 5' a 3' a lo largo de la hebra de ADN no transcrita (es decir, la hebra que tiene una secuencia homóloga al ARNm). Una "molécula de ADN recombinante" es una molécula de ADN que ha experimentado una manipulación biológica molecular. El ADN incluye, pero no de forma limitativa ADNc, ADN genómico, ADN plásmido, ADN sintético, y ADN semisintético.

40 Los términos "fragmento" utilizados junto con una secuencia de polinucleótidos (por ejemplo, "fragmento de polinucleótido") se refiere a una secuencia de nucleótidos de longitud reducida con respecto al ácido nucleico de referencia y que comprende, sobre la porción común, una secuencia de nucleótidos idéntica al ácido nucleico de referencia. Dicho fragmento de ácido nucleico de acuerdo con la invención puede, cuando sea adecuado, incluirse en un polinucleótido más grande del cual es un constituyente. Dichos fragmentos comprenden, o consisten alternativamente de, polinucleótidos con una longitud comprendida desde al menos 6, 8, 9, 10, 12, 15, 18, 20, 21, 22, 23, 24, 25, 30, 39, 40, 42, 45, 48, 50, 51, 54, 57, 60, 63, 66, 70, 75, 78, 80, 90, 100, 105, 120, 135, 150, 200, 300, 500, 720, 900, 1000 o 1500 nucleótidos consecutivos de un ácido nucleico de acuerdo con la invención.

50 El término "quimérico" significa comprendido por fragmentos que no son contiguos en su estado natural. Por ejemplo, un polinucleótido quimérico significa un polinucleótido que comprende fragmentos que no son contiguos en su estado natural.

55 El término "sintético" utilizado junto con una secuencia de polinucleótidos es un polinucleótido (o porción de un polinucleótido) no natural que difiere de una secuencia de polinucleótidos natural. Por ejemplo, un gen (o porción de un gen) sintético puede contener una o más secuencias de ácidos nucleicos no contiguas en la naturaleza (secuencias quiméricas), y/o puede abarcar sustituciones, inserciones, y deleciones y sus combinaciones.

60 Un "gen" se refiere a un polinucleótido que comprende nucleótidos que codifican una molécula afuncional (por ejemplo, un polipéptido o ARN), e incluye ADNc o ácidos nucleicos de ADN genómico. Se entiende generalmente que el ADN genómico que codifica un polipéptido o ARN incluye regiones no codificantes (es decir, intrones) que se cortan y empalman a partir del ARNm maduro, y por tanto no están presentes en el ADNc que codifica el mismo polipéptido o ARN. "Gen" puede comprender un fragmento de ácido nucleico que expresa un ARN, proteína o polipéptido específico. El "gen" puede comprender además secuencias reguladoras que preceden (secuencias 5' no codificantes) y siguen (secuencias 3' no codificantes) a la secuencia de codificación. El "gen" puede comprender también oligonucleótidos formadores de tripletes (TFO). "Gen nativo" se refiere a un gen que se encuentra en la naturaleza con

sus propias secuencias reguladoras. "Gen quimérico" o "gen recombinante" se refiere a cualquier gen que no es un gen nativo, que comprende secuencias reguladoras y/o codificantes que no se encuentran juntas en la naturaleza. De acuerdo con ello, un gen quimérico puede comprender secuencias reguladoras y secuencias codificantes que se derivan de fuentes diferentes, o secuencias reguladoras y secuencias codificantes derivadas de la misma fuente, pero

5 dispuestas de una manera diferente a la que se encuentra en la naturaleza. Un gen quimérico puede comprender secuencias codificantes derivadas de diferentes fuentes y/o secuencias reguladoras derivadas de diferentes fuentes. "Gen endógeno" se refiere a un gen nativo en su localización natural en el genoma de un organismo.

Un gen "extraño" o gen "exógeno" o gen "heterólogo" o "transgén" se refiere a un gen que no se encuentra normalmente en la célula u organismo hospedador, pero que se introduce en la célula u organismo hospedador mediante transferencia génica. Los transgenes pueden comprender genes nativos insertados en un organismo no nativo, o genes quiméricos o sintéticos. Un transgén puede ser también una versión de ADNc de un gen endógeno. Un gen transgén puede ser también una versión no mutada de un gen endógeno mutado o una versión mutada de un gen endógeno no mutado. Un gen transgén puede ser también un gen terapéutico o un gen experimental tal como un

10 indicador. Un transgén puede introducirse directamente en las células diana de un organismo hospedador, o introducirse directamente por la transferencia de células transformadas, por ejemplo, células autólogas, en el organismo hospedador.

La "región 5 prima no traducida" o "5'UTR" de un gen debe entenderse como la parte de un gen que se transcribe en un transcrito de ARN primario (pre-ARNm) y cuya parte se localiza en la dirección 5' de la secuencia de codificación. El transcrito primario es el producto de ARN inicial, que contiene intrones y exones, producidos por la transcripción del ADN. Muchos transcritos primarios deben experimentar el procesamiento del ARN para formar las especies de ARN fisiológicamente activas. El procesamiento a un ARNm maduro puede comprender el recorte de los extremos, la eliminación de intrones, la protección y/o el corte de moléculas de ARN individuales de sus ARN precursores. La 5'UTR de un ARNm es de esta manera la parte del ARNm que no se traduce en proteína y que se localiza en la dirección 5' de la secuencia de codificación. En una secuencia genómica, la 5'UTR se define normalmente como la región entre el sitio de inicio de la transcripción y el codón de inicio. Las regiones 5' no traducidas (5'UTR) de los ARNm de vertebrados pueden tener unas pocas decenas de bases hasta algunos cientos de bases de longitud (Crowe et al., 2006 BMC Genomics 7:16).

Una "5'UTR sintética" es una 5'UTR no natural que difiere de la secuencia del polinucleótido 5'UTR. Una 5' UTR sintética puede contener una o más secuencias de ácidos nucleicos no contiguas en la naturaleza (secuencias quiméricas), y/o puede abarcar sustituciones, inserciones, y delecciones y sus combinaciones.

Una "unión de corte y empalme", "unión intrón-exón de corte y empalme", o el "sitio de corte y empalme" son regiones en los límites de un intrón en un pre-ARNm eucariota reconocidas por el aparato de corte y empalme de la célula donde se unen dos exones adyacentes y se elimina el intrón. Los sitios de corte y empalme se representan por secuencias conservadas en los límites del intrón/exón 5' y 3'. Para la gran mayoría de intrones, la mayoría de secuencias conservadas son GU que flanquean el extremo 5' del intrón y AG que flanquean el extremo 3'. Sin embargo, se conocen también excepciones a estas secuencias consenso tales como intrones con sitios de corte y empalme AU-AC. El sitio de corte y empalme 5' en el límite intrón-exón se conoce como sitio "de corte y empalme donante". El sitio de corte y empalme 3' en un límite intrón-exón se conoce como un sitio "de corte y empalme" receptor.

Un "ayustosoma" es un complejo de ribonucleoproteína grande que sirve como el aparato de corte y empalme de la célula. El ayustosoma está comprendido por subunidades de ribonucleoproteínas nucleares pequeñas (RNPsn) que se ensamblan sobre un sustrato pre-ARNm. Las propias RNPsn están comprendidas por ARN nucleares pequeños (ARNsn) y algunas subunidades de proteínas. Durante la reacción de corte y empalme, el reconocimiento de sitios de corte y empalme en el pre-ARNm se lleva a cabo mediante emparejamiento de bases con los ARNsn.

ADN "heterólogo" se refiere a ADN que no se localiza naturalmente en la célula, o en un sitio cromosómico de la célula. Por lo tanto, el ADN heterólogo incluye un gen extraño a la célula. El ADN "heterólogo" puede incluir también un gen que existe naturalmente en la célula, pero que se localiza en una localización no nativa. Además, una molécula de ADN "heteróloga" puede ser una molécula de ADN que contiene un segmento de ADN que no es del hospedador, unido de manera operativa a un segmento de ADN del hospedador, por ejemplo, un promotor de la transcripción. Por el contrario, una molécula de ADN heteróloga puede comprender un gen endógeno unido de manera operativa con un promotor exógeno. Adicionalmente, "heterólogo" puede referirse a una molécula o fragmento de ADN que se deriva de un gen que no comparte un origen evolutivo común con una molécula o fragmento de ADN de referencia.

El término "genoma" incluye ADN o ARN cromosómico, así como mitocondrial, de cloroplastos y vírico.

El término "sonda" se refiere a una molécula de ácido nucleico monocatenario que puede emparejar sus bases con un ácido nucleico diana monocatenario complementario para formar una molécula bicatenaria.

Una "secuencia codificante" de ADN se refiere a una secuencia de ADN bicatenario que codifica un polipéptido y que se puede transcribir y traducir en un polipéptido en una célula in vitro o in vivo o en el exterior de una célula, por ejemplo, en un tubo, cuando se coloca bajo el control de secuencias reguladoras adecuadas. "Secuencias reguladoras

adecuadas" se refiere a secuencias de nucleótidos localizadas en la dirección 5' (secuencias 5' no codificantes), en su interior, o en la dirección 3' (secuencias 3' no codificantes) de una secuencia codificante, y que alteran la transcripción, el procesamiento o la estabilidad del ARN, o la traducción de la secuencia codificante asociada. Las secuencias reguladoras pueden incluir promotores, secuencias líder de traducción, intrones, secuencias de reconocimiento de poliadenilación, sitio de procesamiento del ARN, sitio de unión al efector y estructura tallo-bucle. Los límites de la secuencia de codificación están determinados por un codón de inicio en el extremo 5' (amino) y un codón de detención de la traducción en el extremo 3' (carboxilo). Una secuencia codificante puede incluir, pero no de forma limitativa, secuencias procariotas, eucariotas, o quiméricas, ADNc procedente de ARNm, secuencias de ADN genómico, e incluso secuencias de ADN sintético.

"Marco de lectura abierto" se abrevia ORF y se refiere a la longitud de una secuencia de ácido nucleico, tanto ADN, ADNc o ARN, que comprende una señal de inicio de la traducción o codón de inicio, tal como un ATG o AUG, y un codón de terminación y puede traducirse potencialmente en una secuencia polipeptídica.

El término "en la dirección 3'" se refiere a una secuencia de nucleótidos que se localiza en 3' respecto a una secuencia de nucleótidos de referencia. En particular, las secuencias de nucleótidos en la dirección 3' se refieren generalmente a secuencias que siguen al punto de inicio de la transcripción. Por ejemplo, el codón de inicio de la traducción de un gen se localiza en la dirección 3' del sitio de inicio de la transcripción.

El término "en la dirección 5'" se refiere a una secuencia de nucleótidos que se localiza en 5' respecto a una secuencia de nucleótidos de referencia. En particular, secuencias de nucleótidos en la dirección 5' se refiere generalmente a secuencias que se localizan en el lado 5' de una secuencia de codificación del punto de partida de la transcripción. Por ejemplo, la mayoría de los promotores se localizan en la dirección 5' del sitio de inicio de la transcripción.

"Sintetizado químicamente", en relación a una secuencia de ADN, significa que los nucleótidos componentes se ensamblaron in vitro. Se puede llevar a cabo la síntesis química manual del ADN utilizando procedimientos bien establecidos, o se puede llevar a cabo la síntesis química automatizada utilizando una de numerosas máquinas comercialmente disponibles. De acuerdo con ello, los genes pueden personalizarse para la expresión génica óptima basándose en la optimización de la secuencia de nucleótidos para reflejar el sesgo del codón de la célula hospedadora. El técnico experto aprecia la probabilidad de una expresión génica satisfactoria si la utilización del codón está sesgada hacia aquellos codones favorecidos por el hospedador. La determinación de los codones preferidos puede basarse en una vigilancia de los genes derivados de la célula hospedadora cuando está disponible información de la secuencia.

Los términos "endonucleasa de restricción" y "enzima de restricción" se usan de manera indistinta y se refieren a una enzima que se une y corta en una secuencia de nucleótidos específica en el ADN bicatenario.

"Polipéptido", "péptido" y "proteína" se usan de manera indistinta y se refieren a un compuesto polimérico comprendido por restos aminoácidos unidos covalentemente.

"Reacción en cadena de la polimerasa" se abrevia PCR y se refiere a un método in vitro para amplificar enzimáticamente secuencias de ácido nucleico específicas. PCR implica una serie repetitiva de ciclos de temperatura comprendiendo cada ciclo tres etapas: desnaturalización del molde de ácido nucleico para separar las hebras de la molécula diana, hibridación de un cebador de oligonucleótidos de la PCR monocatenario para el molde de ácido nucleico, y extensión del (de los) cebador(es) hibridado(s) mediante la ADN polimerasa.

El término "homología" se refiere al porcentaje de identidad entre dos restos de polinucleótidos o dos restos polipeptídicos. La correspondencia entre la secuencia entre un resto y otro puede determinarse mediante técnicas conocidas en la materia. Por ejemplo, se puede determinar la homología mediante una comparación directa de la información de la secuencia entre dos moléculas de polipéptidos alineando la información y utilizando programas informáticos fácilmente disponibles. Como alternativa, se puede determinar la homología mediante la hibridación de polinucleótidos en condiciones en que formen dupletes estables entre regiones homólogas, seguido por digestión con nucleasa(s) específicas monocatenarias y determinación del tamaño de los fragmentos digeridos.

Como se usa en el presente documento, el término "homólogo" en todas sus formas gramaticales y variantes ortográficas se refiere a la relación entre proteínas que poseen un "origen evolutivo común", incluyendo proteínas de superfamilias (por ejemplo, la superfamilia de las inmunoglobulinas) y proteínas homólogas de diferentes especies (por ejemplo, una cadena ligera de miosina, etc.) (Reeck et al., Cell 50:667 (1987)). Dichas proteínas (y sus genes codificantes) tienen homología de secuencia, según se refleja por su elevado grado de similitud de secuencia. Sin embargo, en el uso común y en la presente aplicación, el término "homólogo", cuando se modifica con un adverbio tal como "altamente", puede referirse a la similitud de secuencia y no a un origen evolutivo común. 12

De acuerdo con ello, el término "similitud de secuencias" en todas sus formas gramaticales se refiere al grado de identidad o correspondencia entre secuencias de ácidos nucleicos o aminoácidos de proteínas que pueden compartir o no un origen evolutivo común (véase Reeck et al., Cell 50:667 (1987)). En una realización, dos secuencias de ADN son "sustancialmente homólogas" o "sustancialmente similar" cuando al menos aproximadamente un 21 %

(preferentemente al menos aproximadamente un 50 %, y de forma más preferente al menos un 75 %, 90 %, 95 %, 96 %, 97 %, 98 %, o 99 %) de los nucleótidos se corresponden sobre la longitud definida de las secuencias de ADN. Se pueden identificar secuencias que son sustancialmente homólogas comparando las secuencias utilizando software estándar disponible en los bancos de datos de secuencias, o en un experimento de hibridación Southern en, por ejemplo, condiciones rigurosas como se define para este sistema concreto. Definir las condiciones de hibridación adecuadas está comprendido en los conocimientos del técnico experto (véase, por ejemplo, Sambrook et al., 1989, *infra*).

Como se usa en el presente documento, "sustancialmente similar" se refiere a fragmentos de ácidos nucleicos, en donde los cambios en una o más bases de nucleótidos dan como resultado la sustitución de uno o más aminoácidos, pero que no afectan a las propiedades funcionales de la proteína codificada por la secuencia de ADN. "Sustancialmente similar" se refiere también a fragmentos de ácidos nucleicos, en donde los cambios en una o más bases de nucleótidos no afectan a la capacidad del fragmento de ácido nucleico de mediar en la alteración de la expresión génica mediante la tecnología de sentido contrario o cosupresión. "Sustancialmente similar" se refiere también a modificaciones de los fragmentos de ácidos nucleicos de la presente invención tales como delección o inserción de una o más bases de nucleótidos que no afectan sustancialmente las propiedades funcionales del transcrito resultante. Se entiende por tanto que la invención abarca más de las secuencias ilustrativas específicas. Cada una de las modificaciones propuestas está comprendida en los conocimientos rutinarios del experto en la materia, como es la determinación de la retención de la actividad biológica de los productos codificados.

Además, el técnico experto reconoce que las secuencias sustancialmente similares abarcadas por la presente invención se definen también por su capacidad de hibridarse, en condiciones rigurosas. Una molécula de ácido nucleico es "-hibrizable" con otra molécula de ácido nucleico, tal como ADNc, ADN genómico, o ARN, cuando una forma monocatenaria de la molécula de ácido nucleico puede hibridarse con la otra molécula de ácido nucleico en las condiciones adecuadas de temperatura y fuerza iónica de la solución (véase Sambrook et al., 1989 *infra*). La hibridación y las condiciones de lavado son bien conocidas y se ilustran en Sambrook, J., Fritsch, E. F. y Maniatis, T. Molecular Cloning. A Laboratory Manual, Segunda Edición, Cold Spring Harbor Laboratory Press, Cold Spring Harbor Press (1989), particularmente el Capítulo 11 y la Tabla 11.1 del anterior. Las condiciones de temperatura y fuerza iónica determinan el "rigor" de la hibridación.

Se pueden ajustar las condiciones de rigor para seleccionar fragmentos moderadamente similares, tales como secuencias homólogas de organismos relacionados de forma distante, con fragmentos muy similares, tales como genes que duplican las enzimas funcionales de organismos estrechamente relacionados. Para la selección preliminar de ácidos nucleicos homólogos, se pueden seleccionar condiciones de hibridación de rigor bajo, que corresponden a una Tm de 55 °C, por ejemplo, 5XSSC, SDS al 0,1 %, leche al 0,25 %, y sin formamida; o con formamida al 30 %, 5XSSC, SDS al 0,5 %. Las condiciones de hibridación con un moderado rigor corresponden a una Tm mayor, por ejemplo, formamida al 40 %, con 5X o 6XSSC. Las condiciones de hibridación con un elevado rigor corresponden a la Tm más elevada, por ejemplo, formamida al 50 %, 5X o 6X SSC.

La hibridación requiere que los dos ácidos nucleicos contengan secuencias complementarias, aunque dependiendo del rigor de la hibridación, son posibles emparejamientos incorrectos entre las bases. El término "complementario" se usa para describir la relación entre bases de nucleótidos que pueden hibridarse entre sí. Por ejemplo, con respecto al ADN, adenosina es complementaria a timina y citosina es complementaria a guanina. De acuerdo con ello, la presente invención incluye también fragmentos de ácidos nucleicos aislados que son complementarios de las secuencias completas como se divulga o utiliza en el presente documento, así como con aquellas secuencias de ácidos nucleicos sustancialmente similares.

En una realización, se detectan los polinucleótidos empleando condiciones de hibridación que comprenden una etapa de hibridación a una Tm de 55 °C, y utilizando condiciones como las que se han definido anteriormente. En otra realización, la Tm es de 60 °C; en determinadas realizaciones, la Tm es de 63°C o 65°C.

Los lavados posteriores a la hibridación determinan también las condiciones de rigor. Un conjunto de las condiciones preferidas utiliza una serie de lavados que comienzan con 6XSSC, SDS al 0,5 % a temperatura ambiente durante 15 minutos (min), a continuación se repitió con 2XSSC, SDS al 0,5 % a 45 °C durante 30 minutos, y a continuación se repite dos veces con 0,2 X SSC, SDS al 0,5 % a 50°C durante 30 minutos. Otro ejemplo de condiciones de rigor utiliza temperaturas superiores, en donde los lavados son idénticos a los anteriores excepto por la temperatura de los dos lavados de 30 min finales en 0,2 x SSC, se aumenta el SDS al 0,5 % a 60 °C. Otro ejemplo más de condiciones de elevado rigor utiliza dos lavados finales en 0,1 X SSC, SDS al 0,1 % a 65°C. La hibridación requiere que los dos ácidos nucleicos comprendan secuencias complementarias, aunque dependiendo del rigor de la hibridación, son posibles emparejamientos incorrectos entre las bases.

El rigor adecuado para hibridar ácidos nucleicos depende de la longitud de los ácidos nucleicos y del grado de complementación, variables bien conocidas en la técnica. Cuanto mayor sea el grado de similitud u homología entre dos secuencias de nucleótidos, mayor será el valor de la Tm para híbridos de ácidos nucleicos que tengan aquellas secuencias. La estabilidad relativa (que corresponde a una Tm mayor) de las hibridaciones de ácidos nucleicos disminuye en el siguiente orden: ARN:ARN, ADN:ARN, ADN:ADN. Para híbridos de más de 100 nucleótidos de

longitud, se han derivado las ecuaciones para calcular la T_m (véase Sambrook et al., *supra*, 9,50-0,51). Para la hibridación con ácidos nucleicos más cortos, es decir, oligonucleótidos, la posición de los emparejamientos incorrectos se vuelve más importante, y la longitud del oligonucleótido determina su especificidad (véase Sambrook et al., *supra*, 11,7-11,8).

En una realización, los polinucleótidos se detectan empleando condiciones de hibridación que comprenden una etapa de hibridación en menos de 500 mM de sal y al menos 37 °C, y una etapa de lavado en 2 X SSPE al menos a 63°C. En otra realización, las condiciones de hibridación comprenden menos de 200 mM de sal y al menos 37 °C para la etapa de hibridación. En determinadas realizaciones, las condiciones de hibridación comprenden 2 X SSPE y 63 °C para las etapas de hibridación y lavado.

La longitud de un ácido nucleico hibridable es, por ejemplo, al menos aproximadamente 10 nucleótidos. Una longitud mínima para un ácido nucleico hibridable puede ser al menos aproximadamente 15 nucleótidos; al menos aproximadamente 20 nucleótidos; o al menos 30 nucleótidos. Además, el técnico experto reconocerá que la temperatura y la concentración salina de la solución de lavado se pueden ajustar según sea necesario de acuerdo a factores tales como la longitud de la sonda.

