

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 522 615**

51 Int. Cl.:

C12N 15/85 (2006.01)

C12N 15/86 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **04.06.2008 E 08756678 (2)**

97 Fecha y número de publicación de la concesión europea: **22.10.2014 EP 2150617**

54 Título: **Regiones de expresión y estabilidad potenciadas**

30 Prioridad:

04.06.2007 US 933213 P

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

17.11.2014

73 Titular/es:

REGENERON PHARMACEUTICALS, INC.

(100.0%)

**777 OLD SAW MILL ROAD
TARRYTOWN, NY 10591, US**

72 Inventor/es:

**CHEN, GANG;
BABB, ROBERT y
FANDL, JAMES P.**

74 Agente/Representante:

UNGRÍA LÓPEZ, Javier

ES 2 522 615 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

Regiones de expresión y estabilidad potenciadas

5 **Antecedentes****Campo de la invención**

10 La invención proporciona la expresión de proteínas recombinante en células eucariotas. En particular, la invención se refiere a métodos para la expresión mejorada de proteínas en células eucariotas empleando secuencias de nucleótidos que potencian la expresión. La invención se refiere a secuencias de región de expresión y estabilidad potenciadas (EESYR) que facilitan la expresión potenciada y estable de proteínas recombinantes en células eucariotas, y métodos para usar dichas secuencias.

15 **Descripción de la técnica relacionada**

El desarrollo de sistemas de expresión es un objetivo importante para proporcionar una fuente fiable y eficaz de una proteína determinada para uso de investigación y terapéutico. Con frecuencia se prefiere la expresión de proteínas recombinantes en células de mamífero para fabricar proteínas terapéuticas debido, por ejemplo, a la capacidad de los sistemas de expresión de mamíferos de modificar apropiadamente de forma post-traduccionales las proteínas recombinantes.

25 Se dispone de diversos vectores para su expresión en hospedadores de mamífero, conteniendo cada uno diversas combinaciones de elementos reguladores en *cis* y, en algunos casos, en *trans* para conseguir niveles elevados de proteína recombinante con cortos tiempos de incubación. A pesar de la disponibilidad de muchos de estos vectores, a menudo la expresión de una proteína recombinante realizada en sistemas de mamífero es inaceptablemente baja o, de otro modo, insatisfactoria. Además, desarrollar una línea celular que exprese de forma fiable niveles suficientemente elevados de una proteína deseada a menudo requiere etapas de clonación y amplificación que consumen mucho tiempo. Por consiguiente, existe la necesidad en la técnica de sistemas de expresión mejorados en mamífero.

Schmerling Doron et al. describen la expresión fuerte y ubicua de transgenes dirigidos en el locus de beta-actina por remplazo del casete cre/lox (Journal of Genetics and Development, vol.42, pág. 229 a 235 (2005)).

35 **Breve resumen**

La invención proporciona una célula aislada que comprende una secuencia potenciadora de la expresión seleccionada entre las SEC ID N° 1, 3 ó 5, y al menos un sitio de reconocimiento de recombinasa localizado dentro de una secuencia de nucleótidos de la SEC ID N° 1, 3 ó 5. La invención también proporciona una célula CHO que comprende una secuencia potenciadora de la expresión que es los nucleótidos 6.400-6.500 de la SEC ID N° 5, que comprende adicionalmente al menos un sitio de reconocimiento de recombinasa flanqueado 5' o 3' por un gen humano de interés (GOI), opcionalmente donde el GOI humano está flanqueado 5' por el primer sitio de reconocimiento de recombinasa, y dicha célula CHO comprende adicionalmente un segundo sitio de reconocimiento de recombinasa que flanquea 3' el GOI humano. La invención proporciona adicionalmente una célula aislada que comprende una secuencia potenciadora de la expresión seleccionada entre las SEC ID N° 1, 3 ó 5 y al menos un GOI añadido de forma exógena dentro de la secuencia potenciadora de la expresión.

La invención proporciona adicionalmente un método para preparar una proteína de interés, que comprende: (a) proporcionar una célula hospedadora que comprende una secuencia potenciadora de la expresión seleccionada entre las SEC ID N° 1, 3 ó 5, o los nucleótidos 6.400-6.500 de la SEC ID N° 5, (b) introducir en la célula hospedadora, dentro de la secuencia potenciadora de la expresión un gen de interés (GOI) unido de forma funcional a un promotor; (c) mantener la célula hospedadora de (a) en condiciones que permitan al GOI expresar una proteína de interés; y (d) recuperar la proteína de interés. La invención proporciona adicionalmente el uso de la célula aislada de la invención o la célula CHO aislada de la invención para preparar un anticuerpo o fragmento del mismo.

También se describe en este documento una secuencia de nucleótidos aislada que comprende una secuencia potenciadora de la expresión seleccionada entre una secuencia de las SEC ID N° 1-6, o un fragmento potenciador de la expresión de las mismas.

60 Las secuencias potenciadoras de la expresión descritas en este documento comprenden una secuencia potenciadora de la expresión de la SEC ID N° 5 localizada en una posición dentro de la SEC ID N° 5 seleccionada entre los nucleótidos que abarcan las posiciones numeradas 10-13.515; 20-12.020; 1.020-11.020; 2.020-10.020; 3.020-9.020; 4.020-8.020; 5.020-7.020; 6.020-6.920; 6.120-6.820; 6.220-6.720; 6.320-6.620; 6.420-6.520; 6.460-6.500; 6.470-6.490; y 6.475-6.485. Otras secuencias potenciadoras de la expresión descritas en este documento se seleccionan entre el grupo que consiste en los nucleótidos 5.000-7.400 de la SEC ID N° 5; 5.000-6.500 de la SEC ID N° 5; 6.400-7.400 de la SEC ID N° 5; y los nucleótidos 6.400-6.500 de la SEC ID N° 5.

El sitio de reconocimiento de recombinación puede posicionarse como se ha descrito anteriormente, con la condición de que la secuencia potenciadora de la expresión comprenda una secuencia que es al menos un 90% idéntica, más preferiblemente al menos un 95% idéntica, mucho más preferiblemente al menos un 99% idéntica, a la secuencia potenciadora de la expresión de la SEC ID N° 5 o un fragmento potenciador de la expresión de la misma.

5 Las secuencias potenciadoras de la expresión puede comprender adicionalmente al menos un sitio de reconocimiento de recombinasa que comprende una secuencia independientemente seleccionada entre un sitio loxp, un sitio lox 511, un sitio lox 2272, y un sitio frt. El sitio de reconocimiento de recombinasa puede estar dentro de la secuencia potenciadora de la expresión. El sitio de reconocimiento de recombinasa puede estar inmediatamente adyacente en la dirección 5' al nucleótido terminal del extremo 5' o inmediatamente adyacente en la dirección 3' al nucleótido terminal del extremo 3' de la secuencia potenciadora de la expresión.

15 Pueden estar presentes al menos dos sitios de reconocimiento de recombinasa dentro de la secuencia potenciadora de la expresión. Pueden estar presentes dos sitios de reconocimiento de recombinasa de orientación opuesta dentro de la secuencia potenciadora de la expresión. Pueden estar presentes tres sitios de reconocimiento de recombinasa dentro de la secuencia potenciadora de la expresión.

20 En este documento también se describe una secuencia de nucleótidos aislada que comprende una secuencia potenciadora de la expresión que es al menos un 80% idéntica, preferiblemente al menos un 90% idéntica, más preferiblemente al menos un 95% idéntica, mucho más preferiblemente al menos un 99% idéntica a una secuencia potenciadora de la expresión de las SEC ID N° 1-6 o un fragmento potenciador de la expresión de las mismas. La secuencia potenciadora de la expresión puede presentar la identidad indicada a una secuencia de la SEC ID N° 5 como se ha descrito anteriormente.

25 En este documento se describe una célula eucariota aislada que comprende una secuencia potenciadora de la expresión seleccionada entre las SEC ID N° 1-6 o un fragmento potenciador de la expresión de las mismas. La secuencia potenciadora de la expresión puede comprender una secuencia potenciadora de la expresión de la SEC ID N° 5 como se ha descrito anteriormente.

30 La célula eucariota puede ser una célula de ratón, rata, hámster, o humana. La célula eucariota puede ser una célula CHO.

35 La célula eucariota puede comprender adicionalmente al menos una secuencia de reconocimiento de recombinasa dentro de la secuencia potenciadora de la expresión. La al menos una secuencia de reconocimiento de recombinasa puede seleccionarse independientemente entre un sitio loxp, un sitio lox 511, un sitio lox 2272, y un sitio frt. El sitio de reconocimiento de recombinasa puede estar inmediatamente adyacente en la dirección 5' al nucleótido terminal del extremo 5' o inmediatamente adyacente en la dirección 3' al nucleótido terminal del extremo 3' de la secuencia potenciadora de la expresión.

40 Pueden estar presentes al menos dos sitios de reconocimiento de recombinasa dentro de la secuencia potenciadora de la expresión. Pueden estar presentes dos sitios de reconocimiento de recombinasa de orientación opuesta dentro de la secuencia potenciadora de la expresión. Pueden estar presentes tres secuencias de reconocimiento de recombinasa, y una de las tres secuencias de reconocimiento de recombinasa está en una orientación opuesta a las dos secuencias de reconocimiento de recombinasa restantes.