Los fragmentos de ácidos nucleicos sustancialmente similares de la presente invención son aquellos fragmentos de ácidos nucleicos cuyas secuencias de ADN son al menos idénticas en un 70 % a la secuencia de ADN de los fragmentos de ácidos nucleicos notificados en el presente documento. Los fragmentos de ácidos nucleicos de la presente invención incluyen aquellos fragmentos de ácidos nucleicos cuyas secuencias de ADN son al menos idénticas en un 80 %, 90 %, 95 %, 96 %, 97 %, 98 %, y 99 % a la secuencia de ADN de los fragmentos de ácidos nucleicos notificados en el presente documento.

El término "correspondiente a" se usa en el presente documento para referirse a secuencias similares u homólogas, tanto si la posición exacta es idéntica o diferente de la molécula de la cual se mide la similitud o la homología. La alineación de una secuencia de ácido nucleico o aminoácidos puede incluir espacios. Por lo tanto, el término "correspondiente a" se refiere a la similitud de secuencias, y no a la numeración de los restos de aminoácidos o bases de nucleótidos.

Una "porción sustancial" de una secuencia de aminoácidos o nucleótidos comprende suficiente de la secuencia de aminoácidos de un polipéptido o la secuencia de nucleótidos de un gen para identificar presuntamente este polipéptido o gen, tanto mediante evaluación manual de la secuencia por un experto en la materia, como mediante comparación e identificación de secuencias automatizada informáticamente utilizando algoritmos tales como BLAST (Basic Local Alignment Search Tool (Herramienta de Búsqueda de Alineación Local Básica; Altschul y col., J. Mol. Biol. 215:403 410 (1993)); BLAST está públicamente disponible en la World Wide Web. En general, es necesaria a menudo una secuencia de treinta o más aminoácidos o nucleótidos contiguos para identificar presuntamente una secuencia de polipéptidos o ácidos nucleicos como homóloga a una proteína o gen conocido. Además, con respecto a las secuencias de nucleótidos, las sondas de oligonucleótidos específicas de genes que comprende 20 a 30 nucleótidos contiguos se pueden usar en métodos dependientes de secuencias de identificación de genes (por ejemplo, hibridación Southern) y aislamiento (por ejemplo, hibridación in situ de colonias bacterianas o placas de bacteriófagos). Además, se pueden usar oligonucleótidos cortos de 12 a 15 bases como cebadores de la amplificación por PCR a fin de obtener un fragmento de ácido nucleico concreto que comprenda los cebadores. De acuerdo con ello, una "porción sustancial" de una secuencia de nucleótidos comprende suficiente de la secuencia para identificar y/o aislar específicamente un fragmento de ácido nucleico que comprende la secuencia.

El término "porcentaje de similitud", como se conoce en la materia, es una relación entre dos o más secuencias de polipéptidos o dos o más secuencias de polinucleótidos, que se determina comparando las secuencias. En la técnica, "identidad" significa también el grado de correspondencia entre secuencias de polipéptidos o polinucleótidos, según el caso, que se determina por la correspondencia entre cadenas de dichas secuencias. "Identidad" y "similitud" pueden calcularse fácilmente por métodos conocidos, incluyendo, pero no de forma limitativa, los descritos en: Computational Molecular Biology (Lesk, A. M., ed.) Oxford University Press, Nueva York (1988); Biocomputing: Informatics and Genome Projects (Smith, D. W., ed.) Academic Press, Nueva York (1993); Computer Analysis of Sequence Data, Parte I (Griffin, A. M., y Griffin, H. G., eds.) Humana Press, Nueva Jersey (1994); Sequence Analysis in Molecular Biology (von Heinje, G., ed.) Academic Press (1987); y Sequence Analysis Primer (Gribskov, M. y Devereux, J., eds.) Stockton Press, Nueva York (1991). Los métodos para determinar la identidad están diseñados para proporcionar la mejor correspondencia entre las secuencias ensayadas. Los métodos para determinar la identidad y la similitud están públicamente disponibles en programas informáticos. Se pueden llevar a cabo alineaciones de secuencias y cálculos del porcentaje de identidad utilizando el programa Megalign de la suite de cálculo bioinformático (DNASTAR Inc., Madison, WI). Se pueden llevar a cabo múltiples alineaciones de las secuencias utilizando el método de alineación Clustal (Higgins et al., CABIOS. 5:151 153 (1989)) con los parámetros por defecto (PENALIZACIÓN POR HUECO=10, PENALIZACIÓN POR LONGITUD DE HUECO=10). Se pueden seleccionar los parámetros por defecto para las alineaciones por pares utilizando el método Clustal: KTUPLE 1, PENALIZACIÓN POR HUECO=3, VENTANA=5 y DIAGONALES GUARDADAS=5.

El término "software de análisis de secuencias" se refiere a cualquier algoritmo informático o programa de software que

es útil para el análisis de secuencias de nucleótidos o aminoácidos. El "software de análisis de secuencias" puede estar comercialmente disponible o desarrollarse de forma independiente. El software de análisis típico incluirá, pero no de forma limitativa la suite de programas GCG (Wisconsin Package Versión 9.0, Genetics Computer Group GCG), Madison, WI), BLASTP, BLASTN, BLASTX (Altschul et al., J. Mol. Biol. 215:403-410 (1990)), y DNASTAR (DNASTAR, Inc. 1228 S. Park St. Madison, WI 53715 EE.UU.). En el contexto de la presente solicitud se entenderá que cuando se usa el software de análisis de secuencias para el análisis, los resultados del análisis se basarán en los "valores por defecto" del programa al que se hace referencia, salvo que se especifique otra cosa. Como se usa en el presente documento "valores por defecto" significará cualquier conjunto de valores o parámetros que se cargarán originalmente con el software cuando se inicializa en primer lugar.

Como se usa en el presente documento, los términos "expresión" o "expresión génica" se refieren al proceso de convertir la información genética codificada en un gen en ARN (por ejemplo, ARNm, ARNr, ARNt, o ARNsn) mediante la "transcripción" del gen (es decir, mediante la acción enzimática de una ARN polimerasa), y para los genes que codifican proteínas, en una proteína mediante la "traducción" del ARNm. La expresión génica puede regularse en muchas etapas en el proceso. "Regulación en exceso" o "activación" se refiere a la regulación que aumenta la producción de los productos de expresión génica (es decir, ARN o proteína), mientras que "regulación por defecto" o represión se refiere a la regulación que disminuye la producción. Los factores (por ejemplo, factores de transcripción) que están implicados en la regulación en exceso o la regulación por defecto se denominan a menudo "activadores" y "represores", respectivamente. Para los fines de la invención, un gen diana puede regularse por defecto de forma "posterior a la transcripción" (es decir, en el transcrito de ARN) a través de la interacción específica con una molécula de ARN regulada por defecto.

El término "Secuencias control de la transcripción y la traducción" se refiere a secuencias reguladoras del ADN, tales como promotores, potenciadores, terminadores, y similares, que proporcionan la expresión de una secuencia de codificación en una célula hospedadora.

El término "unido de manera operativa" se refiere a la asociación de secuencias de ácidos nucleicos en un único fragmento de ácido nucleico para que la función de uno se vea afectada por el otro. Por ejemplo, un promotor está unido de manera operativa a una secuencia de codificación cuando puede alterar la expresión de esta secuencia de codificación (es decir, que la secuencia de codificación está bajo el control transcripcional del promotor). Las secuencias de codificación pueden unirse de manera operativa a secuencias reguladoras orientadas en sentido directo o en sentido contrario.

Un "vector" se refiere a cualquier vehículo para la clonación y/o la transferencia de un ácido nucleico a una célula hospedadora. Un vector puede ser un replicón al cual se puede unir otro segmento de ADN con el fin de llevar a cabo la replicación del segmento unido. Un "replicón" se refiere a cualquier elemento genético (por ejemplo, un plásmido, fago, cósmido, cromosoma, virus) que funciona como una unidad autónoma de replicación del ADN in vivo, es decir, capaz de replicación bajo su propio control. El término "vector" incluye vehículos víricos y no víricos para introducir el ácido nucleico en una célula hospedadora in vitro, ex vivo o in vivo. El término "vector" puede incluir también ADN minicirculares. Por ejemplo, el vector puede ser un plásmido con secuencias de ADN bacteriano. La eliminación de secuencias de ADN bacteriano que son ricas en regiones CpG ha mostrado disminuir el silenciamiento de la expresión del transgén y da como resultado una expresión más persistente de vectores de ADN plásmido (véase, por ejemplo, Ehrhardt, A. et al. (2003) Hum Gene Ther 10: 215-25; Yet, N. S. (2002) Mol Ther 5: 731-38; Chen, Z. Y. et al. (2004) Gene Ther 11: 856-64). El término "vector" puede incluir también transposones tales como Sleeping Beauty (Izsvak et al. J. Mol. Biol. 302:93-102 (2000)), o cromosomas artificiales.

Se puede usar un gran número de vectores conocidos en la técnica para manipular ácidos nucleicos, incorporar elementos de respuesta y promotores en genes, etc. o transferir un ácido nucleico a una célula hospedadora. Los vectores posibles incluyen, por ejemplo, plásmidos o virus modificados que incluyen, por ejemplo, bacteriófagos tales como derivados lambda, o plásmidos tales como pBR322 o derivados plásmidos de pUC, o el vector Bluescript. Vectores grandes tales como cromosomas artificiales (de bacterias (BAC), de levaduras (YAC), o de seres humanos (HAC)) se pueden usar para acomodar inserciones grandes. Por ejemplo, la inserción de los fragmentos de ADN que corresponden a elementos de respuesta o promotores en un vector adecuado puede llevarse a cabo ligando los fragmentos de ADN adecuados en un vector seleccionado que tiene términos cohesivos complementarios. Como alternativa, los extremos de las moléculas de ADN pueden modificarse enzimáticamente o se pueden producir en cualquier sitio ligando las secuencias de nucleótidos (enlazadores) en los términos de ADN. Dichos vectores pueden genomanipularse para contener genes marcadores seleccionables que proporcionan la selección de células transfectadas o transformadas con el vector. Un vector recombinante que comprende un polinucleótido de acuerdo con la invención puede incluir uno o más orígenes de replicación en los hospedadores celulares, en donde se desea su amplificación o su expresión, los marcadores o marcadores seleccionables.

El término "marcador seleccionable" se refiere a un factor identificante, usualmente un antibiótico o un gen de resistencia química, que es capaz de seleccionarse basándose en el efecto del gen marcador, es decir, resistencia a un antibiótico, resistencia a un herbicida, marcadores colorimétricos, enzimas, marcadores fluorescentes, y similares, en donde el efecto se usa para rastrear la herencia de un ácido nucleico de interés y/o para identificar o seleccionar una célula u organismo que ha heredado el ácido nucleico de interés. Los ejemplos de genes marcadores

seleccionables conocidos y usados en la técnica incluyen: genes que proporcionan resistencia a la ampicilina, estreptomycin, gentamicina, kanamicina, higromicina, herbicida bialaphos, sulfonamida, y similares; y genes que se utilizan como marcadores fenotípicos, es decir, genes reguladores de antocianinas, gen de la isopentaniil transferasa, y similares.

5 El término "gen indicador se refiere a un ácido nucleico que codifica un factor identificante que se puede identificar según el efecto del gen indicador, en el que el efecto se usa para rastrear la herencia de un ácido nucleico de interés, para identificar una célula u organismo que ha heredado el ácido nucleico de interés, y/o para medir la inducción o transcripción de la expresión génica. Los ejemplos de genes indicadores conocidos y usados en la técnica incluyen:

10 luciferasa (Luc), proteínas fluorescentes tales como proteína fluorescente verde (GFP), cloranfenicol acetiltransferasa (CAT), beta-galactosidasa (LacZ), beta-glucuronidasa (Gus), y similares. Los genes marcadores seleccionables pueden considerarse también genes indicadores.

15 El término "plásmido" se refiere a un elemento extracromosómico que transporta a menudo un gen que no forma parte del metabolismo central de la célula, y usualmente en la forma de moléculas de ADN bicatenarias circulares. Dichos elementos pueden ser secuencias que se replican autónomamente, secuencias que integran genomas, secuencias de fagos o nucleótidos, lineales, circulares, o superenrolladas, de un ADN o ARN monocatenario o bicatenario, derivado de cualquier fuente, en el que numerosas secuencias de nucleótidos se han unido o recombinado en una única construcción que es capaz de introducir un fragmento de promotor y una secuencia de ADN de un producto génico

20 seleccionado junto con una secuencia 3' no traducida adecuada en una célula.

Un "vector de clonación" se refiere a un "replicón" que es una unidad de longitud de un ácido nucleico, por ejemplo, ADN, que se replica secuencialmente y que comprende un origen de replicación, tal como un plásmido, fago o cósmido, al cual se puede unir otro segmento de ácido nucleico con el fin de llevar a cabo la replicación del segmento

25 unido. Los vectores de clonación pueden ser capaces de replicación en un tipo de célula y de expresión en otra ("vector lanzadera").

El término "vector de expresión" se refiere a un vector, plásmido o vehículo diseñado para permitir la expresión de una secuencia de ácido nucleico insertada tras la transformación en una célula hospedadora. El gen clonado, es decir, la

30 secuencia de ácido nucleico insertada, se coloca usualmente bajo el control de elementos control tales como un promotor, un promotor mínimo, un potenciador, o similares. Las regiones control del inicio o promotores, que son útiles para impulsar la expresión de un ácido nucleico en la célula hospedadora son numerosas y familiares para los expertos en la materia.

35 Se pueden introducir vectores en las células hospedadoras deseadas por métodos conocidos en la técnica, por ejemplo, transfección, electroporación, microinyección, transducción, fusión celular, DEAE-dextrano, precipitación con fosfato de calcio, lipofección (fusión de lisosomas), uso de una pistola génica, o un vector transportador de ADN (véase, por ejemplo, Wu y col., J. Biol. Chem. 267:963-967 (1992); Wu y col., J. Biol. Chem. 263:14621-14624 (1988); y Hartmutetal., solicitud de patente canadiense n.º 2.012.311, presentada el 15 de marzo de 1990).

40 Los ejemplos de vectores eucariotas incluyen, pero no de forma limitativa, pW-LNEO, pSV2CAT, pOG44, pXT 1 y pSG disponible de Stratagene; pSVK3, pBPV, pMSG y pSVL disponible de Amersham Pharmacia Biotech; y pCMVDsRed2-express, pIRES2-DsRed2, pDsRed2-Mito, pCMV-EGFP disponible de Clontech. Son bien conocidos otros muchos vectores y están comercialmente disponibles.

45 Por ejemplo, se describen vectores útiles, que comprenden pivotes de inserción molecular para la inserción rápida y la eliminación de elementos de programas génicos, en la solicitud de patente publicada de Estados Unidos N.º 2004/0185556, en la solicitud de patente de Estados Unidos N.º 11/233.246 y las solicitudes internacionales publicadas N.ºs WO 2005/040336 y WO 2005/116231.

50 "Promotor" y "secuencia promotora" se usan de manera indistinta y se refieren a una secuencia de ADN capaz de controlar la expresión de una secuencia codificante o ARN funcional. En general, una secuencia codificante se localiza 3' de una secuencia promotora. Los promotores pueden derivarse en su totalidad de un gen nativo, o componerse de diferentes elementos derivados de diferentes promotores que se encuentran en la naturaleza, o comprenden incluso

55 segmentos de ADN sintético. Los expertos en la materia entienden que diferentes promotores pueden dirigir la expresión de un gen en diferentes tejidos o tipos de células, o en diferentes etapas de desarrollo, o en respuesta a diferentes condiciones ambientales o fisiológicas.

Los promotores que hacen que un gen se exprese en la mayoría de tipos celulares la mayoría de las veces se denominan comúnmente como "promotores constitutivos". Los promotores que hacen que un gen se exprese en un tipo celular específico se denominan comúnmente "promotores condicionales". Los ejemplos no limitantes de

60 promotores condicionales son "promotores específicos de células" o "promotores específicos de tejidos". Los promotores que hacen que un gen se exprese en una etapa específica de desarrollo o diferenciación celular se denominan comúnmente "promotores específicos de desarrollo" o "promotores específicos de la diferenciación celular." Los promotores que se inducen y hacen que un gen se exprese tras la exposición o tratamiento de la célula con un agente, molécula biológica, químicas, ligando, luz, o similar que induce el promotor se denominan comúnmente

65

"promotores inducibles" o "promotores regulables". Un ejemplo no limitante de promotor inducible es un promotor inducible TetO. Se reconoce además que, como en la mayoría de los casos no se han definido completamente los límites exactos de las secuencias reguladoras, fragmentos de ADN de diferentes longitudes pueden tener actividad promotora idéntica.

La secuencia promotora se une de forma típica en su extremo 3' por el sitio de inicio de la transcripción y se extiende hacia adelante (dirección 5') para incluir el número mínimo de bases o elementos necesarios para iniciar la transcripción a los niveles detectables por encima del fondo. En la secuencia promotora se encontrará un sitio de inicio de la transcripción (convenientemente definido, por ejemplo, cartografiando con la nucleasa SI), así como dominios de unión a la proteína (secuencias consenso) responsables de la unión de la ARN polimerasa.

Una secuencia codificante está "bajo el control" de secuencias control de la transcripción y la traducción en una célula cuando la ARN polimerasa transcribe la secuencia codificante como ARNm, que posteriormente se somete a corte y empalme como trans-ARN (si la secuencia codificante contiene intrones) y se traduce a la proteína codificada por la secuencia codificante.

Las regiones de control de la transcripción, es decir, secuencias terminadoras o de poliadenilación, también se pueden derivar de diversos genes nativos de los hospedadores preferidos. Opcionalmente, puede ser innecesario un sitio de terminación, sin embargo, puede estar incluido. En una realización de la invención, la región control de la terminación puede estar comprendida o se puede derivar de una secuencia sintética, una señal de poliadenilación sintética, una señal de poliadenilación de un SV40 tardío, una señal de poliadenilación de un SV40, una señal de poliadenilación de la hormona del crecimiento bovino (BHG), secuencias terminadoras víricas, o similares.

El término "transfección" se refiere a la captación de ARN o ADN exógeno o heterólogo por una célula. Una célula se ha "transfectado" por ARN o ADN exógeno o heterólogo cuando dicho ARN o ADN se ha introducido en el interior de la célula. El ARN o ADN transfectado puede integrarse (unirse covalentemente) en el ADN cromosómico preparando el genoma de la célula hospedadora.

"Transformación" se refiere a la transferencia de un fragmento de ácido nucleico en el genoma de un organismo hospedador, dando como resultado una herencia genéticamente estable.

Los términos "modular" y "modula" significan inducir, reducir o inhibir la expresión del ácido nucleico o génica, dando como resultado la correspondiente inducción, reducción o inhibición de la producción de proteína o polipéptido.

"Transcrito de ARN" se refiere al producto resultante de la transcripción catalizada por la ARN polimerasa de una secuencia de ADN. Cuando el transcrito de ARN es una copia complementaria perfecta de la secuencia de ADN, se denomina transcrito primario o puede ser una secuencia de ARN derivada del procesamiento posterior a la transcripción del transcrito primario y se denomina ARN maduro. "ARN mensajero" (ARNm) se refiere al ARN que está sin intrones y que se puede traducir en proteínas por la célula. "ADNc" se refiere a un ADN bicatenario que es complementario y derivado del ARNm. ARN de "sentido directo" se refiere a un transcrito de ARN que incluye el ARNm y de esta manera se puede traducir en proteínas por la célula.

La invención proporciona un vector de expresión que comprende una construcción de un gen sintético que comprende, cuando se dispone desde 5' a 3', un promotor, un polinucleótido quimérico, y una secuencia de interés que se va a expresar, en el que:

a. el polinucleótido quimérico comprende el segundo intrón de un gen de la calcio ATPasa del retículo sarcoplásmico/endoplasmático, y al menos una porción de la región 5' no traducida (5'UTR) de un gen de la caseína; y

b. el polinucleótido quimérico está situado entre el promotor y la secuencia de interés que se va a expresar, en el que el segundo intrón del gen de la calcio ATPasa del retículo sarcoplásmico/endoplasmático es situado hacia el promotor y el fragmento de polinucleótido de la región 5' no traducida (5'UTR) del gen de la caseína está situado hacia la secuencia de interés que se va a expresar.

Una realización de la divulgación es una 5'UTR sintética que es un polinucleótido quimérico que comprende un primer fragmento de polinucleótido y un segundo fragmento de polinucleótido, en el que:

a. el primer fragmento de polinucleótido comprende el segundo intrón de un gen de la ATPasa de calcio del retículo sarcoplásmico/endoplasmático;

b. el segundo fragmento de polinucleótido comprende al menos una porción de la región 5' no traducida (5'UTR) de un gen de la caseína; y

c. el primer fragmento de polinucleótido se localiza en 5' del segundo fragmento de polinucleótido.

El fragmento de polinucleótido que comprende el segundo intrón de un gen de la calcio ATPasa del retículo

sarcoplásmico/endoplasmático puede derivarse de cualquier gen de la calcio ATPasa del retículo sarcoplásmico/endoplasmático. En una realización de la invención, el fragmento de polinucleótido que comprende el segundo intrón de un gen de la calcio ATPasa del retículo sarcoplásmico/endoplasmático se deriva de un gen SERCA2. En otras realizaciones, se deriva de un gen SERCA1 o SERCA3. El gen de la calcio ATPasa del retículo sarcoplásmico/endoplasmático que es la fuente del fragmento de polinucleótido que comprende el segundo intrón, puede derivar de cualquier especie eucariota. En una realización, el gen de la calcio ATPasa del retículo sarcoplásmico/endoplasmático es de una especie de mamífero. En otra realización, el gen de la calcio ATPasa del retículo sarcoplásmico/endoplasmático es de una especie de ave. En otra realización, el gen de la calcio ATPasa del retículo sarcoplásmico/endoplasmático es de una especie de piscifactoría. En realizaciones específicas, el fragmento de polinucleótido que comprende el segundo intrón se deriva del gen de la calcio ATPasa del retículo sarcoplásmico/endoplasmático de un ser humano, un perro, o un ratón. En otras realizaciones específicas, el fragmento de polinucleótido que comprende el segundo intrón se deriva de un gen de la calcio ATPasa del retículo sarcoplásmico/endoplasmático de una rata, un chimpancé, una gallina, un caballo, una vaca, un alce, un cerdo, un gato, un macaco rhesus, o un pez cebra.