45 Como se ha descrito en este documento, el sitio de reconocimiento de recombinasa en la secuencia potenciadora de la expresión de la SEC ID N° 5 puede estar localizado en una posición dentro de la SEC ID N° 5 seleccionado entre los nucleótidos que abarcan las posiciones numeradas 10-13.515; 20-12.020; 1.020-11.020; 2.020-10.020; 3.020-9.020; 4.020-8.020; 5.020-7.020; 6.020-6.920; 6.120-6.820; 6.220-6.720; 6.320-6.620; 6.420-6.520; 6.460-6.500; 50 6.470-6.490; y 6.475-6.485. El sitio de reconocimiento de recombinasa puede estar en una secuencia que se selecciona entre el grupo que consiste en los nucleótidos 5.000-7.400 de la SEC ID N° 5; 5.000-6.500 de la SEC ID N° 5; 6.400-7.400 de la SEC ID N° 5; y los nucleótidos 6.400-6.500 de la SEC ID N° 5. El sitio de reconocimiento de recombinasa puede estar localizado dentro de los nucleótidos 6400-6500 de la SEC ID N° 5. El sitio de reconocimiento de recombinasa puede insertarse antes, después, o dentro del triplete "act" de los nucleótidos 6471 a 6473 de la SEC ID N° 5 en una secuencia potenciadora de la expresión de la SEC ID N° 5.

55 El sitio de reconocimiento de recombinación puede posicionarse como se ha descrito anteriormente, con la salvedad de que la secuencia potenciadora de la expresión comprende una secuencia que es al menos un 90% idéntica, más preferiblemente al menos un 95% idéntica, mucho más preferiblemente al menos un 99% idéntica, a los nucleótidos 5218 a 6048 de la SEC ID N° 5 o un fragmento potenciador de la expresión de la misma.

60 La célula puede ser una célula CHO donde el sitio de reconocimiento de recombinasa se inserta en el genoma de la célula CHO en o dentro del triplete "act" de los nucleótidos 6.471 a 6.473 de la SEC ID N° 5.

65 Puede insertarse un primer GOI dentro de la secuencia potenciadora de la expresión de la SEC ID N° 5 como se ha descrito anteriormente, y el primer GOI opcionalmente se une de forma funcional a un promotor, donde el GOI ligado

al promotor (o el GOI) está flanqueado 5' por un primer sitio de reconocimiento de recombinasa y 3' por un segundo sitio de reconocimiento de recombinasa. Puede insertarse un segundo GOI 3' del segundo sitio de reconocimiento de recombinasa, y el segundo GOI flanqueado 3' por un tercer sitio de reconocimiento de recombinasa.

- 5 El GOI puede unirse de forma funcional a un promotor capaz de dirigir la expresión del GOI, donde el promotor comprende un promotor eucariota que se puede regular por un activador o inhibidor. El promotor eucariota puede unirse de forma funcional a un operador procariota, y la célula eucariota opcionalmente puede comprender adicionalmente una proteína represora procariota.
- 10 Puede incluirse uno o más marcadores de selección entre el primero y el segundo y/o el segundo y el tercero sitios de reconocimiento de recombinasa. El primero y/o el segundo genes de interés y/o el uno o más marcadores de selección pueden unirse de forma funcional a un promotor, donde el promotor puede el mismo o diferente. El promotor puede comprender un promotor eucariota (tal como, por ejemplo, un promotor CMV), opcionalmente controlado por un operador procariota (tal como, por ejemplo, un operador tet). La célula puede comprender adicionalmente un gen que codifica un represor procariota (tal como, por ejemplo, un represor tet).
- 15

La célula puede comprender adicionalmente un gen capaz de expresar una recombinasa. En una realización específica, la recombinasa es una recombinasa Cre.

- 20 También se describe una célula eucariota aislada en este documento que comprende una secuencia potenciadora de la expresión que es al menos un 80%, más preferiblemente al menos un 90%, más preferiblemente al menos un 95%, mucho más preferiblemente al menos un 99% idéntica a una secuencia potenciadora de la expresión de las SEC ID N° 1-6 o un fragmento potenciador de la expresión de las mismas. La secuencia potenciadora de la expresión puede ser una secuencia dentro de la SEC ID N° 5 como se ha descrito anteriormente.

- 25 Se describe una célula hospedadora eucariota en este documento que comprende una secuencia potenciadora de la expresión seleccionada entre las SEC ID N° 1-6 o una fragmento potenciador de la expresión de las mismas, que comprende un primer sitio de reconocimiento de recombinasa seguido de un primer promotor eucariota, un primer gen marcador, un segundo promotor eucariota, un segundo gen marcador, un segundo sitio de reconocimiento de recombinasa, un tercer promotor eucariota, un tercer gen marcador, y un tercer sitio de reconocimiento de recombinasa. La secuencia potenciadora de la expresión puede estar dentro de la SEC ID N° 5 como se ha descrito anteriormente.

- 30 El primero, segundo, y tercero sitios de reconocimiento de recombinasa pueden ser diferentes. Los sitios de reconocimiento de recombinasa puede seleccionarse entre un sitio loxp, un sitio lox 511, un sitio lox 2272, y un sitio frt.
- 35

- El primer gen marcador puede ser un gen de resistencia a fármaco. El gen de resistencia a fármaco puede ser un gen de resistencia a puomicina. El segundo y tercero genes marcadores puede codificar dos proteínas fluorescentes diferentes. Las dos proteínas fluorescentes diferentes pueden seleccionarse entre coral de *Discosoma* (DsRed), proteína fluorescente verde (GFP), proteína fluorescente verde potenciada (eGFP), proteína fluorescente ciano (CFP), proteína fluorescente ciano potenciada (eCFP), y proteína fluorescente amarilla (YFP). Las dos proteínas fluorescentes diferentes pueden ser eCFP y DsRed.
- 40

- 45 El primero, segundo, y tercero promotores pueden ser iguales. El primero, segundo, y tercero promotores pueden ser diferentes. El primer promotor puede ser diferente del segundo y tercero promotores, y el segundo y tercero promotores son iguales. El primer promotor puede ser un promotor tardío de SV40, y el segundo y tercero promotores son cada uno un promotor CMV humano.

- 50 También se describe una célula hospedadora eucariota en este documento, que comprende una secuencia potenciadora de la expresión seleccionada entre las SEC ID N° 1-6, o un fragmento potenciador de la expresión de las mismas, al menos un sitio de reconocimiento de recombinasa dentro de la secuencia potenciadora de la expresión, y al menos un gen de interés (GOI) dentro de la secuencia potenciadora de la expresión. La secuencia potenciadora de la expresión puede ser una secuencia dentro de la SEC ID N° 5, como se ha descrito anteriormente.

- 55 La célula descrita en este documento puede comprender un primer sitio de reconocimiento de recombinasa seguido de un primer promotor unido de forma funcional a un primer GOI. El primer GOI puede estar seguido de un segundo sitio de reconocimiento de recombinasa. El segundo sitio de reconocimiento de recombinasa puede estar seguido de un segundo promotor unido de forma funcional a un segundo GOI. El segundo GOI puede estar seguido de un tercer sitio de reconocimiento de recombinasa. Al menos un marcador puede unirse de forma funcional a un tercer promotor y localizarse entre el segundo sitio de reconocimiento de recombinasa y el segundo promotor. El primer sitio de reconocimiento de recombinasa puede orientarse en una orientación opuesta al segundo y tercero sitios de reconocimiento de recombinasa. El primero y segundo promotores pueden ser promotores eucariotas unidos de forma funcional a un operador procariota. El primero y segundo promotores pueden ser promotores CMV unidos de forma funcional a secuencias operadoras tet. La célula puede comprender adicionalmente un gen capaz de expresar un represor procariota. El represor procariota puede ser un represor tet. La célula puede comprender un gen capaz
- 60
- 65

de expresar una recombinasa Cre.

Un primero y un segundo gen marcador puede localizarse entre el segundo sitio de reconocimiento de recombinasa y el segundo promotor, y un IRES está entre el primero y segundo genes marcadores. El primer codón (ATG) del primer gen marcador puede estar inmediatamente 5' al segundo sitio de reconocimiento de recombinasa, y el segundo codón del primer gen marcador está inmediatamente 3' al segundo sitio de reconocimiento de recombinasa. El primer gen marcador puede contener un intrón y el segundo sitio de reconocimiento de recombinasa está localizado dentro del intrón de modo que la mitad amino-terminal del primer gen marcador y la mitad 5' del intrón están localizados 5' del segundo sitio de reconocimiento de recombinasa y la mitad 3' del intrón y la mitad carboxi-terminal del primer gen marcador están inmediatamente 3' al segundo sitio de reconocimiento de recombinasa.

El primero, segundo, y tercero sitios de reconocimiento de recombinasa pueden ser diferentes. Los sitios de reconocimiento de recombinasa pueden seleccionarse entre un sitio loxp, un sitio lox 511, un sitio lox 2272, y un sitio frt.

El primer gen marcador puede ser un gen de resistencia a fármaco. El gen de resistencia a fármaco puede ser un gen de resistencia a higromicina. El segundo gen marcador puede codificar una proteína fluorescente. La proteína fluorescente puede seleccionarse entre DsRed, GFP, eGFP, CFP, eCFP, e YFP.

El primero, segundo, y tercero promotores pueden ser iguales. El primero, segundo, y tercero promotores pueden ser diferentes. El tercer promotor puede ser diferente del primero y segundo promotores, y el primero y segundo promotores iguales. El tercer promotor puede ser un promotor tardío de SV40, y el primero y segundo promotores son cada uno un promotor CMV humano.

El primero y segundo promotores pueden unirse de forma funcional a un operador procariota. El operador puede ser un operador tet.

La línea celular hospedadora puede tener un gen añadido de forma exógena que codifica una recombinasa integrada en su genoma, unido de forma funcional a un promotor. La recombinasa puede ser recombinasa Cre. La célula hospedadora puede tener un gen que codifica una proteína reguladora integrada en su genoma, unido de forma funcional a un promotor. La proteína reguladora puede ser una proteína represora tet.

El primer GOI y el segundo GOI pueden codificar una cadena ligera, o fragmento de la misma, de un anticuerpo o una cadena pesada, o fragmento de la misma, de un anticuerpo. El primer GOI puede codificar una cadena ligera de un anticuerpo y el segundo GOI puede codificar una cadena pesada de un anticuerpo.

Se describe un método en este documento para preparar una proteína de interés, que comprende: (a) proporcionar una célula hospedadora que comprende una secuencia potenciadora de la expresión seleccionada entre una secuencia de las SEC ID N° 1-6; (b) introducir en la célula hospedadora, dentro de la secuencia potenciadora de la expresión, un gen de interés (GOI) unido de forma funcional a un promotor; (c) mantener la célula hospedadora de (a) en condiciones que permitan que el GOI exprese una proteína de interés; y, (c) recuperar la proteína de interés.

La célula puede ser una célula CHO y la secuencia de nucleótidos es una secuencia potenciadora de la expresión de la SEC ID N° 5 como se ha descrito anteriormente.

El GOI puede introducirse en la célula empleando un vector de direccionamiento para recombinación homóloga, donde el vector de direccionamiento comprende un brazo de homología 5' homólogo a una secuencia presente en al menos una de las SEC ID N° 1-6, un GOI, y un brazo de homología 3' homólogo a una secuencia presente en al menos una de las SEC ID N° 1-6. La construcción puede comprender adicionalmente dos, tres, cuatro, o cinco o más genes de interés. Uno o más de los genes de interés puede estar unido de forma funcional a un promotor.

El GOI puede introducirse empleando una tecnología de integrasa, por ejemplo, tecnología de integrasa que emplea sitios att tal como los sistemas de clonación Invitrogen's Gateway™ y Multisite Gateway™ que emplean recombinación del sitio att del bacteriófago lambda.

La secuencia potenciadora de la expresión puede comprender uno o más sitios de reconocimiento de recombinasa como se ha descrito anteriormente, e introducirse el GOI en la secuencia potenciadora de la expresión a través de la acción de una recombinasa que reconoce el sitio de reconocimiento de recombinasa.

La secuencia potenciadora de la expresión puede comprender dos sitios de reconocimiento de recombinasa.

La secuencia potenciadora de la expresión puede comprender un primero, un segundo, y un tercero sitios de reconocimiento de recombinasa. El primero, segundo, y tercero sitios de reconocimiento de recombinasa pueden ser diferentes. El primero, segundo, y tercero sitios de reconocimiento de recombinasa pueden no estar en la misma orientación. El primer sitio puede estar 5' al segundo sitio, y el segundo sitio 5' al tercer sitio. El segundo y tercer sitios pueden estar en orientación opuesta con respecto al primer sitio.

Se introducen un primer y segundo GOI en la secuencia potenciadora de la expresión. El primer GOI puede introducirse entre el primero y el segundo sitios de reconocimiento de recombinasa, y el segundo GOI puede introducirse entre el segundo y el tercero sitios de reconocimiento de recombinasa. Los sitios de reconocimiento de recombinasa pueden seleccionarse independientemente entre un sitio loxp, un sitio lox 511, un sitio lox 2272, y un sitio frt.

El primer GOI flanqueado por sitios de reconocimiento de recombinasa en un primer vector y el segundo GOI flanqueado por sitios de reconocimiento de recombinasa en un segundo vector pueden introducirse en una célula que comprende una secuencia potenciadora de la expresión que comprende tres sitios de reconocimiento de recombinasa en una única etapa.

El GOI puede unirse de forma funcional a un promotor eucariota. El promotor eucariota puede unirse de forma funcional a un operador procariota. El promotor eucariota puede ser un promotor CMV y el operador procariota un operador tet.

La célula puede comprender un gen capaz de expresar un represor procariota. El represor procariota puede ser un represor tet.

La célula puede comprender un gen capaz de expresar una recombinasa. La recombinasa puede ser una recombinasa Cre.

También se describe una célula eucariota en este documento, donde la célula eucariota comprende al menos una secuencia potenciadora de la expresión de la SEC ID N° 1-6 y al menos un gen añadido de forma exógena dentro de la secuencia potenciadora de la expresión. El gen añadido de forma exógena puede unirse de forma funcional a un promotor añadido de forma exógena.

La secuencia potenciadora de la expresión puede ser una secuencia de la SEC ID N° 5, y el al menos un gen añadido de forma exógena integrado o insertado dentro de la secuencia potenciadora de la expresión es un gen humano. La célula eucariota puede ser una célula CHO, el gen añadido de forma exógena es un gen humano, y el gen humano se une de forma funcional a un promotor añadido de forma exógena eucariota.

También se describe un vector de direccionamiento para recombinación homóloga en este documento, donde el vector de direccionamiento comprende un brazo de homología 5', un GOI, y un brazo de homología 3', donde cada brazo de homología es homólogo a una secuencia dentro de una de las SEC ID N° 1-6. Los brazos de homología 5' y 3' pueden ser homólogos a secuencias dentro de la SEC ID N° 5. El vector de direccionamiento puede comprender dos, tres, cuatro, o cinco o más genes de interés. El GOI puede unirse de forma funcional a un promotor. Cada uno de los dos, tres, cuatro, o cinco o más genes de interés puede unirse cada uno de forma funcional a un promotor.

Adicionalmente se describe un vector de expresión en este documento, que comprende una secuencia de nucleótidos potenciadora de la expresión seleccionada entre las SEC ID N° 1-6 o un fragmento potenciador de la expresión de las mismas. La secuencia potenciadora de la expresión puede estar dentro de la SEC ID N° 5 como se ha descrito anteriormente.

El vector puede comprender adicionalmente un promotor. El promotor puede ser un promotor CMV humano.

El vector puede comprender adicionalmente un sitio de clonación para un gen de interés (GOI) expresable. La secuencia de nucleótidos seleccionada de la SEC ID N° 1-6 o fragmento potenciador de la expresión de las mismas puede localizarse 3' con respecto al sitio de clonación para el GOI expresable. La secuencia de nucleótidos seleccionada entre las SEC ID N° 1-6 o fragmento potenciador de la expresión de las mismas puede localizarse 5' con respecto al sitio de clonación para el GOI expresable.

El vector puede comprender una secuencia de nucleótidos seleccionada entre las SEC ID N° 1-6 o un fragmento potenciador de la expresión, un promotor CMV humano, un GOI, y una secuencia de terminación, donde la secuencia de nucleótidos seleccionada entre las SEC ID N° 1-6, o el fragmento potenciador de la expresión de las mismas, está localizada 5' con respecto al promotor CMV. El vector puede comprender adicionalmente un intrón seleccionado entre un intrón CMV-MIE y un intrón de β -globina de conejo.

También se describe un vector de expresión en este documento, que comprende una secuencia de nucleótidos potenciadora de la expresión que es al menos un 80% idéntica, preferiblemente al menos un 90% idéntica, más preferiblemente al menos un 95% idéntica, mucho más preferiblemente al menos un 99% idéntica a una secuencia seleccionada entre las SEC ID N° 1-6 o un fragmento potenciador de la expresión de las mismas.

Se describe adicionalmente un método en este documento para preparar una proteína de interés, que comprende: (a) introducir en una célula hospedadora un vector de expresión que comprende una secuencia potenciadora de la expresión seleccionada entre una secuencia dentro de las SEC ID N° 1-6 y un GOI que codifica una proteína de interés, donde el GOI está unido de forma funcional a un promotor y unido de forma funcional a la secuencia

potenciadora de la expresión; (b) cultivar la célula hospedadora de (a) en condiciones que permitan la expresión del GOI; y (c) recuperar la proteína de interés.

5 La secuencia de región de expresión y estabilidad potenciadas puede ser una secuencia potenciadora de la expresión de la SEC ID N° 1-6. La secuencia potenciadora de la expresión puede estar dentro de la SEC ID N° 5 como se ha descrito anteriormente.

10 La proteína recombinante puede seleccionarse entre el grupo que consiste en una subunidad de una inmunoglobulina o fragmento de la misma y un receptor o fragmento de unión a ligando del mismo. La proteína recombinante puede seleccionarse entre el grupo que consiste en una cadena ligera de anticuerpo o fragmento específico de antígeno del mismo, y una cadena pesada de anticuerpo o fragmento específico de antígeno del mismo.

15 La secuencia potenciadora de la expresión puede colocarse en la orientación indicada como se indica en las SEC ID N° 1-6, o en la orientación inversa indicada en las SEC ID N° 1-6.

Cualquiera de los aspectos y realizaciones de la invención pueden usarse junto con cualquier otro aspecto o realización de la invención, salvo que se indique otra cosa o sea evidente a partir del contexto.