En otra realización de la invención, el fragmento de polinucleótido que comprende el segundo intrón de un gen de la calcio ATPasa del retículo sarcoplásmico/endoplasmático comprende una porción del exón 2 flanqueante en el extremo 5' y una porción del exón 3 flanqueante en el extremo 3'. En otra realización, el fragmento de polinucleótido que comprende el segundo intrón de un gen de la calcio ATPasa del retículo sarcoplásmico/endoplasmático la totalidad del exón 2 flanqueante en el extremo 5' y una porción del exón 3 flanqueante en el extremo 3'. El fragmento de polinucleótido que comprende el segundo intrón de un gen de la calcio ATPasa del retículo sarcoplásmico/endoplasmático puede tener al menos 50 nucleótidos de longitud. En otras realizaciones, el fragmento de polinucleótido que comprende el segundo intrón de un gen de la calcio ATPasa del retículo sarcoplásmico/endoplasmático puede ser al menos de aproximadamente 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, o 250 nucleótidos de longitud.

En otra realización de la invención, el fragmento de polinucleótido que comprende el segundo intrón de un gen de la calcio ATPasa del retículo sarcoplásmico/endoplasmático está mutado en un presunto sitio poli A de consenso. En otra realización, el fragmento de polinucleótido que comprende el segundo intrón de un gen de la calcio ATPasa del retículo sarcoplásmico/endoplasmático que comprende una porción flanqueante 5' del exón 2 y una porción flanqueante 3' del exón 3, está mutado en el presunto sitio poli A de consenso, y se deriva de un gen SERCA2 de canino; en una realización específica se representa por la SEQ ID NO: 2.

En otras realizaciones, el fragmento de polinucleótido que comprende el segundo intrón de un gen de la calcio ATPasa del retículo sarcoplásmico/endoplasmático es una secuencia SERCA2 natural o mutada parcialmente. El fragmento de polinucleótido puede derivarse de cualquier gen SERCA2 de longitud completa de cualquier especie. Por ejemplo, en una realización, el fragmento de polinucleótido que comprende el segundo intrón de un gen de la calcio ATPasa del retículo sarcoplásmico/endoplasmático es una secuencia genómica SERCA2 de canino parcialmente natural que comprende una porción del exón 2 flanqueante en el extremo 5' y una porción del exón 3 flanqueante en el extremo 3'; en una realización específica se representa por la SEQ ID NO: 4. La SEQ ID NO: 4 se representa esquemáticamente en la Figura 1A y en la Figura 1C. En otra realización, el fragmento de polinucleótido que comprende el segundo intrón de un gen de la calcio ATPasa del retículo sarcoplásmico/endoplasmático es una secuencia genómica SERCA2 humana parcialmente natural que comprende un exón 2 flanqueante en el extremo 5' y un exón 3 flanqueante en el extremo 3'; en una realización específica, se representa por la SEQ ID NO: 5. En otra realización, el fragmento de polinucleótido que comprende el segundo intrón de un gen de la calcio ATPasa del retículo sarcoplásmico/endoplasmático es una secuencia genómica SERCA2 de murino parcialmente natural que comprende el exón 2 flanqueante en el extremo 5' y el exón 3 flanqueante en el extremo 3'; en una realización específica se representa por la SEQ ID NO: 6. Las SEQ ID NO: 5 y SEQ ID NO: 6 se representan esquemáticamente en la Figura 1B y Figura 1C. En otras realizaciones, el fragmento de polinucleótido que comprende el segundo intrón de un gen de la calcio ATPasa del retículo sarcoplásmico/endoplasmático es una secuencia SERCA2 natural o mutada parcialmente de *Rattus norvegicus*, una secuencia SERCA2 de *Equus caballus*, una secuencia SERCA2 de *Bos taurus*, una secuencia SERCA2 de *Pan troglodytes*, una secuencia SERCA2 de *Felis catus*, una secuencia SERCA2 de *Ortolagus cuniculus*, una secuencia SERCA2 de *Sus scrofa*, una secuencia SERCA2 de *Macaca mulatta*, una secuencia SERCA2 de *Cervus elaphus*, una secuencia SERCA2 de *Gallus*, una secuencia SERCA2 de *Danio rerio*.

El fragmento de polinucleótido que comprende al menos una porción de una región 5' no traducida de un gen de la caseína puede ser de cualquier especie de mamífero. En una realización, el fragmento de polinucleótido que comprende al menos una porción de la región 5' no traducida es de un gen de la beta caseína de bovino; en una realización específica se representa por la SEQ ID NO: 3. En otra realización, el fragmento de polinucleótido que comprende al menos una porción de una región 5' no traducida es de un gen de la beta caseína de ratón; en una realización específica se representa por la SEQ ID NO: 8. En otra realización, el fragmento de polinucleótido que comprende al menos una porción de una región 5' no traducida es de un gen de la beta caseína de rata; en una realización específica se representa por la SEQ ID NO: 9. En otra realización, el fragmento de polinucleótido que comprende al menos una porción de la región 5' no traducida es de un gen de la beta caseína de oveja; en una realización específica se representa por la SEQ ID NO: 10. En otras realizaciones, el fragmento de polinucleótido que comprende al menos una región 5' no traducida es de un gen de la beta caseína de *Bubalus bubalis*, un gen de la beta

caseína de *Capra hircus*, un gen de la beta caseína de *Equus caballus*, un gen de la beta caseína de *Sus scrofa*, un gen de la beta caseína de *Camelus dromedaries*, un gen de la beta caseína de *Oryctolagus cuniculus*, o un gen de la beta caseína de *Canis lupus*.

5 El fragmento de polinucleótido que comprende al menos una porción de una región 5' no traducida del gen de la caseína puede tener al menos aproximadamente 25 nucleótidos de longitud. En otras realizaciones, el fragmento de polinucleótido del gen de la caseína que comprende al menos una porción de la 5'UTR puede tener al menos aproximadamente 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 85, 90, 100 o más nucleótidos de longitud. En otra realización, el fragmento de polinucleótido que comprende al menos una porción de 5'UTR de un gen de la caseína puede representar al menos aproximadamente un 50 % de la secuencia 5'UTR natural. En otras realizaciones, el fragmento de polinucleótido que comprende al menos una porción de 5'UTR de un gen de la caseína puede representar al menos aproximadamente un 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 % o más de la secuencia 5'UTR natural. En otra realización, el fragmento de polinucleótido que comprende al menos una porción de 5'UTR de un gen de la caseína puede representar la secuencia 5'UTR completa natural.

15 En otras realizaciones, las variantes funcionales de los componentes individuales (el fragmento de polinucleótido que comprende el segundo intrón de un gen de la calcio ATPasa del retículo sarcoplásmico/endoplasmático y el fragmento de polinucleótido que comprende al menos una porción de 5'UTR de un gen de la caseína), se usan para crear una 5'UTR sintética. Las variantes funcionales abarcan variantes de sustitución, inserción y delección y sus combinaciones. Las variantes de sustitución son aquellas, en donde se ha eliminado al menos una base en la secuencia de nucleótidos y se ha insertado una base diferente en su lugar. Las variantes de inserción de un ácido nucleico son aquellas, en donde se introducen uno o más nucleótidos en un sitio predeterminado de la secuencia. Las variantes de delección de un ácido nucleico se caracterizan por la eliminación de uno o más nucleótidos del ácido nucleico. Se puede producir cualquier combinación de sustitución(ones), delección(ones) o inserción(ones) con la condición de que la funcionalidad del componente siga siendo esencialmente la misma, es decir, que la variante funcional, cuando se usa en una 5'UTR sintética de la presente invención, dé lugar a una expresión aumentada de una secuencia de interés, un gen sintético, o transgén.

30 Adicionalmente, las secuencias homólogas a las realizaciones específicas del fragmento de polinucleótido que comprende el segundo intrón de un gen de la calcio ATPasa del retículo sarcoplásmico/endoplasmático divulgadas en el presente documento (SEQ ID NO:2, SEQ ID NO: 4, SEQ ID NO: 5, y SEQ ID NO:6) y las secuencias homólogas a las realizaciones específicas del fragmento de polinucleótido que comprende al menos una porción de la 5'UTR de caseína (SEQ ID NO:3, SEQ ID NO: 8, SEQ ID NO: 9, y la SEQ ID NO: 10) se pueden usar para construir una 5'UTR sintética. Como se ha mencionado anteriormente, las fuentes adecuadas de un fragmento para crear una 5'UTR sintética incluyen un gen de la calcio ATPasa del retículo sarcoplásmico/endoplasmático de cualquier especie eucariota y un gen de la caseína de cualquier especie de mamífero. En una realización, el fragmento de polinucleótido que comprende el segundo intrón de un gen de la calcio ATPasa del retículo sarcoplásmico/endoplasmático se deriva de un ortólogo de SERCA2 de canino, SERCA2 de ratón, o SERCA2 de ser humano. En otra realización, el fragmento de polinucleótido que comprende al menos una porción de la 5'UTR de la caseína se deriva de un ortólogo de la beta caseína de bovino, beta caseína de ratón, beta caseína de rata, o beta caseína de oveja.

45 Las personas expertas en la materia conocerán métodos para investigar e identificar los homólogos de la calcio ATPasa de retículo sarcoplásmico/endoplasmático o de los homólogos de 5'UTR de la caseína. Dichos métodos comprenden la comparación de las secuencias representadas por las SEQ ID NO: 2-6 y 8-10, en un formato informáticamente legible, con secuencias que están disponibles en bases de datos públicas en la World Wide Web tales como M IPS, GenBank, o base de datos de secuencias de nucleótidos del EMBL, utilizando algoritmos bien conocidos en la técnica para la alineación o comparación de secuencias, tales como GAP (Needleman y Wunsch, J. Mol. Biol. 48; 443453 (1970)), BESTFIT (Miller, W., Myers, E. W. y Lipman, D. J., J. Mol. Biol. 215:403-410 (1990)), FASTA y TFASTA (W. R. Pearson y D. J. Lipman Proc. Natl. Acad. Sci. USA 85:2444-2448 (1988)). El software para llevar a cabo los análisis BLAST está públicamente disponible del National Centre for Biotechnology Information. Se pueden identificar los homólogos adecuados utilizando los parámetros por defecto de BLAST (matriz BLOSUN62, penalización por apertura de hueco de 11 y penalización por extensión de hueco de 1).

55 Adicionalmente, los homólogos de SERCA 2 de canino, ser humano, o ratón pueden identificarse también investigando sobre secuencias conservadas en el gen SERCA2. Por ejemplo, la secuencia completa del exón 3 de SERCA2 de canino tal como la SEQ ID NO: 11 se puede usar como secuencia de consulta en la investigación mediante BLAST. Se anticipa que utilizando la secuencia del exón como secuencia de consulta en la investigación mediante BLAST se recuperaran mayores cantidades de homólogos de SERCA2 que utilizando una secuencia que comprende la secuencia del intrón. De manera similar, se pueden identificar homólogos de beta caseína de bovino, ratón, rata, u oveja utilizando una porción codificante tal como la secuencia de consulta.

65 Es también posible el análisis de secuencias genómicas para identificar homólogos de la calcio ATPasa de retículo sarcoplásmico/endoplasmático u homólogos de 5'UTR de la caseína. Están disponibles algunos algoritmos y herramientas de software para identificar genes en la secuencia de ADN bruto. Usualmente estas herramientas combinan el análisis de parámetros estadísticos en la secuencia de ADN con métodos basados en la homología para identificar secuencias homólogas en bases de datos. Aunque ninguno de estos métodos en solitario es

suficientemente fiable para una buena predicción, la combinación de diversos programas proporciona usualmente buenos resultados. Ejemplos bien conocidos de dichas herramientas que están públicamente disponibles en la World Wide Web incluyen GeneMark (Borodovsky, M. y McIninch J. GeneMark: Parallel Gene Recognition for both DNA Strands. Computers & Chemistry, 17, 123-133 (1993)), Gene Locator and Interpolated Markov Modeler (GLIMMER) (A. L. Delcher et al. Improved microbial gene identification with GLIMMER. NucleicAcids Research, 27, 4636-4641. (1999)), Gene Recognition and Assembly Internet Link (GRAIL), GenScan (Burge, C. y Karlin, S. Prediction of complete gene structures in human genomic DNA. J. Mol. Biol. 268, 78-94 (1997)), y GeneBuilder (Milanesi L. et al. GeneBuilder: interactive in silico prediction of genes structure. Bioinformatics, 15 (7):612-621 (1999)). Se puede llevar a cabo un análisis combinado con el programa TIGR Combiner (J. E. Allen et al. Computational gene prediction using multiple sources of evidence. Genome Research, 14(1), 142-148 (2004)) que predice modelos de genes utilizando el resultado de otro software de anotación tal como GeneMark, GlimmerM, GRAIL, GenScan, y Fgenes. Este usa un algoritmo estadístico para identificar modelos de evidencia que corresponden a modelos de genes.

Se puede identificar el segundo intrón de un gen de la calcio ATPasa de retículo sarcoplásmico/endoplasmático mediante métodos rutinarios tales como comparar la secuencia de ADN genómico del gen con la de su secuencia de ARNm o ADNc en un programa de alineación por pares. Las regiones de homología representan exones, mientras que las secuencias intervinientes que están ausentes en la secuencia de ADNc pero presentes en el ADN genómico representan intrones. Puede identificarse el comienzo y el final de la secuencia del intrón mediante su 5' GT flanqueante y su 3' AG flanqueante. Utilizando esta solución, la secuencia de ARNm SERCA2 de canino representada por el número de acceso público NM_001003214 es útil para identificar intrones en una secuencia genómica SERCA2 de canino, mientras que se pueden usar las secuencias de ARNm SERCA2 de ser humano y ratón (NM_170665 y NM_009722, respectivamente) para identificar intrones en sus respectivas secuencias genómicas.

Se pueden identificar también homólogos de la calcio ATPasa de retículo sarcoplásmico/endoplasmático u homólogos de 5'UTR de la caseína sondeando una biblioteca de fragmentos genómicos o de ADNc de otras especies. Por ejemplo, el ADN genómico de una especie de interés puede estar fragmentado en fragmentos adecuadamente dimensionados para su inserción en un vector seleccionado tal como un plásmido o vector lambda. El vector se digiere a continuación con una enzima de restricción adecuada y a continuación se liga con la mezcla completa de fragmentos genómicos. Las células bacterianas se transforman con el vector y a continuación se siembran en placas sobre placas de agarosa. A continuación se unen las colonias o la placa de ADN del fago a una membrana. En una realización, los fragmentos representados por las SEQ ID NO: 2-6 y 8-10 o sus porciones se utilizaron como sondas marcadas para la hibridación con el ADN de un clon de la biblioteca que contiene una secuencia homóloga. Se pueden usar procedimientos similares para cribar una biblioteca de ADNc. Adicionalmente, se pueden identificar homólogos utilizando los fragmentos representados por las SEQ ID NO: 2-6 y 8-10 o sus porciones en un experimento de hibridación genómica o un experimento de hibridación Southern del ADNc. En otras realizaciones, las secuencias más conservadas tales como las que comprenden una región codificante de un gen de una calcio ATPasa del retículo sarcoplásmico/endoplasmático (tal como la SEQ ID NO: 11) o un gen de la caseína o una de sus porciones se usan como sondas en un experimento de hibridación Southern o para el cribado de una biblioteca.

Se pueden seleccionar condiciones de hibridación adecuadas para permitir la hibridación de un fragmento con una sonda de otra especie (híbridos sonda-diana emparejados de forma parcialmente incorrecta) reduciendo el rigor del experimento de hibridación mediante una combinación adecuada de temperatura, concentración de sal, o % de formamida. Por ejemplo, puede disminuirse el rigor del experimento de hibridación reduciendo la temperatura o aumentando la concentración de sal. Son bien conocidos en la técnica los procedimientos para identificar condiciones de hibridación adecuadas y se describen en Sambrook (2001) Molecular Cloning: a laboratory manual, 3ª Edición Cold Spring Harbor Laboratory Press, CSH, Nueva York.

Otra realización de la divulgación es un polinucleótido al menos 80 %, 85 %, 90 %, 95 %, 96 %, 97 %, 98 % o 99 % idéntico a un polinucleótido representado por cualquiera de las SEQ ID NO 1 -10.

Otra realización de la divulgación es una 5'UTR sintética que comprende un polinucleótido representado por una de las SEQ ID NO: 2 y 4-6 y un polinucleótido representado por una de las SEQ ID NO: 3 y 8-10.

Otra realización de la divulgación es una 5'UTR sintética que comprende un polinucleótido que es al menos un 80 %, 85 %, 90 %, 95 %, 96 %, 97 %, 98 % o 99 % idéntico a un polinucleótido representado por una de las SEQ ID NO:2 y 4-6 y un polinucleótido que es al menos un 80 %, 85 %, 90 %, 95 %, 96 %, 97 %, 98 % o 99 % idéntico a un polinucleótido representado por una de las SEQ ID NO:3, y 8-10.

Otra realización de la invención es una construcción de gen sintético que comprende un polinucleótido que es al menos un 80 %, 85 %, 90 %, 95 %, 96 %, 97 %, 98 % o 99 % idéntico a un polinucleótido representado por la SEQ ID NO: 1.

Otra realización de la invención es una construcción de gen sintético que comprende un polinucleótido que es al menos un 80 %, 85 %, 90 %, 95 %, 96 %, 97 %, 98 % o 99 % idéntico a un polinucleótido representado por la SEQ ID NO: 7.

En otra realización, la secuencia 5'UTR sintética carece de sitios de restricción que interferirían con la inserción en el Sistema de Producción UltraVector (Intrexon Corp., Blacksburg, VA) como se describe en el documento WO 2007/038276. En una realización específica, la secuencia 5'UTR sintética carece de secuencias de reconocimiento internas para las siguientes endonucleasas de restricción: AsiS I, Pac I, Sbf I, Fse I, Asc I, Mlu I, SnaB I, Not I, Sal I, Swa I, Rsr II, BsiW I, Mfe I, Nhe I, Nsi I, Cla I, Nde I, Nsi I, Kpn I, Nco I y Pst I.

La secuencia 5'UTR sintética incluye opcionalmente sitios de restricción en el extremo 5' y 3' para facilitar la clonación en un vector. En una realización específica, la secuencia 5'UTR sintética incluye secuencias de reconocimiento para Mlu I en el extremo 5' y secuencias de reconocimiento para Mfe I en el extremo 3'.

En una realización específica, la 5'UTR sintética se representa por la SEQ ID NO: 1. La SEQ ID NO: 1 comprende las siguientes características: un sitio de restricción MluI, SEQ ID NO: 2, un sitio de restricción KpnI, SEQ ID NO: 3, un sitio de restricción MfeI. La SEQ ID NO: 1 se representa también esquemáticamente por la Figura 2k.

En otra realización específica, la 5'UTR sintética se representa por la SEQ ID NO: 7. La SEQ ID NO: 7 comprende las siguientes características: un sitio de restricción Ascl, un sitio de restricción MluI, SEQ ID NO: 4, un sitio de restricción KpnI, SEQ ID NO: 3, un sitio de restricción MfeI. La SEQ ID NO: 7 se representa esquemáticamente por la Figura 2B.

En una realización, la secuencia 5'UTR sintética tiene menos de aproximadamente 500 nucleótidos de longitud. En otra realización, la secuencia 5'UTR sintética tiene menos de aproximadamente 400 nucleótidos de longitud. En otra realización, la secuencia 5'UTR sintética tiene menos de aproximadamente 350 nucleótidos de longitud. En otra realización, la secuencia 5'UTR sintética tiene menos de aproximadamente 300 nucleótidos de longitud. En otra realización, la secuencia 5'UTR sintética tiene menos de aproximadamente 240 nucleótidos de longitud. En otra realización, la secuencia 5'UTR sintética tiene menos de aproximadamente 200 nucleótidos de longitud.

En otra realización de la divulgación, el polinucleótido 5'UTR sintético es un componente de un vector de expresión eucariota, que comprende un vector de expresión que comprende una construcción de un gen sintético que comprende, cuando se dispone desde 5' a 3', un promotor, un polinucleótido quimérico, y una secuencia de interés que se va a expresar, en el que:

a. el polinucleótido quimérico comprende el segundo intrón de un gen de la calcio ATPasa del retículo sarcoplásmico/endoplasmático, y al menos una porción de la región 5' no traducida (5'UTR) de un gen de la caseína; y

b. el polinucleótido quimérico es situado entre el promotor y la secuencia de interés que se va a expresar, en el que el segundo intrón del gen de la calcio ATPasa del retículo sarcoplásmico/endoplasmático es situado hacia el promotor y el fragmento de polinucleótido de la región 5' no traducida (5'UTR) del gen de la caseína es situado hacia la secuencia de interés que se va a expresar.

La presente invención proporciona también vectores que comprenden una 5' UTR sintética descrita en el presente documento. Se contempla que los vectores incluyan cualquiera de las realizaciones de la secuencia del polinucleótido 5'UTR sintético descrita en el presente documento. Por ejemplo, una realización de la invención es un vector que comprende una secuencia del polinucleótido 5'UTR sintética que comprende un fragmento de polinucleótido que comprende el segundo intrón del gen de la calcio ATPasa del retículo sarcoplásmico/endoplasmático y un fragmento de polinucleótido que comprende al menos una porción de una 5'UTR de un gen de la caseína.

Otra realización de la invención es un vector que comprende un polinucleótido que es al menos un 80 %, 85 %, 90 %, 95 %, 96 %, 97 %, 98 % o 99 % idéntico a un polinucleótido representado por una de las SEQ ID NO:1-10.

En otra realización, el vector es un vector de expresión que comprende una construcción de un gen sintético que comprende una 5'UTR sintética. La construcción del gen sintético puede comprender un promotor flanqueando un extremo de la 5'UTR sintética y una secuencia de interés que se va a expresar flanqueando el otro extremo de la 5'UTR sintética. La construcción del gen sintético puede comprender además un sitio de poliadenilación.