20 Llegarán a ser evidentes otros objetos y ventajas a partir de una revisión de la siguiente descripción detallada.

Breve descripción de las figuras

25 **Fig.1.** Diagrama esquemático de una construcción retroviral, pTE252, usada para la introducción de construcción de ácido nucleico en un genoma celular. LTR: repetición terminal larga; LoxP: secuencia de reconocimiento de recombinasa Cre: ATAACCTTCGTATAATGTATGCTATACGAAGTTGT (SEC ID N° 7); Lox511: una mutación de secuencias LoxP: ATAACCTTCGTATAATGTATACTATACGAAGTTAG (SEC ID N° 8); la secuencia Lox511 se reconoce por recombinasa Cre pero el sitio Lox511 no se recombina con un sitio LoxP. GFP: proteína fluorescente verde; CMV MIE: promotor temprano inmediato principal de citomegalovirus humano; npt: neomicina fosfotransferasa; bla: beta lactamaasa; IRES: sitio interno de entrada del ribosoma.

30 La **Fig. 2** ilustra una construcción plasmídica usada para identificar una EESYR.

Las **Fig. 3A y 3B** muestran una alineación de secuencias EESYR de ovario de hámster chino (CHO), ratón, ser humano, y rata para un fragmento de la SEC ID N° 5.

35 La **Fig. 4** muestra una alineación de la SEC ID N° 5 con secuencias de ratón, ser humano, y rata.

La **Fig. 5** ilustra que una EESYR, unida de forma funcional a un gen de interés (GOI), muestra expresión potenciada sobre un GOI que no está unido de forma funcional a una EESYR.

La **Fig. 6** muestra EESYR comparadas en su capacidad relativa de potenciar la expresión de un GOI unido de forma funcional.

40 La **Fig. 7** ilustra características clonales de células con respecto a la funcionalidad de EESYR.

La **Fig. 8** ilustra que células EESYR experimentan recombinación específica y eficaz.

La **Fig. 9** ilustra la rareza de eventos de integración aleatoria en células EESYR.

La **Fig. 10** ilustra el ensayo de elementos de acción en cis que emplean secuencias EESYR.

La **Fig. 11** ilustra el ensayo de promotores, intrones, y UTR usando un sistema EESYR.

45 La **Fig. 12** ilustra la optimización de la expresión de proteínas usando un sistema EESYR.

Descripción detallada

50 Antes de describir los presentes métodos, debe entenderse que esta invención no está limitada a métodos particulares, y condiciones experimentales descritas, ya que dichos métodos y condiciones pueden variar. También debe entenderse que la terminología usada en este documento es con el propósito de describir realizaciones particulares solamente, y no pretende ser limitante, ya que el alcance de la presente invención se limitará solamente por las reivindicaciones adjuntas.

55 Como se usa en esta memoria descriptiva y las reivindicaciones adjuntas, las formas singulares "un", "una", y "el", "la" incluyen referencias plurales salvo que el contexto indique claramente otra cosa. Por tanto, por ejemplo, una referencia a "un método" incluye uno o más métodos, y/o etapas del tipo descrito en este documento y/o que llegarán a ser evidentes para los especialistas en la técnica tras leer esta descripción.

60 Salvo que se defina de otro modo, o se especifique de otro modo, todos los términos técnicos y científicos usados en este documento tienen el mismo significado que el comprendido habitualmente por un especialista en la técnica a la que pertenece esta invención.

Aunque puede usarse cualquier método y materiales similares o equivalentes a los descritos en este documento en la práctica o ensayo de la presente invención, ahora se describen métodos y materiales particulares.

65

Definiciones

Las regiones de ADN están unidas de forma funcional cuando están relacionadas funcionalmente entre sí. Por ejemplo, un promotor está unido de forma funcional a una secuencia codificante si el promotor es capaz de participar en la transcripción de la secuencia; un sitio de unión al ribosoma está unido de forma funcional a una secuencia codificante si está posicionado para permitir la traducción. Generalmente, unido de forma funcional puede incluir, pero no requiere, contigüidad. En el caso de secuencias tales como líderes de secreción, la contigüidad y colocación apropiada en una fase de lectura son características típicas. Una EESYR está unida de forma funcional a un GOI cuando está relacionada funcionalmente con gen de interés, por ejemplo, donde su presencia provoca la expresión potenciada del GOI.

El término "potenciado" cuando se usa para describir expresión potenciada incluye una potenciación de al menos aproximadamente 1,5 veces a al menos aproximadamente 2 veces de potenciación en expresión sobre lo que se observa típicamente por integración aleatoria en un genoma, por ejemplo, en comparación con una combinación de integrantes aleatorios de una única copia de la misma construcción de expresión. La potenciación factorial de la expresión observada empleando las secuencias de la invención es en comparación con un nivel de expresión del mismo gen, medida en sustancialmente las mismas condiciones, en ausencia de una secuencia de la invención. Como se usa en este documento, la expresión "potenciación de la expresión" se usa de forma intercambiable con "expresión y estabilidad potenciadas" cuando se refiere a una región o secuencia. Una "región de expresión y estabilidad potenciadas", también mencionada en este documento como una "EESYR", es una región o secuencia que muestra recombinación más eficaz, estabilidad del inserto, y expresión de mayor nivel que lo observado típicamente por integración aleatoria en un genoma.

Eficacia potenciada de recombinación incluye una potenciación de la capacidad de un locus de recombinar (por ejemplo, empleando sitios de reconocimiento de recombinasa). Potenciación se refiere a una eficacia sobre la recombinación aleatoria, que es típicamente del 0,1%. Una eficacia de recombinación potenciada preferida es aproximadamente 10 veces sobre la aleatoria, o de aproximadamente el 1%. Salvo que se especifique, la invención reivindicada no está limitada a una eficacia específica de recombinación.

Cuando se emplea la expresión "gen añadido de forma exógena" o "GOI añadido de forma exógena" con referencia a una EESYR, la expresión se refiere a cualquier gen no presente dentro de la EESYR del modo en que la EESYR se encuentra en la naturaleza. Por ejemplo, un "gen añadido de forma exógena" dentro de una CHO EESYR (por ejemplo, una EESYR que comprende la secuencia de la SEC ID N° 5), puede ser un gen de hámster no encontrado dentro de la CHO EESYR en la naturaleza (es decir, un gen de hámster de otro locus en el genoma de hámster), un gen de cualquier otra especie (por ejemplo, un gen humano), un gen quimérico (por ejemplo, humano/de ratón), o cualquier otro gen no encontrado en la naturaleza existiendo dentro de la CHO EESYR.

Se entiende que el porcentaje de identidad, cuando se describe una EESYR, incluye secuencias homólogas que presentan la identidad indicada a lo largo de regiones de homología contigua, pero la presencia de huecos, deleciones, o inserciones que no tienen homología en la secuencia comparada no se toman en cuenta al calcular el porcentaje de identidad. En la explicación del uso de "porcentaje de identidad" en este contexto, se hará referencia a la siguiente comparación de secuencia:

EESYR	5595	AGATTCTGTGGGCTCTGAGGCAACTTGACCTCAGCCAGATGGTATTTGAATAACCTGCTC	5654
Ch3 de rata	5619	AGATTCTGTGGGTTCTGAGACAACCTTGACTTCAGCCAGATGGCATTTTGAATAAC-----	5672
Ch2 de M Mus	6618	AGATTCTGTGGGTTCTGAGACAACCTTGACTTTAGCCAGATGGTATTTGAGTAATCTGGG-	6676
Ch20 de H. Sap.	6620	AGATTCAGTGGGCTTTGGGACAGCTTGAGTTCAACTAGATGGTATTTGAATAATCTGCT-	6678
		*****	*****

Como se usa en este documento, una determinación de "porcentaje de identidad" entre la secuencia "EESYR" anterior (para una EESYR de célula CHO o fragmento de la misma) con un homólogo de rata ("Ch3 de rata") no incluiría una comparación de las secuencias de CHO de 5649 a 5654, ya que el homólogo de rata no tiene secuencia homóloga para comparar en una alineación (es decir, la CHO EESYR tiene una inserción en ese punto, o el homólogo de rata tiene un hueco o deleción, según pueda ser el caso). Por tanto, en la comparación anterior, la comparación del porcentaje de identidad se extendería desde el "AGATTC" en el extremo 5' hasta el "AATAAC" en el extremo 3'. En ese caso, el homólogo de rata difiere solamente en que tiene una "T" en la posición 5607 de CHO EESYR, una "A" en la posición 5614 de CHO EESYR, una "T" en la posición 524 de CHO EESYR, y una "C" en la posición 5637 de CHO EESYR. Como la comparación es sobre 54 bases contiguas en un tramo de 60 pares de bases, con solamente cuatro diferencias (que no son huecos, deleciones, o inserciones), existe más de un 90% de identidad entre las dos secuencias (CHO y rata) desde la posición 5595 de CHO EESYR hasta la posición 5654 de CHO EESYR (porque el "porcentaje de identidad" no incluye penalizaciones para huecos, deleciones, y inserciones).