Por ejemplo, otra realización de la invención proporciona un vector de expresión que comprende una construcción de un gen sintético que comprende, cuando se dispone desde 5' a 3', un promotor, un polinucleótido quimérico, y una secuencia de interés que se va a expresar, en el que:

a. el polinucleótido quimérico comprende el segundo intrón de un gen de la calcio ATPasa del retículo sarcoplásmico/endoplasmático y al menos una porción de la región 5' no traducida de un gen de la caseína;

b. el polinucleótido quimérico es situado entre el promotor y la secuencia de interés que se va a expresar, en el que el segundo intrón del gen de la calcio ATPasa del retículo sarcoplásmico/endoplasmático es situado hacia el promotor y el fragmento de polinucleótido de la región 5' no traducida del gen de la caseína es situado hacia la secuencia de interés que se va a expresar.

La Figura 3A representa esquemáticamente una realización de una construcción del gen sintético de la invención insertado en una estructura de vector. En esta realización, SPL en la Figura 3A se refiere al fragmento de polinucleótido que comprende al menos un sitio de corte y empalme y la UTR se refiere al fragmento del polinucleótido que comprende al menos una porción de una región 5' no traducida. SPL y UTR juntos constituyen una 5'UTR sintética. El promotor de la construcción del gen sintético es situado para dirigir la expresión del ARN de la 5'UTR sintética y la secuencia de interés. Adicionalmente, la secuencia de interés que se va a expresar comprende un codón de inicio para que comience la traducción.

Los vectores de expresión ilustrativos que comprenden esta arquitectura de vector se representan esquemáticamente en la Figura 10 y en la Figura 11. La secuencia del vector de la Figura 10 se proporciona en la SEQ ID NO: 12; la secuencia del vector de la Figura 11 se proporciona en la SEQ ID NO: 13.

La divulgación describe también un vector de expresión que comprende, cuando se dispone desde 5' a 3', un promotor, un polinucleótido quimérico, y un sitio de clonación, en el que:

a. el polinucleótido quimérico comprende un fragmento de polinucleótido de un primer gen eucariota que comprende al menos un sitio de corte y empalme y un fragmento de polinucleótido de un segundo gen eucariota que comprende al menos una porción de la región 5' no traducida;

b. el polinucleótido quimérico es situado entre el promotor y el sitio de clonación, en el que el fragmento de polinucleótido del primer gen eucariota es situado hacia el promotor y el fragmento de polinucleótido del segundo gen eucariota es situado hacia el sitio de clonación.

El sitio de clonación del vector de expresión puede comprender uno o más sitios de restricción únicos con el fin de que se pueda insertar la secuencia de interés. En otra realización, el sitio de clonación comprende sitios de unión a la recombinasa con el fin de que se pueda insertar la secuencia de interés mediante recombinación específica de sitio.

Una realización del vector de expresión se representa esquemáticamente en la Figura 3B. En la Figura 3B, SPL se refiere al fragmento de polinucleótido que comprende el sitio de corte y empalme y la UTR se refiere al fragmento del polinucleótido que comprende al menos una porción de una región 5' no traducida. SPL y UTR juntos constituyen una 5'UTR sintética.

El fragmento de polinucleótido que comprende el sitio de corte y empalme en el vector de expresión puede ser un fragmento de cualquier gen eucariota. Los fragmentos de polinucleótidos ilustrativos que se pueden usar en el vector de expresión incluyen los que comprenden un sitio de corte y empalme de un gen de la calcio ATPasa del retículo sarcoplásmico/endoplasmático eucariota descrito en el presente documento. Adicionalmente, el fragmento de polinucleótido que comprende al menos una porción de una región 5' no traducida de un gen puede ser de cualquier gen eucariota. Los fragmentos de polinucleótidos ilustrativos que se pueden usar en el vector de expresión incluyen los que comprenden al menos una porción de la 5'UTR de un gen de la caseína descrito en el presente documento.

Los vectores de expresión de la invención pueden comprender además una o más secuencias de polinucleótidos adicionales en la dirección 3' de la secuencia de interés o el sitio de clonación para crear una fusión en marco con el polipéptido codificado por la secuencia de interés. Por ejemplo, los polinucleótidos adicionales en la dirección 3' de la secuencia de interés pueden codificar un epítipo tag, un indicador, o una etiqueta de purificación. Se conocen en la técnica los epítipos tag e incluyen myc, hemaglutinina (HA), y FLAG. Los ejemplos de indicadores incluyen la proteína fluorescente verde y sus variantes, beta-galactosidasa (LacZ), beta-glucuronidasa (Gus) cloranfenicol acetiltransferasa (CAT) y luciferasa. Los ejemplos de etiquetas de purificación incluyen His₆ y GST. El vector de expresión puede comprender también un sitio poliA en la dirección 3' de la secuencia de interés o el sitio de clonación.

Dependiendo del resultado deseado, la porción del promotor del vector de expresión que contiene una 5'UTR sintética puede ser un promotor constitutivo, un promotor no constitutivo, un promotor específico de tejido (constitutivo o no constitutivo), una patogénesis o enfermedad relacionada con el promotor, un promotor específico del desarrollo, o un promotor controlado selectivamente tal como un promotor inducible. Diferentes promotores controlados selectivamente están controlados por mecanismos diferentes. Por ejemplo, un promotor inducible por tetraciclina se activa para expresar una secuencia de codificación en la dirección 3' cuando la célula que contiene el promotor y otros factores celulares necesarios se trata con tetraciclina. Otros promotores inducibles se activan por otros fármacos o factores. RHEOSWITCH es un sistema promotor inducible disponible de New England Biolabs (Ipswich, MA). Pueden utilizarse también promotores sensibles a la temperatura para aumentar o disminuir la expresión génica. Una realización de la invención comprende una construcción génica que contiene una 5'UTR sintética cuya expresión está controlada por un promotor inducible.

La invención incluye realizaciones, en donde la estructura del vector que comprende la 5'UTR sintética comprende secuencias adecuadas para la expresión de una secuencia de interés en una célula eucariota. En una realización, la estructura del vector que comprende la 5'UTR sintética comprende secuencias adecuadas para la expresión de una secuencia de interés en una célula de mamífero. Los vectores de expresión de mamíferos pueden comprender elementos no transcritos tales como un origen de replicación, un promotor y un potenciador adecuados unidos a la

secuencia de interés que se va a expresar, y otras secuencias 5' o 3' flanqueantes no transcritas, y secuencias 5' o 3' no traducidas, tales como los sitios de unión a ribosoma necesarios, un sitio de poliadenilación, y secuencias de terminación de la transcripción. Son bien conocidos en la técnica los ejemplos de vectores de expresión de mamíferos e incluyen pcDNA3 (Invitrogen) y pRSVneo (ATTC).

Por ejemplo, la porción promotora de un vector de expresión que contiene la 5'UTR sintética puede ser un promotor animal o de mamífero. Los promotores animales o de mamíferos incluyen la región promotora temprana de SV40 (SV40e), el promotor contenido en la repetición terminal larga 3' (LTR) del virus del sarcoma de Rous (VSR), los promotores de los genes EIA o del promotor tardío mayor (MLP) de adenovirus (Ad), el promotor temprano del citomegalovirus (CMV), el promotor de la timidina quinasa (TK) del virus del herpes simple (VHS), un promotor del factor 1 alfa de elongación (EF1), un promotor de la fosfoglicerato quinasa (PGK), un promotor de la ubiquitina (Ubc), un promotor de la albúmina, las secuencias reguladoras del promotor de la metalotioneína-L de ratón y regiones control de la transcripción, los promotores ubicuos (HPRT, vimentina, beta-actina, tubulina y similares), los promotores de los filamentos intermedio (desmina, neurofilamentos, queratina, GFAP, y similares), los promotores de genes terapéutico (del MDR, CFTR o el factor de tipo VII, y similares), promotores de la patogénesis o enfermedades relacionadas, y promotores que presentan especificidad tisular y que se han utilizado en animales transgénicos, tales como la región control del gen de la elastasa I que es activa en las células acinares pancreáticas; la región control del gen de la insulina activa en células beta pancreáticas, la región control del gen de la inmunoglobulina activa en células linfoides, la región control del virus del tumor de mama de ratón activa en células de testículo, mama, linfoides y mastocitos; el gen de la albúmina, las regiones control de ApoI y ApoAII activas en hígado, la región control del gen de la alfa-fetoproteína activa en hígado, la región control del gen de la alfa 1-antitripsina activa en el hígado, la región control del gen de la beta-globina activa en células mieloides, la región control de la proteína básica mielina activa en células de oligodendrocitos en el cerebro, la región control del gen de la cadena ligera 2 de miosina activa en el músculo esquelético, y la región control del gen de la hormona de liberación gonadotrófica activa en el hipotálamo, el promotor de la piruvato quinasa, el promotor de la vilina, el promotor de la proteína intestinal de unión a ácido graso, el promotor de la beta-actina de células del músculo liso, y similares. Los promotores del vector de expresión pueden modificarse mediante la adición de un potenciador o secuencias reguladoras y similares.

La secuencia de la porción de interés de un vector de expresión que contiene una 5'UTR sintética puede ser una secuencia de cualquier gen eucariota o una de sus porciones, o la secuencia de codificación o no de codificación natural o no natural. Los ejemplos no limitantes de secuencias de codificación que se pueden usar en la presente invención incluyen secuencias que codifican indicadores (por ejemplo, luciferasa, beta-galactosidasa, proteínas fluorescentes), epitopos, polipéptidos experimentales, o polipéptidos terapéuticos. Los ejemplos adicionales de secuencias de ácidos nucleicos de interés incluyen moléculas de ARN, tales como ARN pequeños, micro ARN, ARN ribosómicos, ARN terapéuticos, y ribozimas.

En otro aspecto de la divulgación se usan 5'UTR sintéticas múltiples no redundantes, en el contexto de construcciones génicas multigénicas dentro de un único vector.

En otra realización, el vector es un vector de tratamiento génico que comprende una construcción de un gen sintético que comprende una 5'UTR sintética. El vector de tratamiento génico puede ser cualquier vector de tratamiento génico conocido en la técnica, incluyendo vectores no víricos o vectores víricos tales como un vector de adenovirus, un vector vírico adenoasociado (VAA), o un vector retrovírico. La construcción del gen sintético puede comprender un promotor flanqueando el extremo 5' de la 5'UTR sintética y un gen terapéutico de interés flanqueando el extremo 3' de la 5'UTR sintética como se muestra en la Figura 3A. El promotor puede ser un promotor constitutivo, un promotor específico de tejido, y un promotor inducible, u otro promotor descrito en el presente documento. Los ejemplos de clases de genes terapéuticos de interés que se pueden incluir en el vector de tratamiento génico incluyen sin limitación, genes que codifican citoquinas, quimioquinas, hormonas, anticuerpos, moléculas similares a inmunoglobulina diseñadas mediante genotecnología, anticuerpos monocatenarios, proteínas de fusión, enzimas, moléculas coestimuladoras inmunes, moléculas inmunomoduladoras, mutantes negativos transdominantes de proteínas diana, toxinas, toxinas condicionales, sensibilizantes de quimioterapia o radioterapia, antígenos, proteínas supresoras de tumores, factores de crecimiento, proteínas de membrana, proteínas y péptidos vasoactivos, proteínas antiviricas o sus variantes.

Los ejemplos específicos de genes terapéuticos de interés son sin número, e incluyen, sin limitación, eritropoyetina, insulina, VEGF, FGF, Factor VIII, Factor IX, endostatina, angiostatina, GDNF, BDNF, NGF, EGF, CFTR, PEGF, IFN-alfa, IFN-gamma, IL-1, IL-2, IL-4, IL-6, IL-7, IL-10, IL-12, IL-21, GM-CSF, G-CSF, M-CSF, TNF- α , TNF- β , TGF- α , TGF- β , CD40, hirudina, y similares.

La presente invención también proporciona kits que comprenden los vectores de expresión de la invención. Por ejemplo, en una realización, la presente invención proporciona un kit que comprende al menos un polinucleótido que es al menos un 80 %, 85 %, 90 %, 95 %, 96 %, 97 %, 98 % o 99 % idéntico a un polinucleótido que está representado por una de las SEQ ID NO:1-10. Los kits pueden comprender los anteriores vectores o componentes que permiten su ensamblaje. Por ejemplo, el kit puede incluir un vector que puede linealizarse a través de la digestión con una enzima de restricción en un sitio que permite al usuario insertar una 5'UTR sintética o componentes individuales de una 5'UTR sintética. El kit puede comprender además componentes adicionales para el ensamblaje del vector, tales como la enzima de restricción, la ligasa, el tampón, y similares.

La presente invención proporciona además métodos para expresar un producto o una secuencia génica de interés en una célula hospedadora.

Por ejemplo, otra realización de la invención es un método para expresar un producto génico que comprende transfectar una célula hospedadora con una construcción génica sintética que comprende la 5'UTR sintética descrita en el presente documento.

Otra realización de la invención es un método para expresar un producto génico que comprende transfectar una célula hospedadora con una construcción génica sintética que comprende un polinucleótido al menos un 80 %, 85 %, 90 %, 95 %, 96 %, 97 %, 98 % o 99 % idéntico a un polinucleótido representado por la SEQ ID NO: 1 o SEQ ID NO:7.

Otra realización de la invención es un método para expresar una secuencia de interés en una célula hospedadora, que comprende las etapas de:

a. transfectar una célula hospedadora con un vector de expresión que comprende una construcción génica sintética descrita en el presente documento; y

b. cultivar dicha célula hospedadora en condiciones adecuadas para obtener la expresión de dicha secuencia de interés.

En otro método para expresar una secuencia de interés en una célula hospedadora, se puede insertar una secuencia de interés en un vector de expresión descrito en el presente documento que comprende un promotor, una 5'UTR sintética, y un sitio de clonación. Por ejemplo, otra realización de la divulgación es un método para expresar una secuencia de interés en una célula hospedadora, que comprende las etapas de:

a. insertar una secuencia de interés que se va a expresar en un vector de expresión descrito en el presente documento que comprende un promotor, una 5'UTR sintética, y un sitio de clonación en el sitio de clonación, en el que la secuencia de interés que se va a expresar incluye un ARN o una secuencia codificante de un polipéptido;

b. transfectar una célula hospedadora con el vector de expresión; y

c. cultivar dicha célula hospedadora en condiciones adecuadas para obtener la expresión de dicha secuencia de interés.

En otro método para expresar una secuencia de interés en una célula hospedadora, se puede insertar una 5'UTR sintética en el vector de expresión entre el promotor y la secuencia de interés, de tal manera que la porción que comprende el elemento de corte y empalme de un gen eucariota en la 5'UTR sintética es situado hacia el promotor y la porción que comprende al menos una porción de una 5'UTR de otro gen eucariota es situado hacia la secuencia de interés.

Por ejemplo, otra realización de la divulgación es un método para expresar una secuencia de interés en una célula hospedadora, que comprende las etapas de:

a. insertar una 5'UTR sintética descrita en el presente documento en un vector de expresión entre un promotor y una secuencia de interés que se va a expresar, en el que la secuencia de interés que se va a expresar incluye un ARN o una secuencia codificante de un polipéptido;

b. transfectar una célula hospedadora con dicho vector de expresión; y

c. cultivar dicha célula hospedadora en condiciones adecuadas para obtener la expresión de dicha secuencia de interés.

Otra realización de la descripción es un método para expresar una secuencia de interés en una célula hospedadora, que comprende las etapas de:

a. insertar un polinucleótido al menos un 80 %, 85 %, 90 %, 95 %, 96 %, 97 %, 98 o 99 % idéntico a un polinucleótido representado por la SEQ ID NO:1 o SEQ ID NO:7 en un vector de expresión entre un promotor y una secuencia de interés que se va a expresar, en el que la secuencia de interés que se va a expresar incluye un ARN o una secuencia codificante de un polipéptido;

b. transfectar una célula hospedadora con dicho vector de expresión; y

c. cultivar dicha célula hospedadora en condiciones adecuadas para obtener la expresión de dicha secuencia de interés.

Los métodos de la invención se llevan a cabo utilizando técnicas rutinarias. Salvo que se indique de otra forma, las

técnicas de ADN recombinante se llevan a cabo de acuerdo a los protocolos normalizados descritos en (Sambrook, *supra*) o en los Volúmenes 1 y 2 de Ausubel et al. (1994), Current Protocols in Molecular Biology, Current Protocols.

La presente invención proporciona también células hospedadoras que comprenden un vector de expresión de la invención.

En una realización, la célula hospedadora es una célula hospedadora de mamífero. En realizaciones específicas, la célula hospedadora es una célula de hámster, una célula de ratón, una célula de rata, una célula de conejo, una célula de gato, una célula de perro, una célula de bovino, una célula de cabra, una célula de vaca, una célula de cerdo, una célula de caballo, una célula de oveja, una célula de simio, una célula de mono, una célula de chimpancé, o una célula de ser humano. Los ejemplos específicos de células hospedadoras de la invención que comprenden una 5'UTR sintética incluyen, pero no de forma limitativa A549, ARPE-19, CH3/10T1/2, C2C12, aco2, COS7, FL-83B, HEK-293, HEPG2, HeLa, HT-1080, MDCK, P19, SH-SY5Y, Sol 8, y U87.

Las células hospedadoras de la invención se transfectan con un polinucleótido que comprende una 5'UTR sintética. La transfección de células hospedadoras es bien conocida en la técnica y se puede conseguir mediante varios métodos que incluyen, pero no de forma limitativa electroporación, infección vírica, transfección de plásmidos/vectores, transfección mediada por vector no vírico, bombardeo de partículas, y similares. La expresión de los productos génicos deseados implica cultivar las células hospedadoras transfectadas en condiciones adecuadas y medir la expresión del gen transfectado. Las condiciones de cultivo y los protocolos de expresión génica en células eucariotas son bien conocidos en la técnica.

La presente invención proporciona también organismos hospedadores que comprende un vector como se reivindica en el presente documento. Las 5'UTR sintéticas se pueden insertar directamente en el genoma de un organismo hospedador in vivo. Por ejemplo, en una realización, una 5'UTR sintética se introduce en un organismo hospedador para sustituir una 5'UTR natural introduciendo directamente en el organismo hospedador un vector que contiene una 5'UTR sintética de la invención flanqueada por secuencias homólogas a secuencias que flanquean la 5'UTR natural que se va a sustituir. En otra realización, una construcción génica sintética que comprende una 5'UTR sintética se inserta en el genoma de un organismo hospedador introduciendo un vector de integración que comprende la construcción génica en el organismo hospedador. Se puede controlar la especificidad tisular de la inserción, por ejemplo, por la ruta de administración del vector. En otra realización, una construcción génica sintética que comprende una 5'UTR sintética se introduce en un organismo hospedador introduciendo un vector no de integración que comprende la construcción génica en el organismo hospedador.

Se puede introducir una 5'UTR o una construcción génica sintética que comprenda una 5'UTR sintética en un organismo hospedador mediante soluciones ex vivo. Por ejemplo, se pueden transformar células autólogas o no autólogas con un vector que comprenda una 5'UTR sintética y a continuación introducirse en el organismo hospedador.

En otra realización, se introduce una 5'UTR sintética en el genoma de una célula u organismo hospedador utilizando un sistema de recombinasa específico de sitio tal como el sistema Cre/loxP. En esta realización, se consigue la integración estable de un fragmento de ADN tal como un fragmento que comprende una 5'UTR sintética o una construcción génica sintética que comprende una 5'UTR sintética en el genoma introduciendo sitios lox modificados en el genoma y el vector donante que evita la reescisión del ADN integrado (véase Metzgerand Feil, CurrentOpinion in Biotechnology, 10:470-476 (1999)). Adicionalmente, se pueden introducir sitios loxP heteroespecíficos ('floxed', secuencia de ADN que se encuentra entre dos sitios loxP) en el genoma que flanquea una región que se va a sustituir (Metzger y Feil, *supra*), tal como una 5'UTR endógena y una 5'UTR sintética o un plásmido donante. Los animales transgénicos con sitios cromosómicos lox P introducidos en el genoma pueden cruzarse con ratones transgénicos que expresan la recombinasa Cre impulsada por un promotor específico de tejido o célula, que son conocidos en la materia. La administración de un plásmido donante que comprende una 5'UTR sintética a la progenie de este cruzamiento dará como resultado la integración en tipos de tejidos o células específicos.

En otras realizaciones, otros sistemas de recombinación específicos de sitio tales como los divulgados en la publicación de patente de EE.UU. N.º 20060172377, se utilizan para introducir una 5'UTR sintética en el genoma de un organismo hospedador.

Otra realización de la invención es un organismo no humano que comprende un polinucleótido que es al menos un 80 %, 85 %, 90 %, 95 %, 96 %, 97 %, 98 % o 99 % idéntico a un polinucleótido representado por las SEQ ID NO:1-10.

Otra realización de la invención es un organismo no humano que comprende una construcción génica sintética que comprende un polinucleótido al menos un 80 %, 85 %, 90 %, 95 %, 96 %, 97 %, 98 % o 99 % idéntico a un polinucleótido representado por las SEQ ID NO:1-10.

Adicionalmente, se puede crear un organismo transgénico tal como un ratón que comprende una 5'UTR sintética o una construcción génica que comprende una 5'UTR sintética. Por ejemplo, se pueden generar ratones transgénicos inyectando un vector adecuado que comprende la 5'UTR sintética en los pronúcleo de oocitos de ratón fertilizados.

Como alternativa, se puede introducir el vector en los embriocitoblastos de ratón, que se microinyectan a continuación en los blastocistos de ratón. Los cigotos o blastocistos transformados se trasplantan a continuación a ratones hembra pseudoembarazadas. Las crías resultantes se seleccionan según la presencia de los polinucleótidos mediante la PCR o el análisis de la transferencia Southern. A continuación, los animales transgénicos heterocigóticos se cruzan entre sí para generar homocigotos. En una realización, la 5'UTR sintética sustituye una 5'UTR endógena de un gen de interés en el animal transgénico mediante la recombinación homóloga.

Por lo tanto, otra realización de la divulgación es un animal transgénico que comprende una secuencia del polinucleótido 5'UTR sintética que comprende un fragmento de un gen de la calcio ATPasa del retículo sarcoplásmico/endoplasmático eucariota que comprende el segundo intrón fusionado a un fragmento que comprende al menos una porción de la región 5'UTR de un gen de la caseína.

Otra realización de la invención es un organismo transgénico que comprende un polinucleótido que es al menos un 80 %, 85 %, 90 %, 95 %, 96 %, 97 %, 98 % o 99 % idéntico a un polinucleótido representado por las SEQ ID NO: 1-10.

Otra realización de la divulgación es un organismo transgénico que comprende una construcción génica sintética que comprende un polinucleótido que es al menos un 80 %, 85 %, 90 %, 95 %, 96 %, 97 %, 98 % o 99 % idéntico a un polinucleótido representado por las SEQ ID NO:1-10.

Los siguientes ejemplos son ilustrativos, de las realizaciones de la presente invención. Otras modificaciones y adaptaciones adecuadas que son obvias a los expertos en la materia están comprendidas en el alcance de la invención.

Ejemplos

Ejemplo 1

Se construyeron tres versiones diferentes de una 5'UTR sintética, y se insertaron en un vector en el que un promotor del citomegalovirus (CMV) humano dirige la expresión de un indicador de la beta-galactosidasa (secuencia codificante de LacZ). En la versión uno, se insertó poliG (12) entre el promotor y el indicador de la beta-galactosidasa como una 5'UTR sintética (Figura 9). En la versión dos, denominada 5U2, Se insertó la SEQ ID NO: 1 entre el promotor y el indicador de la beta-galactosidasa como una 5'UTR sintética (Figura 10). En la versión 3, denominada INXN-1, Se insertó la SEQ ID NO: 7 entre el promotor y el indicador de la beta-galactosidasa como una 5'UTR sintética (Figura 11). Cada transgén tiene una secuencia de poliadenilación de SV40 en la región reguladora 3'. En la Figura 3A se muestra esquemáticamente una versión genérica de cada vector conteniendo una 5'UTR sintética 5U2 o INXN-1. Las células HEK-293 y 1080 se transfectaron transitoriamente con cada vector de expresión.

Se midió la beta-galactosidasa de las células utilizando el Sistema Galacto-Star™ de Applied Biosystems (números de catálogo BM100S, BM300S, BM2500S, BY100S, BY300S, BY2500S) de acuerdo con el siguiente protocolo. Se lisaron las células en 50 µl de solución de lisis (Sistema Galacto-Star™) durante 10 minutos a temperatura ambiente. 100 µl de sustrato Galacto-Star® se sometieron a alícuotas en pocillo de una microplaca opaca blanca de 96 pocillo. 10 µl de solución de lisis celular se añadieron a 100 µl de sustrato Galacto- Star® y se incubaron durante 30 minutos. Se midió la señal luminosa utilizando un luminómetro de microplacas.

En la Figura 4 se muestran los resultados de las células HEK-293. La expresión del indicador de beta-galactosidasa aumentó de forma importante en células HEK-293 transfectadas con vectores que contienen la 5'UTR sintética INXN-1 o 5U2 en comparación con vectores que contienen poliG como la 5'UTR sintética. En la Figura 5 se muestra la expresión aumentada del indicador en células 1080 transfectadas con vectores que contienen INXN-1 o 5U2 en comparación con vectores que contienen poliG como la 5'UTR. El nivel de expresión de transgenes en vectores que contienen una poliG 5'UTR fue similar al control donde no estaba presente la 5'UTR (VVN- 2712, Figura 7, datos en la Figura 12). El control negativo para medir el antecedente del ensayo era un vector sin beta-galactosidasa (VVN-2713, Figura 8, datos en la Figura 12). La Figura 6 muestra que el gen indicador de la beta-galactosidasa sintética contiene la 5'UTR sintética INXN-1 se expresó aproximadamente 7,5 veces más en células HEK-293 y aproximadamente 2,5 veces más en células 1080 en comparación con un gen indicador sintético que contiene poliG como la 5'UTR, mientras que un gen indicador sintético que contenía la 5'UTR sintética 5U2 se expresó aproximadamente 8 veces más en células HEK-293 y aproximadamente 2,5 veces más en células 1080 en comparación con un gen indicador sintético que contenía poliG como la 5'UTR.

Ejemplo 2

Una investigación en la base de datos pública de nucleótidos no redundantes del GenBank, utilizando el algoritmo blastn con parámetros por defecto, utilizando la SEQ ID NO: 4 como secuencia de consulta dio como resultado los siguientes homólogos SERCA2 representativos relacionados en la Tabla 1. La SEQ ID NO: 4 representa un componente de la 5'UTR sintética representada por la SEQ ID NO: 7.

Tabla 1

Número de acceso	Descripción	Cobertura de consulta	Valor E	Identidad Máxima
EU365364	Gen twitch 2 (ATP2A2) que ralentiza el transporte de ATPasa CA++ del músculo cardíaco de Homo sapiens, exones 2, 3 y cds parcial.	99 %	2e-16	77 %
AM 137440	Gen atp2A2 para la calcio ATPasa 2 del retículo sarcoplásmico/endoplasmático de Equus caballus, exones 1-20.	51 %	3e-14	90 %
M33834	Gen Ca-2+■ ATPasa (SERCA2) de retículo sarcoplásmico/endoplasmático de Oryctolagus cuniculus, exones 1-3 y cds parcial	53 %	5e-05	80 %

Un fragmento de genoma de Equus caballus identificado a partir de los resultados de la investigación de BLAST indicada en la Tabla 1 (número de acceso público AM 137440) se comparó con la secuencia NM_001081765 del ARNm de SERCA2 de Equus caballus utilizando la función de alineamiento por pares en el programa DNA Strider 1.4f17 públicamente disponible (véase Marck, C, Nucleic Acids Research 16(5):1829-1837 (1988) y Douglas, S, Molecular Biotechnology 3(1):37-45 (1995) para las descripciones del DNA Strider). Las secuencias se alinearon utilizando el método Blocks utilizando parámetros por defecto para la penalización por emparejamiento incorrecto y la penalización por hueco. En la Figura 13 se muestra una porción de la alineación de la secuencia genómica de Equus caballus y ARNm, con flechas marcando el comienzo y el final del segundo intrón del gen SERCA2 de Equus caballus. La región de homología que es 5' para el intrón representa el exón 2, y la región de homología que es 3' para el intrón representa el exón 3.

A continuación se sintetizó un oligonucleótido que representaba un fragmento del gen SERCA2 de Equus caballus comprendiendo el segundo intrón, que se fusionó con la SEQ ID NO: 3 utilizando técnicas de ADN recombinante para crear una 5'UTR sintética. A continuación se insertó la 5'UTR sintética en un vector entre un promotor de citomegalovirus humano (CMV) y un gen indicador de la luciferasa. A continuación se usaron el vector y un vector control con una poliG 5'UTR para transfectar células 3T3. Se midió la actividad de la luciferasa utilizando un luminómetro y se compararon unidades de luz relativas entre ambos conjuntos de células. La expresión del indicador en células transfectadas con el vector que contenía la 5'UTR sintética se elevó en comparación con las células transfectadas con el vector que contenía la poliG 5'UTR.

Ejemplo 3

Una investigación en la base de datos pública de nucleótidos no redundantes del GenBank, utilizando el algoritmo blastn con parámetros por defecto, utilizando la SEQ ID NO: 3 como secuencia de consulta dio como resultado los siguientes homólogos representativos relacionados en la Tabla 2. La SEQ ID NO: 3 representa el componente de la 5'UTR de la caseína de bovino de la 5'UTR sintética representada por la SEQ ID NO: 1 y la SEQ ID NO: 7.

Tabla 2

Número de acceso	Descripción	Cobertura de consulta	Valor E	Identidad Máxima
DQ317447	ARNm de la beta caseína de Bubalus bubalis, cds completo	77 %	5e-10	97 %
NM_001009373	Beta caseína de Ovis aries (CSN2), ARNm.	77 %	6e-09	95 %
AY311384	Gen de la beta caseína de Capra hircus, promotor y exón 1.	56 %	5e-04	96 %
NM_001081852	Beta caseína de Equus caballus (CSN2), ARNm.	58 %	0,002	93 %
AY452035	Gen de la beta caseína de Sus scrofa, Región promotora, exón 1, y secuencia parcial.	56 %	0,006	93 %
AJ409279	Gen parcial de la beta caseína de Camelus dromedarius, región 5' flanqueante.	56 %	0,006	93 %
NM_001082759	Pre-beta-caseína de (AA-15 to 213) (LOC100009539) de Oryctolagus cuniculus, ARNm.	66 %	0,006	88 %
NM_001003086	Beta caseína (CSN2) de Canis lupus familiaris, ARNm.	67 %	0,88	83 %

El segundo intrón de cada gen SERCA2 representativo relacionado en la Tabla 1 se identifica usando un programa de

alineación por pares comparando la secuencia genómica con su respectiva secuencia de ARNm. Se sintetizó un primer conjunto de oligonucleótidos que comprendía el segundo intrón de SERCA2 para cada homólogo de SERCA2. A continuación se sintetizó un segundo conjunto de oligonucleótidos que representaba al menos una porción de 5'UTR de cada identificador homólogo de la beta caseína en la Tabla 2. La porción de 5'UTR comprende la porción de la cobertura de consulta identificada en los resultados de BLAST de la Tabla 2. A continuación se construyó un conjunto de 5'UTR sintética utilizando técnicas de ADN recombinante fusionando un único miembro del primer conjunto de oligonucleótidos que comprende el segundo intrón de SERCA2, o un oligonucleótido representado por SEQ ID NO:2 y 4-6 con un único miembro del segundo conjunto de oligonucleótidos que comprende al menos una porción de 5'UTR o un oligonucleótido representado por las SEQ ID NO:3 y 8-10, en donde el oligonucleótido que comprende el segundo intrón de SERCA2 se fusionó en 5' del oligonucleótido que comprendía la porción de 5'UTR. A continuación se añadieron sitios de restricción a cada extremo de la 5'UTR sintética mediante la PCR. Cada 5'UTR sintética única se insertó a continuación en un vector entre un promotor de CMV humano y un promotor lacZ. A continuación se transfectaron células 1080 en placas de 96 pocillos con cada vector así como un vector control con una poliG 5'UTR. Se midió a continuación la actividad de la beta galactosidasa utilizando el ensayo descrito en el Ejemplo 1 y a continuación se compararon los niveles de expresión relativos a la poliG 5'UTR.

Habiendo descrito completamente ahora la invención, las personas normalmente expertas en la técnica entenderán que la misma se puede llevar a cabo en un intervalo de condiciones y otros parámetros amplio y equivalente sin influir sobre el alcance de la invención o cualquiera de sus realizaciones.

LISTADO DE SECUENCIAS

<110> Intrexon Corporation Reed, Thomas D.

<120> 5'UTR SINTÉTICAS, VECTORES DE EXPRESIÓN, Y MÉTODOS PARA AUMENTAR LA EXPRESIÓN TRANSGÉNICA

<130> RS010

<150> 60/975.408

<151> 26-09-2007

<160> 14

<170> PatentIn versión 3.4

<210> 1

<211> 191

<212> ADN

<213> Artificial

<220>

<223> Sintética

<400> 1

```

acgcgctcgaa gaagggtgagt aatcttaaca tgctcttttt tttttttttt gctaattccct      60
tttgtgtgct gatgttagga tgacatttac aacaaatggt tgttcctgac aggaaaaacc      120
ttgctgggta ccttcggttc cggacacttc ttgtcctcta ctttggaata aaggaattga      180
gagcccaatt g                                191

```

<210> 2

<211> 120

<212> ADN

<213> Artificial

<220>

<223> Sintética

<400> 2

ES 2 614 402 T3

		cgaagaaggt gagtaatctt aacatgctct tttttttttt ttttgctaata cccctttgtg	60
		tgctgatgtt aggatgacat ttacaacaaa tgtttgttcc tgacaggaaa aacottgctg	120
5	<210> 3 <211> 53 <212> ADN <213> Artificial		
10	<220> <223> Sintética		
	<400> 3 ttcgtgccg gacactctt gtccttact ttggaaaaa ggaattgaga gcc 53		
15	<210> 4 <211> 120 <212> ADN <213> Artificial		
20	<220> <223> Sintética		
	<400> 4		
		cgaagaaggt gagtaatctt aacatgttct tttttttttt tttttttaata cccctttgtg	60
		tgctgatgtt aggatgacat ttacaacaaa tgtttgttcc tgacaggaaa aacottgctg	120
25	<210> 5 <211> 192 <212> ADN <213> Artificial		
30	<220> <223> Sintética		
35	<400> 5		
		agttaccggc tgaagaaggt aatcttaaca tgctgtttct gttttttttc ctctgttggt	60
		gtgctgatgg taagatgaca gttaaaacac atgtgtttgt ttcttacagg aaaaaccttg	120
		ctggaacttg tgattgagca gtttgaagac ttgctagtta ggattttatt actggcagca	180
		tgtatatctt tt	192
40	<210> 6 <211> 218 <212> ADN <213> Artificial		
45	<220> <223> Sintética		
	<400> 6		
		aattgccggc tgaagaaggt aaataatatt aacatgttat tttggagaga tgatgtgtgc	60
		aggctgattg atgtggacaa ttgaaacaaa tggtttgttt tttttttttt tttttttcct	120
		ttcctttttt aacaggaaaa accttgctgg aacttgatgat tgagcagttt gaagacttac	180
		tagttagaat ttactgctg gcagcatgta tatctttc	218

ES 2 614 402 T3

<210> 7
 <211> 199
 <212> ADN
 <213> Artificial
 5
 <220>
 <223> Sintética
 <400> 7
 10
 ggcgcgccac gcgtcgaaga aggtgagtaa tcttaacatg ctcttttttt tttttttttt 60
 taatcccttt tgtgtgctga tgtaggatg acatttaca caaatgtttg ttcctgacag 120
 gaaaaacctt gctgggtacc ttcgttgccg gacacttctt gtcctctact ttggaaaaaa 180
 ggaattgaga gcccaattg 199
 <210> 8
 <211> 43
 <212> ADN
 <213> Artificial
 15
 <220>
 <223> Sintética
 20
 <400> 8
 gcttcacctc ctctctgtc ctccactaaa ggacttgaca gcc 43
 <210> 9
 <211> 52
 <212> ADN
 <213> Artificial
 25
 <220>
 <223> Sintética
 30
 <400> 9
 atcctctgag ctctccttc tctctgtcc tccgctaaag gacttgacag cc 52
 <210> 10
 <211> 60
 <212> ADN
 <213> Artificial
 35
 <220>
 <223> Sintética
 40
 <400> 10
 atccattcag ctctccttc actctcttc ctctacttg gaaaaaagga atcgagagcc 60
 45
 <210> 11
 <211> 83
 <212> ADN
 <213> Artificial
 50
 <220>
 <223> Sintética
 <400> 11
 55
 gaaaaacctt gctggaactt gtgatcgagc agtttgaaga ctactagtt cgaatattat 60
 tgctggcagc gtgtatatct ttt 83

ES 2 614 402 T3

<210> 12
<211> 8011
<212> ADN
<213> Artificial
5
<220>
<223> Sintética
10
<400> 12
taaacaata ggggttccgc gcacatttcc ccgaaaagtg ccacctgacg tcacctggtc 60

ES 2 614 402 T3

ttaagggcta	tggcagggcc	tgccgccccg	acgttggctg	cgagccctgg	gccttcaccc	120
gaacttgggg	ggtggggtgg	ggaaaaggaa	gaaacgcggg	cgtattggcc	ccaatggggt	180
ctcgggtggg	tatcgacaga	gtgccagccc	tgggaccgaa	ccccgcgttt	atgaacaaac	240
gacccaacac	cgtgcgtttt	attctgtctt	tttattgccc	tcatagcgcg	ggttccttcc	300
ggtattgtct	ccttccgtgt	ttcactcgag	tcagaagaac	tcgtcaagaa	ggcgatagaa	360
ggcgatgcgc	tgcgaatcgg	gagcggcgat	accgtaaagc	acgaggaagc	ggtcagccca	420
ttcgccgcca	agctcttcag	caatatcacg	ggtagccaac	gctatgtcct	gatagcggtc	480
cgccacaccc	agccggccac	agtcgatgaa	tccagaaaag	cggccatttt	ccaccatgat	540
attcggcaag	caggcatcgc	catgggtcac	gacgagatcc	tcgccgtcgg	gcatgcgcgc	600
cttgagcctg	gcgaacagtt	cggctggcgc	gagcccctga	tgctcttcgt	ccagatcatc	660
ctgatcgaca	agaccggctt	ccatccgagt	acgtgctcgc	tcgatgcgat	gtttcgcttg	720
gtggtcgaat	gggcaggtag	ccggatcaag	cgtatgcagc	cggcgcattg	catcagccat	780
gatggatact	ttctcggcag	gagcaagggt	agatgacagg	agatcctgcc	ccggcacttc	840
gccccaatagc	agccagtccc	ttcccgcctc	agtgacaacg	tcgagcacag	ctgcgcgaagg	900
aacgcccgtc	gtggccagcc	acgatagccg	cgtgcctcgc	tcctgcagtt	cattcagggc	960
acoggacagg	tcggtcttga	caaaaagaac	cgggcgcccc	tcgcgtgaca	gccggaacac	1020
ggcggcatca	gagcagccga	ttgtctgttg	tgcccagtca	tagccgaata	gcctctccac	1080
ccaagcggcc	ggagaacctg	cgtgcaatcc	atcttgttca	atggccgatc	ccatggtggc	1140
tctagaggaa	tagctcagag	gccgaggcgg	cttcggcctc	tgcataaata	aaaaaaatta	1200
gtcagccatg	gggcggagaa	tgggcggaac	tgggcggagt	taggggcggg	atgggcggag	1260
ttaggggcgg	gactatggtt	gctgactaat	tgagatgcat	gctttgcata	cttctgcctg	1320
ctggggagcc	tggggacttt	ccacacctgg	ttgctgacta	attgagatgc	atgctttgca	1380
tacttctgcc	tgctggggag	cctggggact	ttccacaccc	taacaagctt	gcccgggcag	1440
cggataaaca	tttcacacag	gaaacagcta	tgaccatgat	tagctgagct	aactataacg	1500
gtcctaaggt	agcgaatcga	tgcgatcgct	taattaacct	gcagggatat	cccatggggg	1560
cgcgagctc	ttccccgggg	gaacagcatg	cgtgtcatgc	catggcctgg	gactagttct	1620
agagcaccgg	tgggcccga	gatctggatc	ccgaactgca	ttagttatta	atagtaatca	1680
attacggggt	cattagttca	tagcccatat	atggagttcc	gcgttacata	acttacggta	1740
aatggccgc	ctggctgacc	gccaacgac	ccccgcccat	tgacgtcaat	aatgacgtat	1800
gttcccatag	taacgccaat	agggactttc	cattgacgtc	aatgggtgga	gtatttacgg	1860
taaactgccc	acttggcagt	acatcaagtg	tatcatatgc	caagtacgcc	ccctattgac	1920

gtcaatgacg gtaaattggcc cgcctggcat tatgcccagt acatgacctt atgggacttt 1980
 cctacttggc agtacatcta cgtattagtc atcgctatta ccatggtgat gcggttttgg 2040
 cagtacatca atgggcgtgg atagcggttt gactcacggg gatttccaag tctccacccc 2100
 attgacgtca atgggagttt gttttggcac caaaatcaac gggactttcc aaaatgtcgt 2160
 aacaactccg ccccatgac gcaaattggc ggtaggcgtg tacggtggga ggtctatata 2220
 agcagagctg gtttagtgaa ccgtcagatc cgctagcctt ggggtacccc ggccggccgg 2280
 cgcgccacgc gtcgaagaag gtgagtaatc ttaacatgct cttttttttt ttttttgcta 2340
 atcccttttg tgtgctgatg ttaggatgac atttacaaca aatgtttgtt cctgacagga 2400
 aaaaccttgc tgggtacctt cgttgccgga cacttcttgt cctctacttt ggaaaaaagg 2460
 aattgagagc ccaattggcc accatgagat ctggcgaccc cgtggtgctg cagaggaggg 2520
 actgggagaa ccccggcgtg acccagctga acaggctggc cggccacccc cccttcgcca 2580
 gctggaggaa cagcgaggag gccaggaccg acaggcccag ccagcagctg aggagcctga 2640
 acggcgagtg gaggttcgcc tggttccccc ccccagaggc cgtgcccag agctggctgg 2700
 agtgcgacct gcccagggcc gacaccgtgg tggtgcccag caactggcag atgcacggct 2760
 acgacgcccc catctacacc aacgtgacct accccatcac cgtgaacccc cccttcgtgc 2820
 ccaccgagaa ccccaccggc tgctacagcc tgaccttcaa cgtggacgag agctggctgc 2880
 aggagggcca gaccaggatc atcttcgacg gcgtgaacag cgccttcac ctgtggtgca 2940
 acggcaggtg ggtgggctac ggccaggaca gcaggctgcc cagcgagttc gacctgagcg 3000
 ccttcctgag ggctggcgag aacaggctgg ccgtgatggt gctgaggtgg agcgacggca 3060
 gctacctgga ggaccaggat atgtggagga tgagcggcat cttcagggac gtgagcctgc 3120
 tgcacaagcc caccacccag atcagcgact tccatgtggc caccaggttc aacgacgact 3180
 tcagcagggc cgtgctggag gccgaggtgc agatgtgcgg cgagctgagg gactacctga 3240
 ggggtgaccgt gagcctgtgg cagggcgaga ccagggtggc cagcggcacc gcccccttcg 3300
 gcggcgagat catcgacgag aggggcggct acgcccagag ggtgaccctg aggctgaacg 3360
 tggagaaccc caagctgtgg agcgccgaga tccccaacct gtacagggcc gtggtggagc 3420
 tgcacaccgc cgacggcacc ctgatcgagg ccgaagcctg cgacgtgggc ttcaggagg 3480
 tgaggatcga gaacggcctg ctgctgctga acggcaagcc cctgctgatc aggggcgtga 3540
 acaggcacga gcaccacccc ctgcacggcc aggtgatgga cgagcagaca atggtgcagg 3600
 acatcctgct gatgaagcag aacaacttca acgcccgtgag gtgcagccac taccccaacc 3660
 accccctgtg gtacaccctg tgcgacaggt acggcctgta cgtggtggac gaggccaaca 3720
 tcgagaccca cggcatggtg cccatgaaca ggctgaccga cgaccccagg tggctgcccg 3780
 ccatgagcga gaggggtgacc aggatggtgc agagggacag gaaccacccc agcgtgatca 3840