Descripción general

La invención se basa al menos en parte en el descubrimiento de que hay secuencias en un genoma que muestran recombinación más eficaz, estabilidad del inserto, y expresión de mayor nivel que otras regiones o secuencias en el

genoma. La invención también se basa al menos en parte en el hallazgo de que cuando se identifican dichas secuencias potenciadoras de la expresión, puede añadirse de forma exógena un gen o construcción adecuada en o cerca de las secuencias y que el gen añadido de forma exógena puede expresarse ventajosamente. Dichas secuencias, llamadas regiones de expresión y estabilidad potenciadas ("EESYR"), pueden modificarse por ingeniería para que incluyan sitios de reconocimiento de recombinasa para la colocación de genes de interés para crear líneas celulares que sean capaces de expresar proteínas de interés. Las EESYR también pueden incluirse como en construcciones de expresión tales como, por ejemplo, vectores de expresión. Pueden usarse vectores de expresión que comprenden EESYR para expresar proteínas de forma transitoria, o pueden integrarse en un genoma por recombinación aleatoria o dirigida tal como, por ejemplo, recombinación homóloga o recombinación mediada por recombinasas que reconocen sitios específicos de recombinación (por ejemplo, recombinación mediada por Cre-lox). También pueden usarse vectores de expresión que comprenden EESYR para evaluar la eficacia de otras secuencias de ADN, por ejemplo, secuencias reguladoras de acción en *cis*.

La CHO EESYR descrita en detalle en este documento se identificó por integración aleatoria de ADN que comprende sitios lox en un genoma de célula CHO, seguido de selección para identificar secuencias donde estaba potenciada la expresión. La integración aleatoria e introducción del sitio lox se consiguió usando una construcción retroviral. La selección y exploración se consiguieron usando marcadores de resistencia a fármaco y marcadores de detectables (por ejemplo, proteínas fluorescentes con exploración FACS), empleando métodos de recombinación que usaban recombinación específica de sitio (por ejemplo, sitios lox y recombinasa Cre). La selección continuó hasta una expresión potenciada de 1,5 a 2 veces sobre la expresión observada cuando se integra aleatoriamente una construcción de expresión en el genoma de la célula CHO. Después de la identificación de la EESYR, se mantuvieron sitios de reconocimiento de recombinasa (en el ejemplo proporcionado, sitios lox) en la EESYR para introducir casetes de expresión que comprenden un GOI expresable, junto con cualquier otro elemento deseable tal como, por ejemplo, promotores, potenciadores, marcadores, operadores, *etc.*

Se muestra una ilustración de una construcción plasmídica usada en la identificación de una EESYR descrita en esta solicitud en la Fig. 2. La construcción de rescate plasmídica comprende un casete de expresión dirigido por un promotor, donde el casete está flanqueado en los extremos 5' y 3' con sitios de reconocimiento de recombinasa (representados por una esfera y una vara en la Fig. 2). Se muestra la inserción dentro de un locus EESYR, donde la inserción produce una construcción de rescate plasmídica que reemplaza un casete de expresión que comprende un promotor y un marcador, donde el casete de expresión dentro del locus EESYR está flanqueado en sus extremo 5' y 3' por sitios de reconocimiento de recombinasa (véase la Fig. 2).

Se describen composiciones y métodos en este documento para integrar de forma estable una secuencia de ácido nucleico en una célula eucariota, donde la secuencia de ácido nucleico tiene capacidad de expresión potenciada en virtud de integrarse en o cerca de una EESYR. Se describen células en este documento que contienen una secuencia de reconocimiento de recombinasa dentro o cerca de una EESYR, conveniente para la inserción de un GOI, para conseguir la expresión de una proteína de interés a partir del GOI. También se describen composiciones y métodos en este documento para usar EESYR en conexión con construcciones de expresión, por ejemplo, vectores de expresión, y para añadir una EESYR exógena en una célula eucariota de interés.

Caracterización física y funcional de una EESYR

Las secuencias de ácido nucleico mencionadas como EESYR se identificaron empíricamente por secuencias cadena arriba y cadena abajo del sitio de integración de una construcción de ácido nucleico (que comprende un casete de expresión) de una línea celular que expresa una proteína indicadora a un elevado nivel. Las secuencias de ácido nucleico EESYR descritas en este documento son secuencias con una nueva funcionalidad asociada con expresión potenciada de un ácido nucleico (por ejemplo, un ácido nucleico exógeno que comprende un GOI) que parece funcionar de forma diferente a lo descrito previamente para elementos de acción en *cis* tales como promotores, potenciadores, regiones de control de locus, regiones de unión a la estructura o regiones de unión a la matriz. Las EESYR no parecen tener ninguna fase de lectura abierta (ORF), haciendo improbable que las EESYR codifiquen nuevas proteínas activadoras en *trans*. Experimentos de transfección demostraron que las secuencias EESYR presentan algunas características de elementos de acción en *cis*. La actividad de EESYR no se detecta en ensayos de transfección transitoria; las secuencias EESYR también parecen ser distintas de los elementos promotores y potenciadores, que se detectan con estos métodos.

Aunque las secuencias EESYR descritas en detalle en este documento se aislaron del genoma de dos líneas celulares, las secuencias EESYR de estas dos líneas celulares son iguales. La actividad de EESYR se identificó en un fragmento de 6,472 kb de ADN genómico de CHO 5' con respecto a un sitio de integración único de un vector retroviral que comprenden una secuencia codificante de indicador DsRed y en un fragmento de 7,045 kb de ADN genómico de CHO 3' con respecto al sitio de integración. Los vectores de expresión que comprenden la región aislada de 6,472 kb y la región aislada de 7,045 kb y fragmentos más cortos de las mismas fueron capaces de conferir a las células CHO transfectadas con las mismas elevados niveles de expresión de proteínas recombinantes.

También se describen en este documento vectores de expresión que comprenden fragmentos EESYR orientados de forma inversa. Los fragmentos EESYR orientados de forma inversa también fueron capaces de conferir a células

CHO transfectadas con los mismos elevados niveles de expresión de proteínas recombinantes.

También pueden desarrollarse otras combinaciones de los fragmentos descritos en este documento. Ejemplos de otras combinaciones de los fragmentos descritos en este documento que también pueden desarrollarse incluyen secuencias que incluyen múltiples copias de la EESYR descrita en este documento, o secuencias derivadas de combinar la EESYR descrita con otras secuencias de nucleótidos para conseguir combinaciones óptimas de elementos reguladores. Dichas combinaciones pueden unirse de forma contigua o disponerse para proporcionar espaciado óptimo de los fragmentos EESYR (por ejemplo, mediante la introducción de nucleótidos espaciadores entre los fragmentos). Los elementos reguladores también pueden disponerse para proporcionar espaciado óptimo de una EESYR con respecto a los elementos reguladores.

Las secuencias de EESYR descritas en este documento se aislaron de células CHO. Se espera que existan elementos homólogos de potenciación de la expresión en células de otras especies de mamíferos (tales como, por ejemplo, seres humanos; véase la Fig. 3) así como en líneas celulares derivadas otros tipos tisulares, y pueden aislarse por técnicas que son bien conocidas en la técnica, por ejemplo, por hibridación de especies cruzadas o técnicas basadas en PCR. Además, pueden hacerse cambios en la secuencia de nucleótidos expuesta en las SEC ID N° 1-6 por técnicas de mutagénesis dirigida al sitio o aleatoria que son bien conocidas en la técnica. Las variantes EESYR resultantes después pueden ensatarse para la actividad de EESYR como se describe en este documento. Los ADN que son al menos aproximadamente un 80% idénticos, preferiblemente al menos aproximadamente un 90% idénticos, más preferiblemente al menos aproximadamente un 95% idénticos, mucho más preferiblemente al menos aproximadamente un 99% idénticos en secuencia de nucleótidos a las SEC ID N° 1-6 o fragmentos de las mismas que tienen actividad de EESYR se pueden aislar por experimentación rutinaria, y se espera que muestren actividad de EESYR. Para fragmentos de EESYR, el porcentaje de identidad se refiere a esa parte de la secuencia nativa de referencia que se encuentra en el fragmento de EESYR. Por consiguiente, también se describen homólogos de EESYR y variantes de EESYR en este documento. Las Fig. 3A y B muestran una alineación de secuencias de ratón, ser humano, y rata con homología variable respecto a un fragmento de la SEC ID N° 5.

Pueden desarrollarse poblaciones celulares que expresan niveles potenciados de una proteína de interés usando los métodos proporcionados en este documento. El nivel absoluto de expresión variará con la proteína específica, dependiendo de lo eficazmente que se procese la proteína por la célula. Combinaciones de células desarrolladas con EESYR son estables en el tiempo, y pueden tratarse como líneas celulares estables para la mayoría de los propósitos. Las etapas de clonación pueden retardarse hasta más tarde en el proceso de desarrollo que lo habitual para proteínas recombinantes.