tctggagcct	gggcaacgag	agcggccacg	gcgccaacca	cgacgccctg	tacaggtgga	3900
tcaagagcgt	ggaccccagc	aggcccgtag	agtagcaggg	cggcggcgcc	gacaccaccg	3960
ccaccgacat	catctgcccc	atgtacgcca	gggtggacga	ggaccagccc	ttccccgccg	4020
tgcccaagtg	gagcatcaag	aagtggctga	gcctgcccgg	cgagaccagg	cccctgatcc	4080
tgtgcgagta	cgcccacgct	atgggcaaca	gcctgggcgg	cttcgccaag	tactggcaag	4140
ccttcaggca	gtaccccagg	ctgcagggcg	gcttcgtgtg	ggactgggtg	gaccagagcc	4200
tgatcaagta	cgacgagaac	ggcaaccctt	ggagcgccca	cggcggcgac	ttcggcgaca	4260
cccccaacga	caggcagttc	tgcataaacg	gcctgggtgt	cgccgacagg	acccccacc	4320
ccgccctgac	cgaggccaag	caccagcagc	agttcttcca	gttcaggctg	agcggccaga	4380
ccatcgaggt	gaccagcgag	tacctgttca	ggcacagcga	caacgagctg	ctgcactgga	4440
tggtggccct	ggacggcaag	cccctggcca	gcggcgaggt	gcccctggac	gtggccccc	4500
agggcaagca	gctgatcgag	ctgcccagag	tgcccagccc	cgagagcgcc	ggacagctgt	4560
ggctgaccgt	gaggggtgtg	cagcccaacg	ccaccgcctg	gagcgaggct	ggccacatca	4620
gcgcctggca	gcagtggagg	ctggccgaga	acctgagcgt	gacctgccc	gccgccagcc	4680
acgccatccc	ccacctgacc	acaagcgaga	tggacttctg	catcgagctg	ggcaacaaga	4740
ggtggcagtt	caacaggcag	agcggcttcc	tgagccagat	gtggatcggc	gacaagaagc	4800
agctgctgac	ccccctgagg	gaccagttca	ccaggggccc	cctggacaac	gacatcgggc	4860
tgagcgaggc	caccaggatc	gaccccaacg	cctgggtgga	gaggtggaag	gccgctggcc	4920
actaccaggc	cgaggccgcc	ctgctgcagt	gtaccgccga	cacctggccc	gaogccgtgc	4980
tgatcaccac	cgcccacgcc	tggcagcacc	agggcaagac	cctgttcac	agcaggaaga	5040
cctacaggat	cgacggcagc	ggccagatgg	ctatcacctg	ggacgtggag	gtggccagcg	5100
acacccccca	ccccgccagg	atcggcctga	actgccagct	ggcccagggt	gccgagaggg	5160
tgaactggct	gggcctgggc	ccccaggaga	actaccccca	caggctgacc	gccgcctgct	5220
tcgacagggt	ggacctgccc	ctgagcgata	tgtacacccc	ctacgtgttc	cccagcgaga	5280
acggcctgag	gtgcggcacc	aggagctga	actacggccc	ccaccagtgg	aggggcgact	5340
tccagttcaa	catcagcagg	tacagccagc	agcagctgat	ggagaccagc	cacaggcacc	5400
tgctgcacgc	cgaggagggc	acctggctga	acatcgacgg	ctccacatg	ggcatcggcg	5460
gcgacgacag	ctggagcccc	agcgtgagcg	ccgagttcca	gctgagcgct	ggcagatacc	5520
actaccagct	ggtgtggtgc	cagaagtaac	atatgtgact	aactaggtac	gtagcggccg	5580
cgtcgacgat	caggtaaagt	tacccaattc	gccctatagt	gagtcgtatt	acaattcact	5640
cgatcgccct	tcccaacagt	tgcgcagcct	gaatggcgaa	tggagatcca	atttttaagt	5700

gtataatgtg ttaaactact gattctaatt gtttgtgtat tttagattca cagtcccaag 5760
 gctcatttca ggccccctcag tcctcacagt ctgttcatga tcataatcag ccataaccaca 5820
 tttgtagagg ttttacttgc tttaaaaaac ctcccacacc tccccctgaa cctgaaacat 5880
 aaaatgaatg caattgttgt tgtaacttg tttattgcag cttataatgg ttacaaataa 5940
 agcaatagca tcacaaattt cacaaataaa gcattttttt cactgcattc tagttgtggt 6000
 ttgtccaaac tcatcaatgt atcttaacgc gctgatttaa atcggtcgcg gtacgatgca 6060
 tattaccctg ttatccctac cgcggttact ggccgtcgtt ttacaacgtc gtgactggga 6120
 aaaccctggc gatgctcttc tcccggtgaa aacctctgac acatggctct tctaaatccg 6180
 gagtttaaac gcttccttca tgtgagcaaa aggccagcaa aaggccagga accgtaaaaa 6240
 ggccgcggtg ctggcggttt tccataggct ccgccccct gacgagcatc acaaaaatcg 6300
 acgctcaagt cagagggtggc gaaacccgac aggactataa agataccagg cgtttcccc 6360
 tggaagctcc ctctgogct ctctgttcc gacctgccg cttaccggat acctgtccgc 6420
 ctttctccct tcgggaagcg tggcgcttc tcatagctca cgctgtagg atctcagttc 6480
 ggtgtaggtc gtctgctcca agctgggctg tgtgcacgaa cccccgttc agcccgaccg 6540
 ctgcgcctta tccggtaact atcgtcttga gtccaacccg gtaagacacg acttatcgcc 6600
 actggcagca gccactggta acaggattag cagagcgagg tatgtaggcg gtgctacaga 6660
 gttcttgaag tgggtggcta actacggcta cactagaagg acagtatttg gtatctgcgc 6720
 tctgctgaag ccagttacct tcggaaaaag agttggtagc tcttgatccg gcaaacaac 6780
 caccgctggt agcgggtggt tttttgttg caagcagcag attacgcgc gaaaaaagg 6840
 atctcaagaa gatcctttga tctttctac ggggtctgac gctcagtga acgaaaactc 6900
 acgttaaggg attttggta tgcctagggt gcaaacagct attatgggtt ttatgggtct 6960
 accggtgcat gagattatca aaaaggatct tcacctagat ctttttaa ataaaaatgaa 7020
 gttttaaatc aatctaaagt atatatgagt aaacttggc tgacagttac caatgcttaa 7080
 tcagtgaggc acctatctca gcgatctgtc tatttcgttc atccatagtt gcctgactcc 7140
 ccgtcgtgta gataactacg atacgggagg gcttaccatc tggccccagt gctgcaatga 7200
 taccgcgaga cccacgctca ccggctccag atttatcagc aataaaccag ccagccggaa 7260
 gggccgagcg cagaagtggc cctgcaactt tatccgcctc catccagttc attaatgtt 7320
 gccgggaagc tagagtaagt agttcgccag ttaatagttt gcgcaacggt gttgccattg 7380
 ctacaggcat cgtgggtgtc cgctcgtcgt ttggtatggc ttcattcagc tccggttccc 7440
 aacgatcaag gcgagttaca tgatcccca tgttgtgcaa aaaagcgggt agctccttcg 7500
 gtcctccgat cgttgtcaga agtaagttgg ccgcagtggt atcactcatg gttatggcag 7560
 cactgcataa ttctcttact gtcatgcat ccgtaagatg cttttctgtg actggtgagt 7620

ES 2 614 402 T3

actcaaccaa gtcattctga gaatagtgtg tgcggcgacc gagttgctct tgccccggcgt 7680
 caatacggga taatacogcg ccacatagca gaactttaaa agtgctcatc attggaaaac 7740
 gttcttcggg gcgaaaactc tcaaggatct taccgctgtt gagatccagt tcgatgtaac 7800
 ccactcgtgc acccaactga tcttcagcat cttttacttt caccagcgtt tctgggtgag 7860
 caaaaacagg aaggcaaaat gccgcaaaaa agggaataag ggcgacacgg aaatgttgaa 7920
 tactcatact cttccttttt caatattatt gaagcattta tcagggttat tgtctcggga 7980
 gcggatacat atttgaatgt atttagaaaa a 8011

<210> 13
 <211> 8011
 <212> ADN
 <213> Artificial

<220>
 <223> Sintética

<400> 13

taaacaaata ggggttccgc gcacatttcc ccgaaaagtg ccacctgacg tcacctggtc 60
 ttaagggcta tggcagggcc tgccgccccg acgttggtctg cgagccctgg gccttcaccc 120
 gaacttgggg ggtggggtgg ggaaaaggaa gaaacgcggg cgtattggcc ccaatggggt 180
 ctcggtgggg tatcgacaga gtgccagccc tgggaccgaa ccccgcgttt atgaacaaac 240
 gacccaacac cgtgcgtttt attctgtctt tttattgccg tcatagcgcg ggttccttcc 300
 ggtattgtct ccttccgtgt ttactcgag tcagaagaac tcgtcaagaa ggcgatagaa 360
 ggcgatgcgc tgcaaatcgg gagcggcgat accgtaaagc acgaggaagc ggtcagccca 420
 ttcgccgcca agctcttcag caatatcacg ggtagccaac gctatgtcct gatagcggtc 480
 cgccacaccc agccggccac agtcgatgaa tccagaaaag cggccatttt ccaccatgat 540
 attcggcaag caggcatcgc catgggtcac gacgagatcc tcgccgtcgg gcattgcgcg 600
 cttgagcctg gcgaacagtt cggctggcgc gagccctga tgctcttcgt ccagatcatc 660
 ctgatcgaca agaccggctt ccatccgagt acgtgctcgc tcgatgcgat gtttcgcttg 720
 gtggtcgaat gggcaggtag ccggatcaag cgtatgcagc cgccgcattg catcagccat 780
 gatggatact ttctcggcag gagcaagggt agatgacagg agatcctgcc ccggcacttc 840
 gcccaatagc agccagtccc ttcccgcttc agtgacaacg tcgagcacag ctgcgcaagg 900
 aacgcccgtc gtggccagcc acgatagccg cgtgcctcg tcctgcagtt cattcagggc 960
 accggacagg tcggtcttga caaaaagaac cgggcgcccc tcgcttgaca gccggaacac 1020
 ggccggcatca gagcagccga ttgtctgttg tgcccagtc tagccgaata gcctctccac 1080
 ccaagcggcc ggagaacctg cgtgcaatcc atcttgttca atggccgatc ccatggtggc 1140

tctagaggaa tagctcagag gccgaggcgg cttcggcctc tgcataaata aaaaaatta	1200
gtcagccatg gggcgagaaa tgggcgggaa tgggcggagt tagggcgagg atggcgagg	1260
ttagggcggg gactatgggt gctgactaat tgagatgcat gctttgcata cttctgcctg	1320
ctggggagcc tggggacttt ccacacctgg ttgctgacta attgagatgc atgctttgca	1380
tacttctgcc tgctggggag cctggggact ttccacaccc taacaagctt gcccgggcag	1440
cggataacaa tttcacacag gaaacagcta tgaccatgat tagctgagct aactataacg	1500
gtcctaaggt agcgaatcga tgcgatcgct taattaacct gcagggatat cccatggggg	1560
ccgcgagctc tccccgggg gaacagcatg cgtgtcatgc catggcctgg gactagttct	1620
agagcaccgg tgggcccga gatctggatc ccgaactgca ttagttatta atagtaatca	1680
attacggggt cattagttca tagcccatat atggagttcc gcgttacata acttacggta	1740
aatggccgc ctggctgacc gcccaacgac ccccgcccat tgacgtcaat aatgacgtat	1800
gttcccatag taacgccaat agggactttc cattgacgtc aatgggtgga gtatttacgg	1860
taaactgcc acttggcagt acatcaagt tatcatatgc caagtacgcc ccctattgac	1920
gtcaatgacg gtaaattggc cgcctggcat tatgccagc acatgacctt atgggacttt	1980
cctacttggc agtacatcta cgtattagtc atcgtatta ccatggtgat gcggttttgg	2040
cagtacatca atgggcgtgg atagcggttt gactcacggg gatttccaag tctccacccc	2100
attgacgtca atgggagttt gttttggcac caaaatcaac gggactttcc aaaatgtcgt	2160
aacaactccg ccccatgac gcaaattggc ggtaggcgtg tacggtggga ggtctatata	2220
agcagagctg gtttagtgaa ccgtcagatc cgttagcctt ggggtacccc ggccggccgg	2280
cgcgccacgc gtcgaagaag gtgagtaatc ttaacatgct cttttttttt ttttttttta	2340
atcccttttg tgtgctgatg ttaggatgac atttacaaca aatgtttgtt cctgacagga	2400
aaaaccttgc tgggtacctt cgttgccgga cacttcttgt cctctacttt ggaaaaaagg	2460
aattgagagc ccaattggcc accatgagat ctggcgaccc cgtggtgctg cagaggaggg	2520
actgggagaa ccccgcgctg acccagctga acaggctggc cggccacccc cccttgcga	2580
gctggaggaa cagcgaggag gccaggaccg acaggcccag ccagcagctg aggagcctga	2640
acggcgagtg gaggttcgcc tggttccccg cccccgaggc cgtgcccag agctggctgg	2700
agtgcgacct gcccgaggcc gacaccgtgg tggtgcccag caactggcag atgcacggct	2760
acgacgcccc catctacacc aacgtgacct accccatcac cgtgaacccc cccttcgtgc	2820
ccaccgagaa cccacccggc tgctacagcc tgaccttcaa cgtggacgag agctggctgc	2880
aggagggcca gaccaggatc atcttcgacg gcgtgaacag cgccttccac ctgtggtgca	2940
acggcaggtg ggtgggctac ggccaggaca gcaggctgcc cagcgagttc gacctgagcg	3000
ccttcctgag ggctggcgag aacaggctgg ccgtgatggt gctgaggtgg agcgacggca	3060

gctacctgga	ggaccaggat	atgtggagga	tgagcgccat	cttcagggac	gtgagcctgc	3120
tgcacaagcc	caccacccag	atcagcgact	tccatgtggc	caccagggtc	aacgacgact	3180
tcagcagggc	cgtgctggag	gccgaggtgc	agatgtgcgg	cgagctgagg	gactacctga	3240
gggtgaccgt	gagcctgtgg	cagggcgaga	cccaggtggc	cagcggcacc	gcccccttcg	3300
gcggcgagat	catcgacgag	aggggcggct	acgccgacag	ggtgaccctg	aggctgaacg	3360
tggagaacct	caagctgtgg	agcgccgaga	tccccaacct	gtacagggcc	gtggtggagc	3420
tgcacaccgc	cgacggcacc	ctgatcgagg	ccgaagcctg	cgacgtgggc	ttcagggagg	3480
tgaggatcga	gaacggcctg	ctgctgctga	acggcaagcc	cctgctgac	aggggcgtga	3540
acaggcacga	gcaccacccc	ctgcacggcc	aggtgatgga	cgagcagaca	atggtgcagg	3600
acatcctgct	gatgaagcag	aacaacttca	acgccgtgag	gtgcagccac	tacccaacc	3660
accccctgtg	gtacacccctg	tgcgacaggt	acggcctgta	cgtggtggac	gaggccaaca	3720
tcgagaccca	cggcatggtg	cccatgaaca	ggctgaccga	cgaccccagg	tggctgcccg	3780
ccatgagcga	gagggtgacc	aggatggtgc	agagggacag	gaaccacccc	agcgtgatca	3840
tctggagcct	gggcaacgag	agcggccacg	gcgccaacca	cgacgccctg	tacaggtgga	3900
tcaagagcgt	ggaccccagc	aggcccgtgc	agtacgaggg	cgggcgccgc	gacaccaccg	3960
ccaccgacat	catctgcccc	atgtacgcca	gggtggacga	ggaccagccc	ttccccgccc	4020
tgcccaagtg	gagcatcaag	aagtggctga	gcctgcccgg	cgagaccagg	cccctgatcc	4080
tgtgcgagta	cgcccacgct	atgggcaaca	gcctggcgcg	cttcgccaag	tactggcaag	4140
ccttcaggca	gtaccccagg	ctgcagggcg	gcttcgtgtg	ggactgggtg	gaccagagcc	4200
tgatcaagta	cgacgagaac	ggcaacccct	ggagcgccta	cggcggcgac	ttcggcgaca	4260
cccccaacga	caggcagttc	tgcataaacg	gcctgggtgt	cgccgacagg	acccccacc	4320
ccgccctgac	cgaggccaag	caccagcagc	agttcttcca	gttcaggctg	agcggccaga	4380
ccatcgaggt	gaccagcgag	tacctgttca	ggcacagcga	caacgagctg	ctgcactgga	4440
tgggtggccct	ggacggcaag	cccctggcca	gcggcgaggt	gcccctggac	gtggcccccc	4500
agggcaagca	gctgatcgag	ctgcccagag	tgcccagacc	cgagagcgcc	ggacagctgt	4560
ggctgaccgt	gaggggtggtg	cagcccaacg	ccaccgcctg	gagcgaggct	ggccacatca	4620
gcgcctggca	gcagtggagg	ctggccgaga	acctgagcgt	gaccctgccc	gccgccagcc	4680
acgccatccc	ccacctgacc	acaagcgaga	tggacttctg	catcgagctg	ggcaacaaga	4740
ggtggcagtt	caacaggcag	agcggtctcc	tgagccagat	gtggatcggc	gacaagaagc	4800
agctgctgac	ccccctgagg	gaccagttca	ccagggcccc	cctggacaac	gacatcggcg	4860
tgagcgaggc	caccaggatc	gacccaacg	cctgggtgga	gaggtggaag	gccgctggcc	4920

actaccaggc cgagggccgc ctgctgcagt gtaccgccga caccctggcc gacgccgtgc 4980
 tgatcaccac cgcccacgcc tggcagcacc agggcaagac cctgttcac agcaggaaga 5040
 cctacaggat cgacggcagc ggccagatgg ctatcacccg ggacgtggag gtggccagcg 5100
 acacccccca ccccgccagg atcggcctga actgccagct ggcccagggt gccgagaggg 5160
 tgaactggct gggcctgggc ccccaggaga actaccccga caggctgacc gccgcctgct 5220
 tcgacagggt ggacctgccc ctgagcgata tgtacacccc ctacgtgttc ccagcgaga 5280
 acggcctgag gtgcggcacc agggagctga actacggccc ccaccagtgg agggggcgact 5340
 tccagttcaa catcagcagg tacagccagc agcagctgat ggagaccagc cacaggcacc 5400
 tgctgcacgc cgaggagggc acctggctga acatcgacgg ctccacatg ggcatcggcg 5460
 gcgacgacag ctggagcccc agcgtgagcg ccgagttcca gctgagcgct ggcagatacc 5520
 actaccagct ggtgtggtgc cagaagtaac atatgtgact aactaggtac gtagcggccg 5580
 cgtcgacgat caggtaagtg tacccaattc gccctatagt gagtctgatt acaattcact 5640
 cgatcgccct tcccaacagt tgcgcagcct gaatggcgaa tggagatcca atttttaagt 5700
 gtataatgtg ttaaactact gattctaatt gtttgtgtat tttagattca cagtcccaag 5760
 gctcatttca ggcccctcag tcctcacagt ctgttcacga tcataatcag ccataccaca 5820
 tttgtagagg ttttacttgc tttaaaaaac ctcccacacc tccccctgaa cctgaaacat 5880
 aaaatgaatg ccattgttgt tgtaacttg tttattgcag cttataatgg ttacaaataa 5940
 agcaatagca tcacaaattt cacaataaaa gcattttttt cactgcattc tagttgtggt 6000
 ttgtccaaac tcatcaatgt atcttaacgc gctgatttaa atcggtcgcg gtacgatgca 6060
 tattaccctg ttatccctac cgcggttact ggccgtcgtt ttacaacgtc gtgactggga 6120
 aaaccctggc gatgctcttc tcccggtgaa aacctctgac acatggctct tctaaatccg 6180
 gagtttaaac gcttccttca tgtgagcaaa aggccagcaa aaggccagga accgtaaaaa 6240
 ggccgcgttg ctggcgtttt tccataggct ccgccccctt gacgagcatc acaaaaatcg 6300
 acgctcaagt cagaggtggc gaaacccgac aggactataa agataccagg cgtttccccc 6360
 tggaagctcc ctctgctgct ctctgttcc gacctgccc cttaccggat acctgtccgc 6420
 ctttctccct tcgggaagcg tggcgcttct tcatagctca cgctgtagggt atctcagttc 6480
 ggtgtaggtc gttcgtcca agctgggctg tgtgcacgaa cccccgctc agcccgaccg 6540
 ctgcgcctta tccggtaact atcgtcttga gtccaacccg gtaagacacg acttatcgcc 6600
 actggcagca gccactggta acaggattag cagagcgagg tatgtaggcg gtgctacaga 6660
 gttcttgaag tgggtggccta actacggcta cactagaagg acagtatttg gtatctgcgc 6720
 tctgctgaag ccagttacct tcggaaaaag agttggtagc tottgatccg gcaaacaaac 6780
 caccgctggt agcgggtggt tttttgtttg caagcagcag attacgcgca gaaaaaaag 6840

ES 2 614 402 T3

```

atctcaagaa gatcctttga tcttttctac ggggtctgac gctcagtggg acgaaaactc 6900
acgttaaggg attttgggtca tgcctaggtg gcaaacagct attatgggta ttatgggtct 6960
accggtgcat gagattatca aaaaggatct tcacctagat ccttttaaat taaaaatgaa 7020
gttttaaatc aatctaaagt atatatgagt aaacttgggtc tgacagttac caatgcttaa 7080
tcagtgaggc acctatctca gcgatctgtc tatttcgttc atccatagtt gcctgactcc 7140
ccgtcgtgta gataactacg atacgggagg gcttaccatc tggccccagt gctgcaatga 7200
taccgcgaga cccacgctca ccggctccag atttatcagc aataaaccag ccagccggaa 7260
gggccgagcg cagaagtggg cctgcaactt tatccgcctc catccagtct attaattgtt 7320
gccgggaagc tagagtaagt agttcgccag ttaatagttt gcgcaacggt gttgccattg 7380
ctacaggcat cgtgggtgtca cgctcgtcgt ttggtatggc ttcattcagc tccggttccc 7440
aacgatcaag gcgagttaca tgatccccc tgttgtgcaa aaaagcgggt agctccttcg 7500
gtcctccgat cgttgtcaga agtaagttgg ccgcagtgtt atcactcatg gttatggcag 7560
cactgcataa ttctcttact gtcatgccat ccgtaagatg cttttctgtg actggtgagt 7620
actcaaccaa gtcattctga gaatagtgtg tgcggcgacc gagttgctct tgccccgct 7680
caatacggga taataccgcg ccacatagca gaactttaaa agtgctcatc attggaaaac 7740
gttcttcggg gcgaaaactc tcaaggatct taccgctgtt gagatccagt tcgatgtaac 7800
ccactcgtgc acccaactga tcttcagcat cttttacttt caccagcgtt tctgggtgag 7860
caaaaacagg aaggcaaaat gccgcaaaa agggaataag ggcgacacgg aaatgttgaa 7920
tactcatact cttccttttt caatattatt gaagcattta tcagggttat tgtctcggga 7980
gcggatacat atttgaatgt atttagaaaa a 8011

```

<210> 14
 <211> 7844
 <212> ADN
 <213> Artificial

<220>
 <223> Sintética

<400> 14

```

taaacaaata ggggttccgc gcacatttcc ccgaaaagtg ccacctgacg tcacctggtc 60
ttaagggcta tggcagggcc tgccgccccg acgttggctg cgagccctgg gccttcaccc 120
gaacttgggg ggtggggtgg ggaaaaggaa gaaacgcggg cgtattggcc ccaatggggt 180
ctcgggtggg tatcgacaga gtgccagccc tgggaccgaa ccccgcttt atgaacaaac 240
gaaccaacac cgtgcgtttt attctgtctt ttatttgcgc tcatagcgcg gggttccttc 300
gggtattgtc cttccgtgt ttactcagag tcagaagaac tcgtcaagaa ggcgatagaa 360

```

ggcgatgcgc tgcgaatcgg gagcggcgat accgtaaagc acgaggaagc ggtagccca 420
 ttccgcccca agctcttcag caatatcacg ggtagccaac gctatgtcct gatagcggtc 480
 cggcacaccc agccggccac agtcgatgaa tccagaaaag cggccatttt ccaccatgat 540
 attcggcaag caggcatcgc catgggtcac gacgagatcc tcgccgtcgg gcatgcgcgc 600
 cttgagcctg gcgaacagtt cggctggcgc gagccctga tgcctctcgt ccagatcatc 660
 ctgatcgaca agaccggctt ccatccgagt acgtgctcgc tcgatgcgat gtttcgcttg 720
 gtggtcgaat gggcaggtag ccgatcaag cgtatgcagc cggcgatttg catcagccat 780
 gatggatact ttctcggcag gagcaagggt agatgacagg agatcctgcc ccggcacttc 840
 gcccataagc agccagtcct ttcccgttc agtgacaacg tcgagcacag ctgcgcaagg 900
 aacgcccgtc gtggccagcc acgatagccg cgtgcctcgc tctgcagtt cattcagggc 960
 accggacagg tcggtcttga caaaaagaac cgggcgcccc tcgctgaca gccggaacac 1020
 ggcggcatca gagcagccga ttgtctgttg tgcccagtc tagccgaata gcctctccac 1080
 ccaagcggcc ggagaacctg cgtgcaatcc atcttgttca atggccgatc ccatggtggc 1140
 tctagaggaa tagctcagag gccgaggcgg cttcggcctc tcgataaata aaaaaatta 1200
 gtcagccatg ggcgagagaa tgggcggaac tgggcggagt tagggcgagg atggcgagg 1260
 ttagggcgag gactatggtt gctgactaat tgagatgcat gctttgcata cttctgcctg 1320
 ctggggagcc tggggacttt ccacacctgg ttgctgacta attgagatgc atgctttgca 1380
 tacttctgcc tgcctgggag cctggggact ttccacaccc taacaagctt gcccgggcag 1440
 cggataacaa ttccacacag gaaacagcta tgaccatgat tagctgagct aactataacg 1500
 gtcctaaggt agcgaatcga tgcgatcgt taattaacct gcagggatat cccatggggg 1560
 ccgagagctc tccccgggg gaacagcatg cgtgtcatgc catggcctgg gactagttct 1620
 agagcacccg tgggcccga gatctggatc ccgaactgca ttagttatta atagtaatca 1680
 attacggggt cattagttca tagcccatat atggagttcc gcgttacata acttacggta 1740
 aatggcccgc ctggctgacc gcccaacgac ccccgcccat tgacgtcaat aatgacgtat 1800
 gttcccatag taacgccaat agggactttc cattgacgtc aatgggtgga gtatttacgg 1860
 taaactgcc acttggcagt acatcaagt tatcatatgc caagtacgcc ccctattgac 1920
 gtcaatgacg gtaaatggcc cgcctggcat tatgccagc acatgacctt atgggacttt 1980
 cctacttggc agtacatcta cgtattagtc atcgctatta ccatggtgat gcggttttgg 2040
 cagtacatca atggcggtg atagcggttt gactcacggg gatttccaag tctccacccc 2100
 attgacgtca atgggagttt gttttggcac caaatcaac gggactttcc aaaatgtcgt 2160
 aacaactccg cccattgac gcaaatgggc ggtaggcgtg tacgggtggga ggtctatata 2220
 agcagagctg gtttagtgaa ccgtcagatc cgctagcctt ggggtacccc ggccggccgg 2280

ES 2 614 402 T3

cgcgccacgc	gtgggggggg	ggggcaattg	gccaccatga	gatctggcga	ccccgtggtg	2340
ctgcagagga	gggactggga	gaaccccggc	gtgacccagc	tgaacaggct	ggccgcccac	2400
cccccccttcg	ccagctggag	gaacagcgag	gaggccagga	ccgacaggcc	cagccagcag	2460
ctgaggagcc	tgaacggcga	gtggagggtc	gcctgggttc	ccgccccga	ggccgtgccc	2520
gagagctggc	tggagtgcga	cctgcccag	gccgacaccg	tgggtggtgcc	cagcaactgg	2580
cagatgcacg	gctacgacgc	ccccatctac	accaacgtga	cctaccccat	caccgtgaac	2640
cccccccttcg	tgcccaccga	gaaccccacc	ggctgctaca	gcctgacctt	caacgtggac	2700
gagagctggc	tgcaggaggg	ccagaccagg	atcatcttcg	acggcgtgaa	cagcgccttc	2760
cacctgtggt	gcaacggcag	gtgggtgggc	tacggccagg	acagcaggct	gccagcgag	2820
ttcgacctga	gcgccttcct	gagggctggc	gagaacaggc	tggccgtgat	ggtgctgagg	2880
tggagcgacg	gcagctacct	ggaggaccag	gatattgtga	ggatgagcgg	catcttcagg	2940
gacgtgagcc	tgctgcacaa	gccaccacc	cagatcagcg	acttccatgt	ggccaccagg	3000
ttcaacgacg	acttcagcag	ggccgtgctg	gaggccgagg	tgcagatgtg	cggcgagctg	3060
agggactacc	tgagggtgac	cgtgagcctg	tggcaggggc	agaccaggt	ggccagcggc	3120
accgccccct	tcggcggcga	gatcatcgac	gagagggggc	gctacgccga	cagggtgacc	3180
ctgaggctga	acgtggagaa	ccccagctg	tggagcgccg	agatcccaa	cctgtacagg	3240
gccgtggtgg	agctgcacac	cgccgacggc	accctgatcg	aggccgaagc	ctgcgacgtg	3300
ggcttcaggg	aggtgaggat	cgagaacggc	ctgctgctgc	tgaacggcaa	gccctgctg	3360
atcagggggc	tgaacaggca	cgagcaccac	cccctgcaog	gccaggtgat	ggacgagcag	3420
acaatggtgc	aggacatcct	gctgatgaag	cagaacaact	tcaacgccgt	gaggtgcagc	3480
cactaccca	accaccccc	gtggtacacc	ctgtgcgaca	ggtacggcct	gtacgtggtg	3540
gacgaggcca	acatcgagac	ccacggcatg	gtgcccata	acaggctgac	cgacgacccc	3600
aggtggctgc	ccgccatgag	cgagaggggt	accaggatgg	tgcagaggga	caggaaccac	3660
cccagcgtga	tcatctggag	cctgggcaac	gagagcggcc	acggcgccaa	ccacgacgcc	3720
ctgtacaggt	ggatcaagag	cgtggacccc	agcaggcccg	tgcagtacga	gggcggcggc	3780
gccgacacca	ccgccaccga	catcatctgc	cccatgtacg	ccagggtgga	cgaggaccag	3840
cccttccccg	ccgtgcccc	gtggagcatc	aagaagtggc	tgagcctgcc	cggcgagacc	3900
aggccccctga	tcctgtgcga	gtacgcccac	gctatgggca	acagcctggg	cggtctcgcc	3960
aagtactggc	aagccttcag	gcagtacccc	aggctgcagg	gcggcttcgt	gtgggactgg	4020
gtggaccaga	gcctgatcaa	gtacgacgag	aacggcaacc	cctggagcgc	ctacggcggc	4080
gacttcggcg	acacccccaa	cgacaggcag	ttctgcatga	acggcctggt	gttcgcccgc	4140

```

aggaccccc accccgcctt gaccgaggcc aagcaccagc agcagttctt ccagttcagg 4200
ctgagcggcc agaccatcga ggtgaccagc gagtaoctgt tcaggcacag cgacaacgag 4260
ctgctgcact ggatggtggc cctggacggc aagcccttgg ccagcggcga ggtgcccttg 4320
gacgtggccc cccagggcaa gcagctgac gagctgcccg agctgcccc a gcccagagagc 4380
gccggacagc tgtggctgac cgtgaggggt gtgcagccca acgccaccgc ctggagcgag 4440
gctggccaca tcagcgcttg gcagcagtg aggtggccg agaacctgag cgtgaccttg 4500
cccgccgcca gccacgcat cccccacctg accacaagcg agatggactt ctgcatcgag 4560
ctgggcaaca agaggtggca gttcaacagg cagagcggct tcctgagcca gatgtggatc 4620
ggcgacaaga agcagctgct gacccccctg agggaccagt tcaccagggc cccctggac 4680
aacgacatcg gcgtgagcga ggccaccagg atcgacccca acgcctgggt ggagaggtgg 4740
aaggccgctg gccactacca ggccgaggcc gccctgctgc agtgtaccgc cgacaccctg 4800
gccgacggcg tgctgatcac caccgcccac gctggcagc accagggcaa gacctgttc 4860
atcagcagga agacctacag gatcgacggc agcggccaga tggctatcac cgtggacgtg 4920
gaggtggcca gcgacacccc ccaccccgcc aggatcggcc tgaactgcc gctggcccag 4980
gtggccgaga ggggtgaactg gctgggcctg ggccccagg agaactacc cgacaggctg 5040
accgcccgtt gcttcgacag gtgggacctg cccctgagcg atatgtacac cccctacgtg 5100
ttccccagcg agaacggcct gaggtgcggc accagggagc tgaactacgg cccccaccag 5160
tggaggggcg acttccagtt caacatcagc aggtacagcc agcagcagct gatggagacc 5220
agccacaggc acctgctgca cgccgaggag ggcaacctggc tgaacatcga cggcttccac 5280
atgggcatcg gcggcgacga cagctggagc cccagcgtga gcgcccagtt ccagctgagc 5340
gctggcagat accactacca gctggtgtgg tgccagaagt aacatatgtg actaactagg 5400
tacgtagcgg ccgctcgac gatcaggtaa gtgtacccaa ttcgccctat agtgagtcgt 5460
attacaattc actcgatcgc ccttcccaac agttgcgcag cctgaatggc gaatggagat 5520
ccaattttta agtgtataat gtgttaaact actgattcta attgtttgtg tatttttagat 5580
tcacagtccc aaggctcatt tcaggccctt cagtcctcac agtctgttca tgatcataat 5640
cagccatacc acattttagt aggttttact tgctttaaaa aacctccac acctccccct 5700
gaacctgaaa cataaaatga atgccattgt tgttgtaaac ttgtttattg cagcttataa 5760
tggttacaaa taaagcaata gcatcacaaa tttcacaaat aaagcatttt tttcactgca 5820
ttctagttgt ggtttgtcca aactcatcaa tgtatcttaa cgcgctgatt taaatcggtc 5880
cgcgtacgat gcatattacc ctgttatccc taccgcggtt actggccgtc gttttacaac 5940
gtcgtgactg ggaaaaccct ggcgatgctc ttctcccgtt gaaaacctct gacacatggc 6000
tcttctaaat ccggagttta aacgcttcct tcatgtgagc aaaaggccag caaaaggcca 6060

```

ES 2 614 402 T3

```

ggaaccgtaa aaagggcgcg ttgctggcgt ttttccatag gctccgcccc cctgacgagc 6120
atcacaaaaa tcgacgctca agtcagaggt ggcgaaaccc gacaggacta taaagatacc 6180
aggcgtttcc ccctggaagc tccctcgtgc gctctcctgt tccgaccctg ccgcttaccg 6240
gatacctgtc cgcctttctc ccttcgggaa gcgtggcgct ttctcatagc tcacgctgta 6300
ggtatctcag ttcggtgtag gtcgttcgct ccaagctggg ctgtgtgcac gaaccccccg 6360
ttcagcccga ccgctgcgcc ttatccggtc actatcgtct tgagtccaac ccgtaagac 6420
acgacttata gccactggca gcagccactg gtaacaggat tagcagagcg aggtatgtag 6480
gcggtgctac agagttcttg aagtgggtggc ctaactacgg ctacactaga aggacagtat 6540
ttggtatctg cgctctgctg aagccagtta ccttcggaaa aagagttggg agctcttgat 6600
ccggcaaaaca aaccaccgct ggtagcgggtg gtttttttgt ttgcaagcag cagattacgc 6660
gcagaaaaaa aggatctcaa gaagatcctt tgatcttttc tacgggggtct gacgctcagt 6720
ggaacgaaaa ctacggttaa gggatttttg tcatgcctag gtggcaaaca gctattatgg 6780
gtattatggg tctaccggtg catgagatta tcaaaaagga tcttcaccta gatcctttta 6840
aattaaaaat gaagttttta atcaatctaa agtatatatg agtaaacttg gtctgacagt 6900
taccaatgct taatcagtga ggcacctatc tcagcgatct gtctatttcg ttcattcata 6960
gttgctgac tccccgtcgt gtagataact acgatacggg agggcttacc atctggcccc 7020
agtgctgcaa tgataccgcg agaccacgc tcaccggctc cagatttatc agcaataaac 7080
cagccagccg gaagggccga gcgcagaagt ggtcctgcaa ctttatccgc ctccatccag 7140
tctattaatt gttgccggga agctagagta agtagttcgc cagttaatag tttgcgcaac 7200
gttggtgcca ttgctacagg catcgtgggt tcacgctcgt cgtttggtat ggcttcattc 7260
agctccggtt cccaacgata aaggcgagtt acatgatccc ccatggtgtg caaaaaagcg 7320
gttagctcct tcggtcctcc gatcgttgtc agaagtaagt tggccgcagt gttatcactc 7380
atggttatgg cagcaactga taattctctt actgtcatgc catccgtaag atgcttttct 7440
gtgactggtg agtactcaac caagtcattc tgagaatagt gtatgcggcg accgagttgc 7500
tcttgcccgg cgtcaatacg ggataatacc gcgccacata gcagaacttt aaaagtgtc 7560
atcattggaa aacgttcttc ggggcgaaaa ctctcaagga tcttaccgct gttgagatcc 7620
agttcgatgt aaccactcgt tgcacccaac tgatcttcag catcttttac tttcaccagc 7680
gtttctgggt gagcaaaaac aggaaggcaa aatgccgcaa aaaagggaat aaggcgac 7740
cggaaatgtt gaatactcat actcttcctt tttcaatatt attgaagcat ttatcagggt 7800
tattgtctcg ggagcggata catatttgaa tgtatttaga aaaa 7844

```

REIVINDICACIONES

1. Un vector de expresión que comprende una construcción de un gen sintético que comprende, cuando se dispone desde 5' a 3', un promotor, un polinucleótido quimérico y una secuencia de interés que se va a expresar, en donde:
 - a. el polinucleótido quimérico comprende el segundo intrón de un gen de la calcio ATPasa del retículo sarcoplásmico/endoplasmático y al menos una porción de la región 5' no traducida (5'UTR) de un gen de la caseína; y
 - b. el polinucleótido quimérico está situado entre el promotor y la secuencia de interés que se va a expresar, en donde el segundo intrón del gen de la calcio ATPasa del retículo sarcoplásmico/endoplasmático está situado hacia el promotor y el fragmento de polinucleótido de la región 5' no traducida (5'UTR) del gen de la caseína está situado hacia la secuencia de interés que se va a expresar.
2. Un método para expresar una secuencia de interés en una célula hospedadora, que comprende las etapas de:
 - a. transfectar una célula hospedadora con el vector de expresión de la reivindicación 1; y
 - b. cultivar dicha célula hospedadora en condiciones adecuadas para obtener la expresión de dicha secuencia de interés.
3. El vector de expresión de la reivindicación 1, en donde el vector de expresión es un vector de expresión eucariota.
4. El vector de expresión de la reivindicación 1,
 - en donde el vector comprende un fragmento de un gen SERCA2 canino, un gen SERCA2 humano o un gen SERCA2 de ratón, y
 - un fragmento de un gen de la beta caseína de bovino, un gen de la beta caseína de ratón, un gen de la beta caseína de rata o un gen de la beta caseína de oveja.
5. El vector de expresión de la reivindicación 4, en donde el vector comprende una secuencia de polinucleótidos que es al menos idéntica en un 80 % a una de las SEQ ID NO: 2 y 4-6 y una secuencia de polinucleótidos que es al menos idéntica en un 80 % a una de las SEQ ID NO: 3 y 8-10.
6. El vector de expresión de la reivindicación 1, en donde el vector comprende la SEQ ID NO: 2 y comprende la SEQ ID NO: 3.
7. El vector de expresión de la reivindicación 1, en donde el vector comprende la SEQ ID NO: 4 y la SEQ ID NO: 3.
8. El vector de expresión de la reivindicación 6, en donde el vector comprende una secuencia de polinucleótidos que es al menos idéntica en un 80 % a la SEQ ID NO: 1.
9. El vector de expresión de la reivindicación 7, en donde el vector comprende una secuencia de polinucleótidos que es al menos idéntica en un 80 % a la SEQ ID NO: 7.
10. El vector de expresión de la reivindicación 6, en donde el vector comprende la SEQ ID NO: 1.
11. El vector de expresión de la reivindicación 7, en donde el vector comprende la SEQ ID NO: 7.
12. El vector de expresión de la reivindicación 1,
 - en donde el vector comprende una secuencia de polinucleótidos que es al menos idéntica en un 80 % a un polinucleótido que comprende al menos 60 nucleótidos consecutivos de la SEQ ID NO: 4, y una secuencia de polinucleótidos que es al menos idéntica en un 80 % a una secuencia de polinucleótidos que comprende al menos 30 nucleótidos consecutivos de la SEQ ID NO: 3.
13. Una célula hospedadora que comprende el vector de expresión de una cualquiera de las reivindicaciones 1 o 3 a 12.
14. Un organismo no humano que comprende el vector de expresión de una cualquiera de las reivindicaciones 1 o 3 a 12.
15. Un kit que comprende el vector de expresión de una de las reivindicaciones 1 o 3 a 12.