EESYR y fragmentos de potenciación de la expresión de las mismas

El locus genómico EESYR está conservado entre los genomas humano, de ratón y de rata. La Fig. 4 muestra el porcentaje de identidad entre secuencias de EESYR. Se identificaron secuencias de EESYR, homólogas a las 13,515 kb del ADN de CHO EESYR clonado de la SEC ID N° 5, entre los genomas publicados humano, de rata y de ratón usando BLAST. Se alinearon las secuencias para determinar el porcentaje de homología usando MacVector (9.0). Se representaron en gráfico incrementos de veinticinco pb de la alineación como el porcentaje de identidad entre las secuencias de EESYR de CHO, humana, de ratón y de rata para cada segmento consecutivo de 25 pb. Como se muestra en la Fig. 4, la línea vertical marca la localización de eventos de recombinación específica de sitio para expresar genes de interés de proteína recombinante. El porcentaje de identidad de secuencias de EESYR adyacentes a una localización de recombinación específica de sitio en una EESYR se muestra en el panel B de la Fig. 4. Se representaron en gráfico incrementos de diez pares de bases de las secuencias alineadas correspondientes a los nt 5022-6110 de una secuencia de EESYR de célula CHO (nucleótidos 5022 a 6110 de la SEC ID N° 5) como el porcentaje de identidad entre secuencias de EESYR de CHO, humana, de ratón y de rata para cada segmento consecutivo de 10 pb. Las secuencias se alinearon usando MacVector™ 9.0. Como se muestra en el panel B, está presente una identidad significativa de secuencia en este fragmento de la EESYR clonada a partir de células CHO. Debe apreciarse que la comparación de la Fig. 4 indica una longitud de aproximadamente 1400 bases, mientras que la secuencia de la SEC ID N° 5 contiene 13.515 bases. Las bases de la Fig. 4 parecen extenderse sobre un tramo más largo debido a la existencia de huecos. Las extensiones de nucleótidos indicadas son aquellas correspondientes a la numeración de la SEC ID N° 5 salvo que se indique otra cosa. La extensión de nucleótidos de aproximadamente 6200 a aproximadamente 7600 como se muestra en la Fig. 4B corresponde a nucleótidos de la SEC ID N° 5 numerados de aproximadamente 5.200 a aproximadamente 6.000.

Por consiguiente, se describe en este documento un fragmento potenciador de la expresión de una secuencia de nucleótidos de la SEC ID N° 5, donde el fragmento potenciador de la expresión incluye la secuencia de nucleótidos indicada por las posiciones de aproximadamente el resto 5022 a aproximadamente 6110 de la SEC ID N° 5, o de aproximadamente 5218 a aproximadamente 6048 de la SEC ID N° 5; o de aproximadamente 6200 a aproximadamente 7600, de aproximadamente 6500 a aproximadamente 7400, o de aproximadamente 6400 a aproximadamente 6500 mostrados en el panel B de la Fig. 4. También se describe en este documento un fragmento potenciador de la expresión de una secuencia de nucleótidos que es al menos un 80% idéntica, preferiblemente al menos un 90% idéntica, más preferiblemente al menos un 95% idéntica, mucho más preferiblemente al menos un 99% idéntica a la secuencia de nucleótidos indicada por las posiciones aproximadamente 6200 a aproximadamente

7600, o aproximadamente 6500 a aproximadamente 7400, o aproximadamente 6400 a aproximadamente 6500 mostradas en el panel B de la Fig. 4. Se describen adicionalmente en este documento vectores que comprenden dicho fragmento, incluyendo para transfección transitoria o estable. Además se describe en este documento una célula eucariota que comprende dicho fragmento donde el fragmento es exógeno y se integra en el genoma celular, y células que comprenden dicho fragmento que tiene al menos un sitio de reconocimiento de recombinasa que está en, inmediatamente 5', o inmediatamente 3' al fragmento.

Un fragmento potenciador de la expresión de la SEC ID N° 5 puede localizarse en una posición dentro de la SEC ID N° 5 seleccionada entre los nucleótidos que abarcan las posiciones numeradas 10-13.515; 20-12.020; 1.020-11.020; 2.020-10.020; 3.020-9.020; 4.020-8.020; 5.020-7.020; 6.020-6.920; 6.120-6.820; 6.220-6.720; 6.320-6.620; 6.420-6.520; 6.460-6.500; 6.470-6.490; y 6.475-6.485.

En una realización, la EESYR se emplea para potenciar la expresión de un GOI, como se ilustra en la Fig. 5. La Fig. 5 muestra un GOI unido de forma funcional con un promotor (con un marcador cadena arriba que tiene su propio promotor) integrado en una posición no EESYR en el genoma de una célula CHO, y una lectura FACS que muestra la distribución de expresión en una población transfectada de forma estable de células. En comparación, se muestra un GOI unido de forma funcional a un promotor integrado en una posición EESYR en el genoma de una célula CHO, y también se muestra una lectura FACS que muestra la distribución de expresión en una población transfectada de forma estable de células. En esta realización, el GOI expresado dentro del locus EESYR muestra una expresión potenciada de aproximadamente dos veces en comparación con el GOI expresado en un locus no EESYR.

En diversas realizaciones, la expresión de un GOI puede potenciarse colocando el GOI dentro de una EESYR, 5' a una EESYR, o 3' a una EESYR. La distancia precisa entre el GOI y la EESYR, donde el GOI está 5' o 3' a la EESYR, debe ser tal que la EESYR esté unida de forma funcional al GOI. Una EESYR está unida de forma funcional al GOI cuando la expresión del GOI - a la distancia seleccionada desde la EESYR (en la dirección 5' o 3') - retiene la capacidad de potenciar la expresión del GOI sobre, por ejemplo, la expresión típicamente observada debida a un evento de integración aleatoria. En diversas realizaciones, la potenciación es al menos aproximadamente 1,5 veces a aproximadamente 2 veces o más. Preferiblemente, la potenciación en la expresión en comparación con una integración aleatoria, o expresión aleatoria, es aproximadamente 2 veces o más.

La Fig. 6 muestra una realización donde la SEC ID N° 1 ("EESYR 5'") y SEC ID N° 2 ("EESYR 3'") se comparan en su capacidad relativa de potenciar la expresión de un GOI unido de forma funcional, donde el GOI está unido de forma funcional a un promotor también (se muestran un marcador y un promotor unidos de forma funcional al gen marcador 5' al promotor del GOI). La orientación de la EESYR se muestra por la dirección de la flecha debajo el término "EESYR". Las construcciones se integran de forma aleatoria en el genoma de una célula CHO. La expresión es relativa a la construcción integrada de forma aleatoria que no comprenden ninguna EESYR. Las lecturas FACS que muestran la expresión relativa se muestran a la derecha. En la Fig. 6, la primera construcción EESYR emplea la SEC ID N° 1; la segunda construcción EESYR emplea la SEC ID N° 3; la tercera construcción EESYR es la SEC ID N° 2; la cuarta construcción EESYR es la SEC ID N° 4. Como se muestra en la figura, la SEC ID N° 3 presenta una potenciación de 2,4 veces en la expresión, y la SEC ID N° 1 presenta una potenciación de 2 veces de la expresión.

Combinaciones de células recombinantes EESYR presentan características clonales. La Fig. 7 ilustra las características clonales de combinaciones recombinantes EESYR. En el diagrama FACS de dos colores que representa un histograma de parámetros dual de células marcadas con marcadores rojos o verdes (las células rojas son células hospedadoras CHO; las células verdes son recombinantes que expresan un GOI), las combinaciones recombinantes EESYR muestran agrupamiento en el diagrama que refleja crecimiento y expresión, perfil de citometría de flujo, y análisis de Southern (no mostrado) sustancialmente idénticos. Un histograma de ELISA de células individuales demuestra expresión uniforme en todos los clones. La estabilidad clonal también es elevada después de tres meses sin selección.

Los recombinantes EESYR experimentan recombinación específica y eficaz, como se muestra en la Fig. 8. El panel A muestra dos marcadores separados por un IRES y flanqueados por sitios de reconocimiento de recombinasa, y un tercer marcador no flanqueado por sitios de reconocimiento de recombinasa como un control de integración aleatoria. Cuando se recombinan en un locus EESYR que comprende un marcador flanqueado por dos sitios de reconocimiento de recombinasa, la recombinación es específica. El panel B muestra poca integración aleatoria en ausencia de recombinasa, pero integración eficaz y específica de sitio en presencia de recombinasa.

La integración aleatoria usando recombinación específica de sitio en una EESYR es rara (véase la Fig. 9). La Fig. 9 muestra que cuando se visualizan eventos de integración aleatoria, dichos eventos representan solamente una pequeña fracción de eventos de integración.

Pueden evaluarse elementos de acción en *cis* de EESYR usando los métodos y composiciones descritos en este documento. Como se muestra en la Fig. 10, las células recombinantes EESYR, permiten la comparación e elementos de acción en *cis* de forma equivalente. Como los recombinantes EESYR se comportan como una población clonal, pueden compararse directamente las diferencias en la expresión génica como resultado de, por ejemplo, la presencia o ausencia de elementos sospechosos de acción en *cis*. Las líneas celulares isogénicas

5 permiten la comparación directa de elementos de acción en *cis*. Usando un sistema EESYR, los elementos de acción en *cis*, tales como, por ejemplo, promotores, intrones, y UTR, se localizan preferiblemente entre sitios de recombinación. La optimización de la expresión también puede conseguirse incluyendo, por ejemplo, orientación del casete de expresión y optimización de codones. A modo de ejemplo, la Fig. 11 muestra casetes flanqueados con sitios de reconocimiento de recombinación que contienen un promotor, un marcador, diversos elementos en *cis* (aquí, intrones en el panel A; UTR en panel el B), y un GOI que se integraron en una EESYR (SEC ID N° 1 en el extremo 5', SEC ID N° 2 en el extremo 3'). La expresión relativa del GOI se muestra a la derecha.

10 La Fig. 12 muestra un ejemplo del modo en que puede conseguirse la optimización de proteínas usando los métodos y composiciones descritos en este documento. La Fig. 12 confirma que la colocación óptima de un ADNc para un gen de anticuerpo de cadena ligera es 5' al ADNc para un gen de anticuerpo de cadena pesada.

Proteínas de interés

15 Una secuencia de ácido nucleico que codifica una proteína de interés puede integrarse convenientemente en una célula que comprende una EESYR que tiene un sitio de reconocimiento de recombinasa a través, por ejemplo, de un proceso de intercambio de casete mediado por recombinasa (RMCE). Puede usarse cualquier proteína de interés adecuada para su expresión en células eucariotas. Por ejemplo, la proteína de interés puede ser un anticuerpo o fragmento del mismo, un anticuerpo quimérico o fragmento del mismo, un ScFv o fragmento del mismo, una proteína marcada con Fc o fragmento de la misma, un factor de crecimiento o un fragmento del mismo, una citoquina o un fragmento de la misma, o un dominio extracelular de un receptor de superficie celular o fragmento del mismo.

Construcciones de ácido nucleico

25 Los vectores de expresión recombinantes pueden comprender fragmentos de ADN sintético o derivado de ADNc que codifican una proteína, unidos de forma funcional a un elemento adecuado regulador de la transcripción y/o la traducción derivado de genes de mamífero, virales o de insecto. Dichos elementos reguladores incluyen promotores de la transcripción, potenciadores, secuencias que codifican sitios adecuados de unión ribosómica a ARNm, y secuencias que controlan la terminación de la transcripción y la traducción, como se describe en detalle a continuación. Los vectores de expresión de mamífero también pueden comprender elementos no transcritos tales como un origen de replicación, otras secuencias no transcritas flanqueantes 5' o 3', y secuencias no traducidas 5' o 3' tales como sitios donantes y aceptores de corte y ajuste. También puede incorporarse un gen marcador de selección para facilitar el reconocimiento de transfectantes.

35 Las secuencias de control de la transcripción y la traducción en vectores de expresión útiles para transfectar células de vertebrado pueden proporcionarse por fuentes virales. Por ejemplo, promotores y potenciadores habitualmente usados se obtienen de virus tales como poliovirus, adenovirus 2, virus de simio 40 (SV40), y citomegalovirus (CMV) humano. Pueden utilizarse promotores genómicos virales, secuencias de control y/o señal para dirigir la expresión, con la condición de que dichas secuencias de control sean compatibles con la célula hospedadora elegida. También pueden usarse promotores celulares no virales (por ejemplo, los promotores de β -globina y EF-1 α), dependiendo del tipo celular en que tiene que expresarse la proteína recombinante.

45 Pueden usarse secuencias de ADN derivadas del genoma viral de SV40, por ejemplo, el origen de SV40, el promotor temprano y tardío, potenciador, sitios de corte y ajuste, y poliadenilación para proporcionar otros elementos genéticos útiles para la expresión de una secuencia de ADN heterólogo. Los promotores tempranos y tardíos son particularmente útiles porque ambos se obtienen fácilmente del virus SV40 como un fragmento que también comprende el origen de replicación viral de SV40 (Fiers *et al.*, Nature 273:113, 1978). También pueden usarse fragmentos más pequeños o más grandes de SV40. Típicamente, se incluye la secuencia de aproximadamente 250 pb que se extiende desde el sitio Hind III hacia el sitio BglII localizado en el origen de replicación de SV40.

50 Se han descrito previamente vectores de expresión bicistronicos usados para la expresión de múltiples transcritos (Kim S. K. y Wold B. J., Cell 42:129, 1985; Kaufman *et al.* 1991, *supra*) y pueden usarse en combinación con una secuencia de EESYR descrita en este documento. También serán útiles otros tipos de vectores de expresión, por ejemplo, aquellos descritos en la patente de Estados Unidos N° 4.634.665 (Axel *et al.*; y la patente de Estados Unidos N° 4.656.134 (Ringold *et al.*).

Células hospedadoras y transfección

60 Las células hospedadoras eucariotas usadas en los métodos de la invención son células hospedadoras de mamífero incluyendo, por ejemplo, células CHO y células de ratón. En este documento se describe una secuencia de ácido nucleico que codifica una secuencia de EESYR de célula CHO. Puede colocarse un sitio de integración, por ejemplo, un sitio de reconocimiento de recombinasa, dentro de una EESYR, o 5' o 3' a la secuencia EESYR. Un ejemplo de un sitio de integración adecuado es un sitio lox p. Otro ejemplo de un sitio adecuado de integración es dos sitios de reconocimiento de recombinasa, por ejemplo, un sitio lox p y un sitio lox 5511. La secuencia de EESYR puede localizarse en el cromosoma 6 del genoma de célula CHO. La secuencia de EESYR puede localizarse dentro de una

secuencia seleccionada entre el grupo que consiste en ácidos nucleicos que comprenden los nucleótidos 1-6473 y 4607-6473 de la SEC ID N° 1; y 1-7045, 1-3115, 1-2245, 1-935, y 1-465 de la SEC ID N° 2.

También se describe en este documento una célula hospedadora de mamífero transfectada con un vector de expresión descrito en este documento. Aunque puede usarse cualquier célula de mamífero, la célula hospedadora es preferiblemente una célula CHO.

Las células hospedadoras transfectadas incluyen células que se han transfectado con vectores de expresión que comprenden una secuencia que codifica una proteína o polipéptido. Las proteínas expresadas preferiblemente se secretarán en el medio de cultivo, dependiendo de la secuencia de ácido nucleico seleccionada, pero pueden retenerse en la célula o depositarse en la membrana celular. Pueden emplearse diversos sistemas de cultivo celular de mamífero para expresar proteínas recombinantes. Ejemplos de líneas celulares hospedadoras de mamífero adecuadas incluyen las líneas COS-7 de células renales de mono, descritas por Gluzman (1981) Cell 23:175, y otras líneas celulares capaces de expresar un vector apropiado incluyendo, por ejemplo, CV-1/EBNA (ATCC CRL 10478), células L, C127, 3T3, CHO, HeLa y líneas celulares BHK. También serán útiles otras líneas celulares desarrolladas para esquemas específicos de selección o amplificación con los métodos y composiciones proporcionados en este documento. Una línea celular preferida es la línea celular CHO denominada K1. Para conseguir el objetivo de elevado volumen de producción de proteínas recombinantes, la línea celular hospedadora debe estar pre-adaptada para medio de biorreactor en el caso apropiado.

Se conocen en la técnica varios protocolos de transfección, y se revisan en Kaufman (1988) Meth. Enzymology 185:537. El protocolo de transfección elegido dependerá del tipo de célula hospedadora y la naturaleza del GOI, y puede elegirse en base a experimentación rutinaria. Las necesidades básicas de cualquiera de estos protocolos son primero introducir ADN codificante de la proteína de interés en una célula hospedadora adecuada, y después identificar y aislar células hospedadoras que han incorporado el ADN heterólogo de un modo relativamente estable, expresable.

Un método habitualmente usado para introducir ADN heterólogo en una célula es precipitación con fosfato de calcio, por ejemplo, como se describe por Wigler et al. (Proc. Natl. Acad. Sci. USA 77:3567, 1980). El ADN introducido en una célula hospedadora por este método frecuentemente experimenta reordenamiento, haciendo que este procedimiento sea útil para la cotransfección de genes independientes.

La fusión inducida por polietileno de protoplastos bacterianos con células de mamífero (Schaffner et al., (1980) Proc. Natl. Acad. Sci. USA 77:2163) es otro método útil para introducir ADN heterólogo. Los protocolos de fusión de protoplastos frecuentemente producen múltiples copias del ADN plasmídico integrado en el genoma de la célula hospedadora de mamífero, y esta técnica requiere que el marcador de selección y amplificación estén en el mismo plásmido que el GOI.

También puede usarse electroporación para introducir ADN directamente en el citoplasma de una célula hospedadora, por ejemplo, como se describe por Potter et al. (Proc. Natl. Acad. Sci. USA 81:7161, 1988) o Shigekawa et al. (BioTechniques 6:742, 1988). A diferencia de la fusión de protoplastos, la electroporación no requiere que el marcador de selección y el GOI estén en el mismo plásmido.

Más recientemente, se han descrito varios reactivos útiles para introducir ADN heterólogo en una célula de mamífero. Éstos incluyen reactivo Lipofectin™ y reactivo Lipofectamine™ (Gibco BRL, Gaithersburg, Md.). Estos dos reactivos son reactivos disponibles en el mercado usados para formar complejos lípido-ácido nucleico (o liposomas) que, cuando se aplican a cultivos celulares, facilitan la captación del ácido nucleico en las células.

También es deseable un método para amplificar el GOI para la expresión de la proteína recombinante, y típicamente implica el uso de un marcador de selección (revisado en Kaufman *supra*). La resistencia a fármacos citotóxicos es la característica más frecuentemente usada como marcador de selección, y puede ser el resultado de un rasgo dominante (por ejemplo, puede usarse independiente del tipo de célula hospedadora) o un rasgo recesivo (por ejemplo, útil en tipos particulares de célula hospedadora que son deficientes en cualquier actividad que se esté seleccionando). Son adecuados varios marcadores amplificables para su uso en los vectores de expresión de la invención (por ejemplo, como se describe en Maniatis, Molecular Biology: A Laboratory Manual, Cold Spring Harbor Laboratory, NY, 1989; pág. 16.9-16.14).