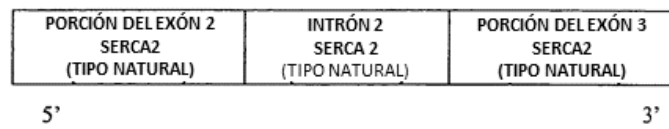


FIGURA 1A



FIGURA 1B

SEQ ID NO:2

cgaagaagggtgagtaatcctaacaatgctcttttttttttttgcctaatccctttgtgtgctgatgtaggatgacatttac
aacaatgtttgttcctgacaggaacaccttgctg

SEQ ID NO:4

cgaagaagggtgagtaatcctaacaatgctcttttttttttttgcctaatccctttgtgtgctgatgtaggatgacatttac
aacaatgtttgttcctgacaggaacaccttgctg

SEQ ID NO:5

agttaccggctgaagaaggtaatcctaacaatgctgtttctgttttttttccctctgttggtgtgctgatggtaagatgacaat
taaaacacatgtgtttgttttttacaggaaaaaccttgctggaacttgtagtgagcagttgaagacttgctagtagg
atattactggcagcatgtatatcttt

SEQ ID NO:6

aattgccggctgaagaaggtaataatattaacatgttatgttgagagatgatgtgtgcaggctgattgatgtggaca
attgaacaatgtgtttgttttttttttttttcccttcctttcctaacaggaaaaaccttgctggaacttgtagtgagc
agttgaagacttactagttagaatttactgctggcagcatgtatatctttc

FIGURA 1C

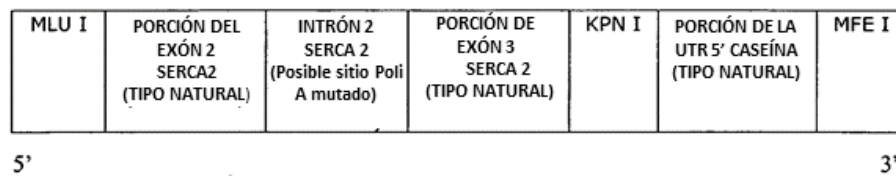


FIGURA 2A

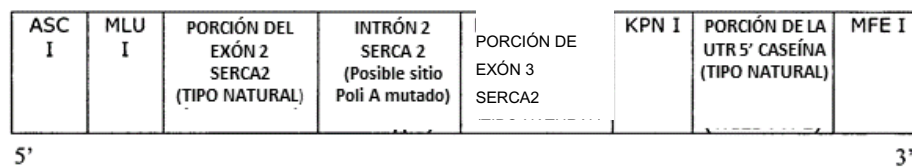


FIGURA 2B

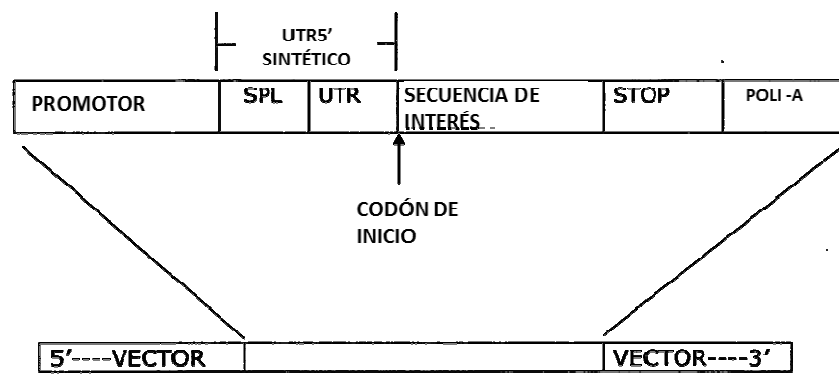


FIGURA 3A

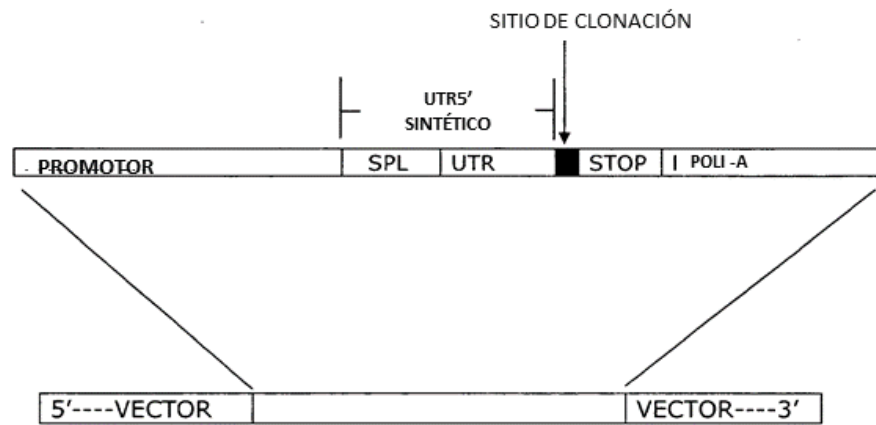


FIGURA 3B

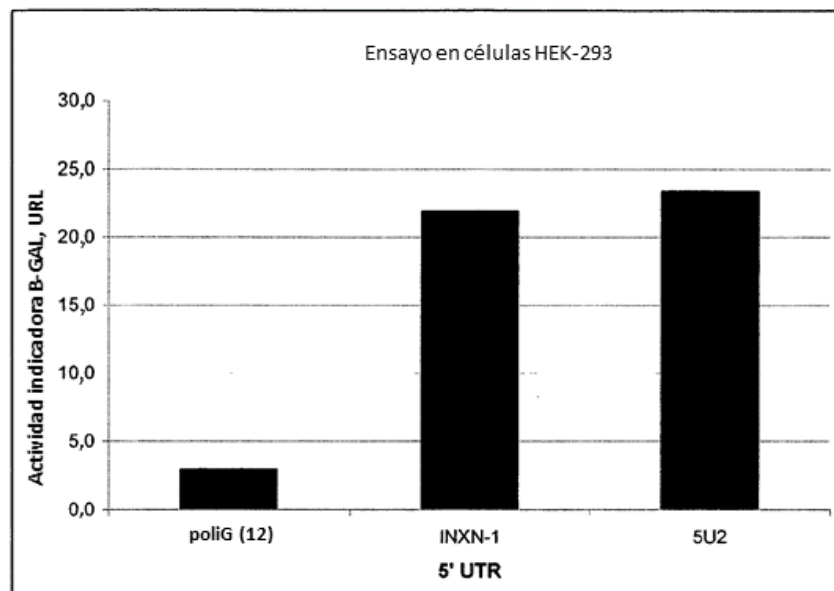


FIGURA 4

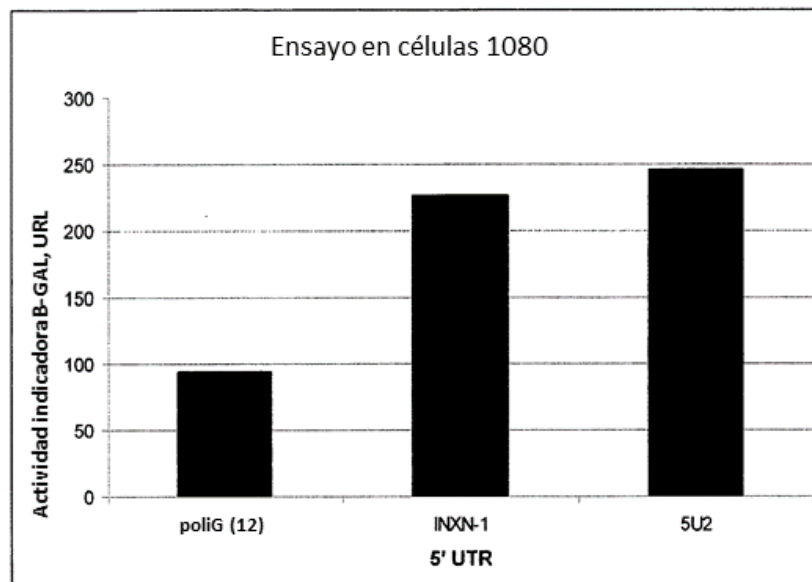


FIGURA 5

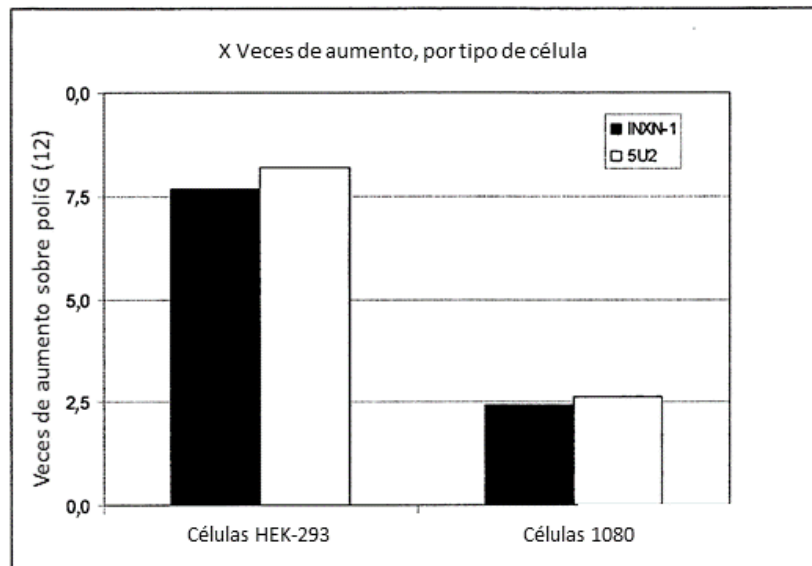


FIGURA 6

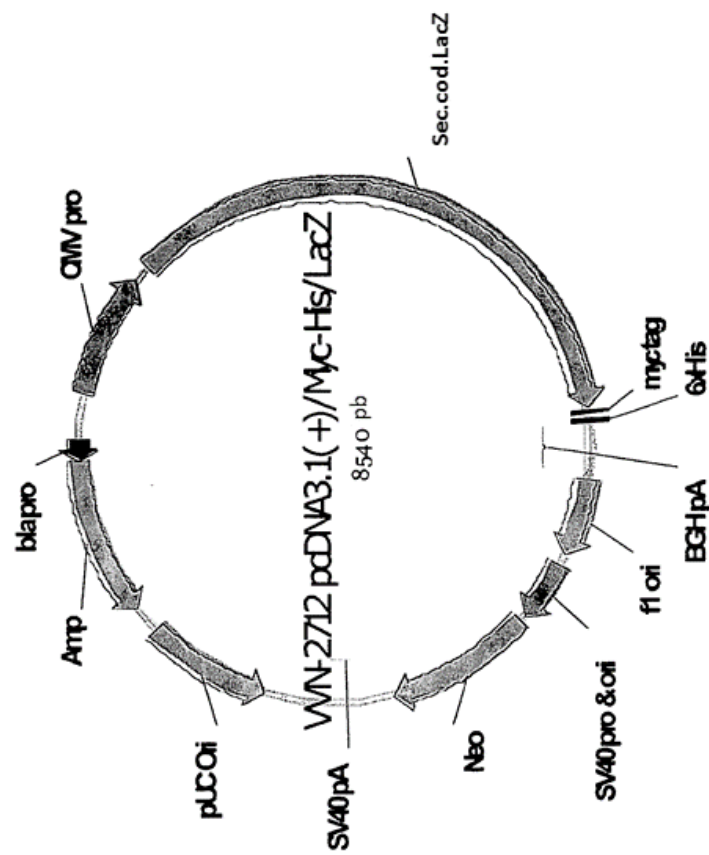


FIGURA 7

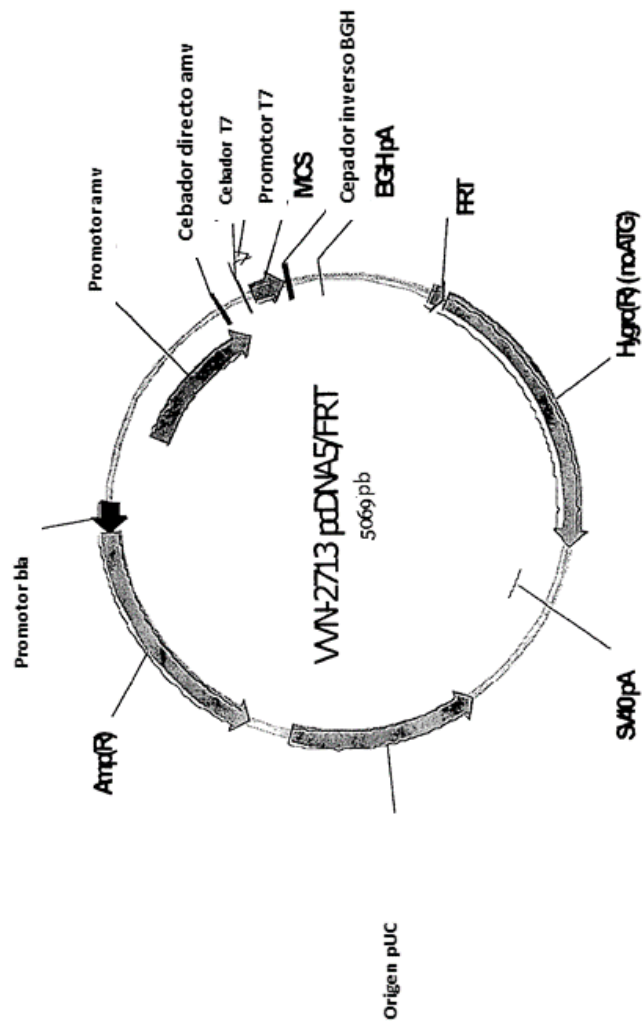


FIGURA 8

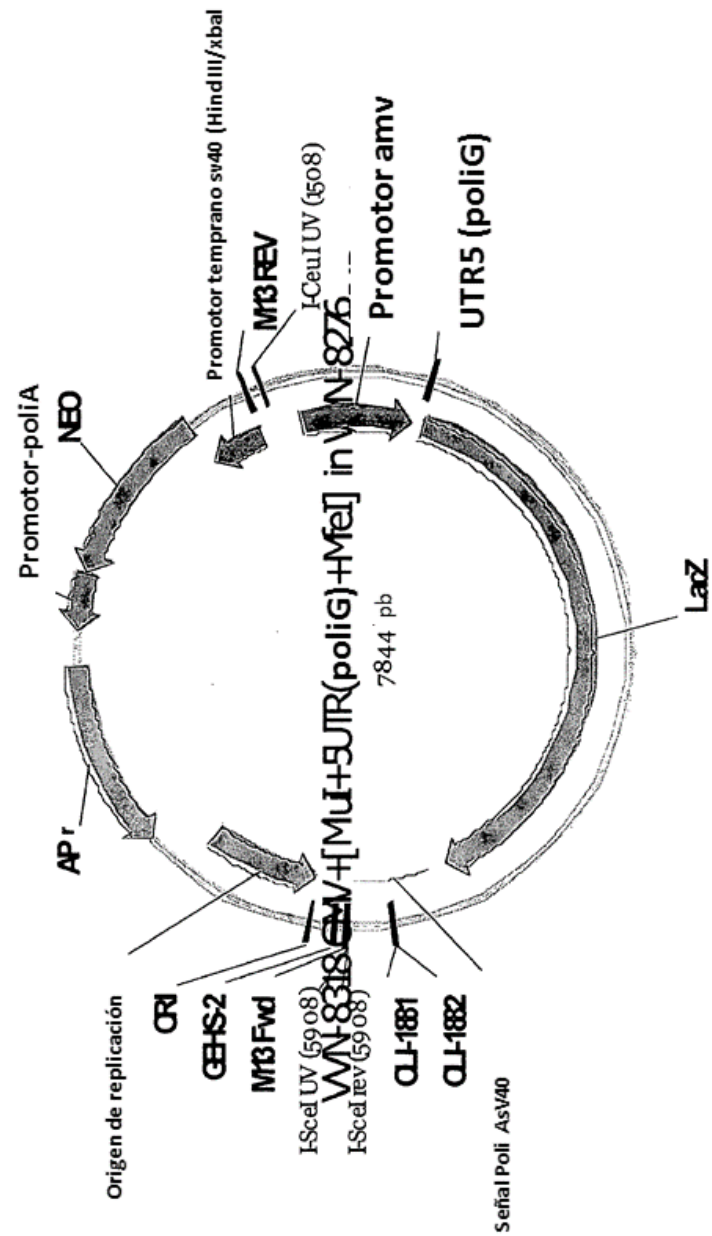


FIGURA 9

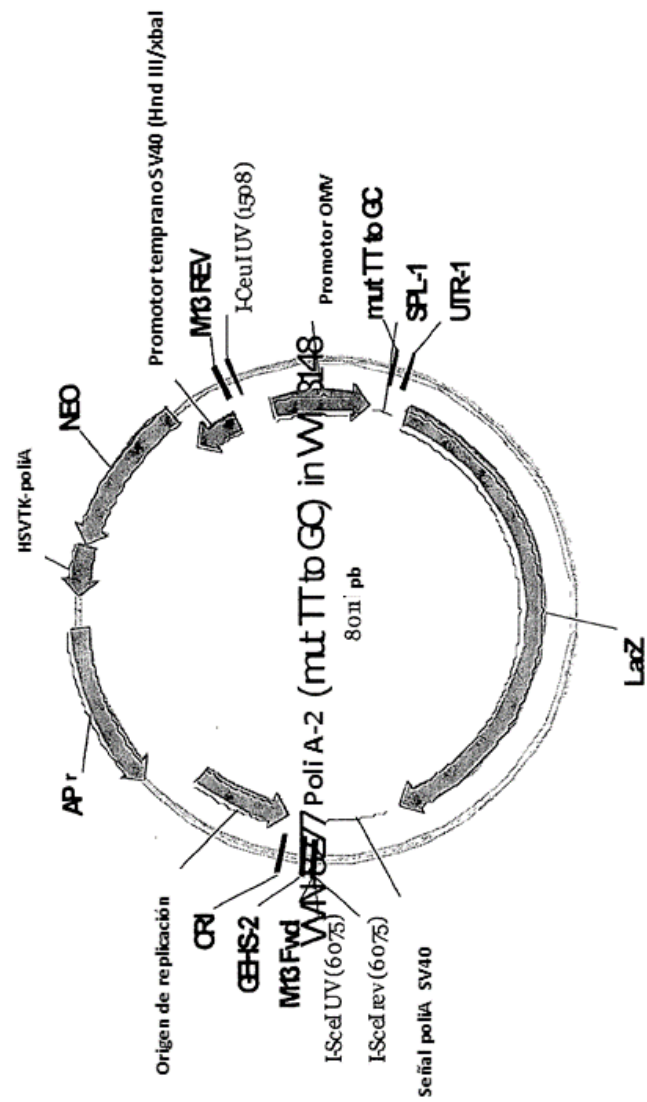


FIGURA 10

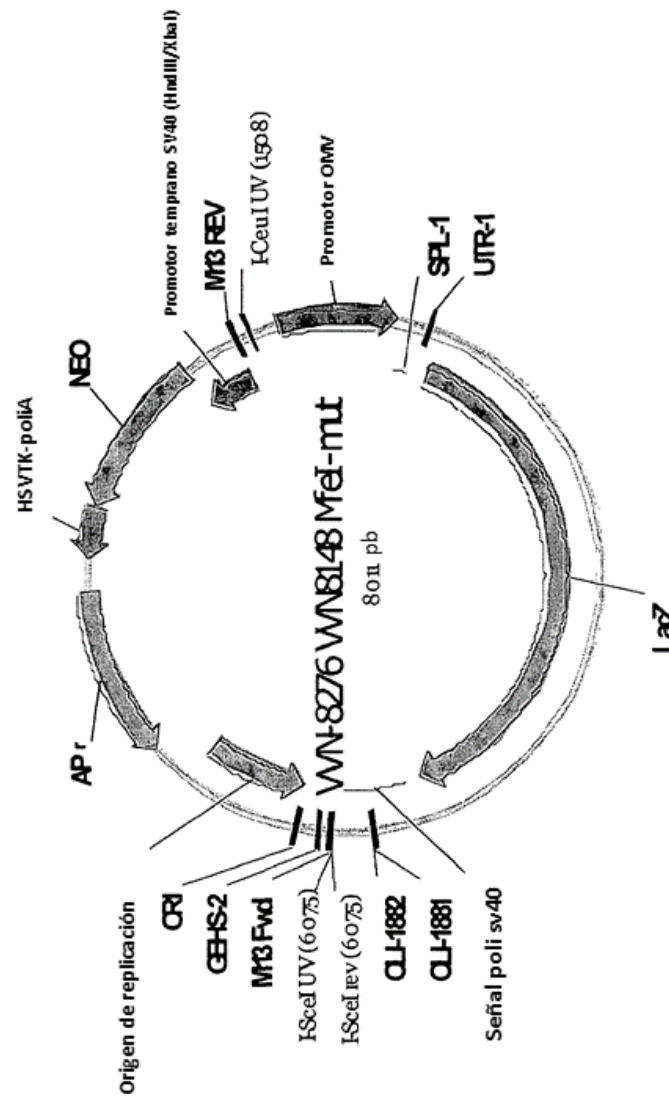


FIGURA 11

Ensayo en células HEK293									
N.º ID WN	N.º ID VEC	Tamaño (pb)	Conc (µg/ul)	Unidades relativas de luz (URL)				RLU PRO	D.E.
								RLU Normalizada	5' UTR
WN-8276	VEC-3159	8001 bp	0,5	39,24	32,802	34,227	32,15	34,6038	3,2106
WN-8318	VEC-3295	7844 bp	0,5	4,5651	4,9187	4,0627	4,731	4,5695	0,3674
WN-8277	VEC-3164	8011 bp	0,5	39,77	36,636	31,986	39,4	36,9478	3,5912
WN-2712 (Pos)	VEC-3386	8540 bp	0,5	4,1371	4,8691	4,8263	4,415	4,5619	0,3494
WN-2713 (Neg)	VEC-3387	5069 bp	0,5	0,0102	0,0157	0,0195	0,014	0,0149	0,0038

Ensayo en células 1080									
N.º ID WN	N.º ID VEC	Tamaño (pb)	Conc (µg/ul)	Unidades relativas de luz (URL)				RLU PRO	D.E.
								RLU Normalizada	5' UTR
WN-8276	VEC-3159	8001 bp	0,5	374,29	325,22	343,81	388,3	357,9040	28,6318
WN-8318	VEC-3295	7844 bp	0,5	140,92	148,52	142,94	149,3	145,4223	4,1338
WN-8277	VEC-3164	8011 bp	0,5	383,31	394,92	369,89	406,7	388,7005	15,7617
WN-2712 (Pos)	VEC-3386	8540 bp	0,5	89,46	80,038	72,098	74,14	78,9333	7,7837
WN-2713 (Neg)	VEC-3387	5069 bp	0,5	0,1789	0,2671	0,1681	0,167	0,1952	0,0482

FIGURA 12

```

281 -----ag 282
1251 gtctagggtgtctttcaagacagcagatgatttttaggaattcttttttagaagatagagtgttaacgtgtttttcccccctctggacacgttgccctggcggaattttacatcctcgagag 1375
1260 . 1280 . 1300 . 1320 . 1340 . 1360 . 320
283 ttaccggctgaaga-----agggaaaaaccttgctggaaccttggtgattgag 327
|||||
1376 ttaccggctgaagaaggtgagcgaatcgaacgtgtttctcccgcttttgctgctgatggttaagtagacgttttacacacagatgtttgttttaacaggggaaaaaccttgctggaaccttggtgattgag 1500
1380 . 1400 . 1420 . 1440 . 1460 . 1480 . 1500
. . . . .
328 cagtttgaagacttactagttagaattttattgctggcagcagtgtatatcttt----- 381
|||||
1501 cagtttgaagacttactagttagaattttattgctggcagcagtgtatatcttttgaagtataaagaattttattttgtctccaaaaagttgggacggttccatagatgaanaagcggggaagta 1625
1520 . 1540 . 1560 . 1580 . 1600 . 1620

```

FIGURA 13