Se muestran marcadores de selección útiles para amplificación génica en células de mamífero resistentes a fármaco en la Tabla 1 de Kaufman, R. J., *supra*, e incluyen resistencia a DHFR-MTX, P-glucoproteína y resistencia a múltiples fármacos (MDR)-diversos agentes citotóxicos lipófilos (por ejemplo, adriamicina, colchicina, vincristina), y adenosina desaminasa (ADA)-Xyl-A o adenosina y 2'-desoxicoformicina.

Otros marcadores de selección dominantes incluyen genes de resistencia a antibióticos derivados de forma microbiana, por ejemplo resistencia a neomicina, kanamicina o higromicina. Sin embargo, estos marcadores de selección no han demostrado ser amplificables (Kaufman, R. J., *supra*). Existen varios sistemas de selección adecuados para hospedadores de mamífero (Maniatis *supra*, pág. 16.9-16.15). También se han descrito protocolos

de co-transfección que emplean dos marcadores de selección dominantes (Okayama y Berg, Mol. Cell Biol 5:1136, 1985).

También pueden incluirse elementos reguladores útiles, descritos previamente o conocidos en la técnica, en las construcciones de ácido nucleico usadas para transfectar células de mamífero. El protocolo de transfección elegido y los elementos seleccionados para su uso en el mismo dependerán del tipo de célula hospedadora usada. Los especialistas en la técnica son conscientes de numerosos protocolos y células hospedadoras diferentes, y pueden seleccionar un sistema apropiado para la expresión de una proteína deseada, en base a las necesidades del sistema de cultivo celular usado.

Otras características de la invención llegarán a ser evidentes en el curso de las siguientes descripciones de realizaciones ejemplares que se dan para ilustración de la invención y no pretenden ser limitantes de la misma.

Ejemplos

Los siguientes ejemplos se presentan para proporcionar a los especialistas en la técnica el modo de hacer y usar los métodos y composiciones descritos en este documento, y no pretenden limitar el alcance de los que los inventores consideran como su invención. Se han hecho esfuerzos por asegurar la precisión con respecto a los números usados (por ejemplo, cantidad, temperatura, etc.) pero debe considerarse algo de error experimental y desviación. Salvo que se indique de otro modo, las partes son partes en peso, el peso molecular es peso molecular promedio, la temperatura es en grados centígrados, y la presión es en o cerca de atmosférica.

Ejemplo 1. Generación de líneas celulares RGC9 y RGC16.

Se infectaron células CHO K1 (1×10^7) con retrovirus pantrópico producido con plásmido pTE252 (Fig.1), que tienen un sitio lox p en el mismo, a una MOI de menos de 1 para generar una combinación estable de células con principalmente una inserción retroviral por célula. Las células en la combinación estable que expresaban una proteína marcadora a un elevado nivel se seleccionaron y expandieron. Se aislaron treinta y seis clones y se expandieron en 36 líneas celulares. Los clones que mostraban la mayor expresión y eficacia de recombinación se identificaron y clonaron, donde los clones contenían cada uno al menos un sitio de reconocimiento de recombinasa en un locus de expresión potenciada. Se seleccionaron ocho poblaciones celulares con la mejor expresión y eficacia de recombinación, y se volvieron a evaluar para expresión potenciada de proteínas. Se seleccionaron dos poblaciones celulares y se denominaron RGC9 y RGC16. El análisis de transferencia de Southern de las poblaciones celulares de las 36 líneas celulares originales que correspondían a estas dos líneas celulares mostró que se integraba una única copia de una construcción indicadora en el genoma de la célula CHO en el mismo locus en el caso de ambas poblaciones celulares, y se determinó que la localización de integración estaba en el triplete "act" en la posición de nucleótido 6.471 a 6.473 de la SEC ID N° 5. Se empleó al menos una de estas dos líneas celulares en experimentos descritos a continuación.

Ejemplo 2. Expresión de la proteína FcFP1 en medio de producción sin suero

Las células RGC38 se obtuvieron de células RGC9 y se adaptaron para crecer en suspensión en un medio de producción sin suero. Las células RGC38 se usaron como células hospedadoras para la expresión de proteína FcFP1 (proteína de fusión Fc-1). Las células RGC38 se transfectaron en una placa de diez centímetros con un vector de expresión de FcFP1, pTE851 y un plásmido Cre, pRG858. El plásmido de FcFP1 tiene, en dirección 5' a 3', un sitio LoxP, un promotor tardío de SV40, un gen de resistencia a higromicina, un IRES, un eGFP, un promotor CMV MIE, un gen que codifica una proteína FcFP1, y un sitio Lox511. Las células se cultivaron en medio F12 con 400 µg/ml de higromicina durante dos semanas después de la transfección. Se aislaron las células que expresaban eGFP pero no DsRed usando citometría de flujo y se denominaron RS421-1. Las células aisladas eran esencialmente isogénicas, aunque derivadas de diferentes células fundadoras. Las células RS421-1 se expandieron en cultivos de suspensión en medio de producción sin suero. Se examinó la proteína FcP1 en medio condicionado de cultivos de 3 días de edad en SDS-PAGE con tinción con azul de Coomassie. La proteína FcF1 en el medio condicionado era abundante y podía observarse sin purificación.

Ejemplo 3. Expresión regulada de proteína FcFP1 en medio de producción sin suero

Las células RGC49 se obtuvieron de RGC16, se adaptaron para crecer en medio de producción sin suero, y contenían un plásmido de expresión tetR-YFP integrado de forma estable, y pcDNA6/TR. La proteína TetR permite la regulación de la transcripción a partir de promotores que comprenden una secuencia operadora tet por tetraciclina o doxiciclina. Las células RGC49 se co-transfectaron con pTE851 y pRG858. Las células transfectadas se seleccionaron con 400 µg/ml de higromicina durante dos semanas. Se aislaron las células que expresaban eGFP pero no YFP usando citometría de flujo y se denominaron RS569-1. Las células RS569-1 se expandieron en cultivos de suspensión en medio de producción sin suero en presencia o ausencia de doxiciclina. Se examinó la proteína FcP1 en medio condicionado de cultivos de 3 días de edad por SDS-PAGE y tinción con azul de Coomassie. Las células RS569-1 expresaban proteína FcFP1 de forma similar a RS421-1 en presencia de 1 µg/ml de doxiciclina en el medio de cultivo. Se detectó muy poca proteína FcFP1 a partir de las células RS569-1 en ausencia de doxiciclina.

Ejemplo 4. Uso de un tercer sitio Lox (Lox2272) en EESYR para crear casete de expresión dual

Se transfectan células de cualquier línea celular que porte un gen DsRED flanqueado por un sitio LoxP y un sitio Lox511 en el locus EESYR con pRG858 y un vector que comprende, en dirección 5' a 3', un sitio LoxP, un primer promotor, una YFP, un sitio Lox2272, un segundo promotor, una eGFP, y un Lox511. Las células que expresan eGFP e YFP, pero no DsRed se aíslan. Las células aisladas después se transfectan con pRG858, pRG1167 (un vector que tiene, en dirección 5' a 3', un sitio LoxP, un promotor tardío de SV40, un gen de resistencia a higromicina, un promotor CMV MIE, un DsRed y un sitio Lox2272) o pRG1234 (un vector que tiene, en dirección 5' a 3', un sitio Lox2272, un promotor tardío de SV40, un gen de resistencia a higromicina, un promotor CMV MIE, un DsRed y un sitio Lox511). Se aíslan las células capaces de expresar DsRED y eGFP pero no YFP, o DsRED e YFP pero no eGFP.

Ejemplo 5. Expresión de anticuerpo a partir de células hospedadoras RGC38

Las células RGC38 se transfectaron con pTE963 y pRG858. pTE963 tiene, en dirección 5' a 3', un sitio LoxP, un promotor tardío de SV40, un gen de resistencia a higromicina, un IRES, una eGFP, un promotor CMV MIE, el gen de cadena ligera de un anticuerpo, un promotor CMV MIE, el gen de cadena pesada de un anticuerpo, y un sitio Lox511. Los cultivos transfectados se seleccionaron con higromicina a 400 µg/ml cada uno durante dos semanas. Las células que expresaban eGFP pero no DsRed se aislaron por citometría de flujo. Las células aisladas se expandieron en suspensión en medio de producción sin suero y se llamaron células RS533. Para la producción de anticuerpos, se cultivaron células RS533 en un biorreactor durante 10 días. Se recogieron alícuotas de medio usado desde el día seis hasta el día diez y se analizó su composición proteica por SDS-PAGE. El péptido de cadena pesada y cadena ligera del anticuerpo en el medio usado de biorreactor se detectaron fácilmente por tinción con azul de Coomassie.

Ejemplo 6. Rescate y subclonación de secuencias de EESYR

Se seleccionó una línea celular CHO (denominada RGC21) que expresaba elevados niveles de un gen indicador, DsRed, para el aislamiento de secuencias de EESYR, ya que el análisis de transferencia de Southern indicó que la expresión elevada de expresión de DsRed observada para esta línea celular está dirigida por una única integración de un casete de expresión que codifica DsRed. Se rescataron secuencias genómicas 5' al casete de expresión por transfección de células RGC21 con plásmidos pTE494 linealizados, un vector que tiene, en dirección 5' a 3', un sitio LoxP, un gen de resistencia a ampicilina, un origen bacteriano de replicación, un sitio de clonación múltiple (EcoRI, BglII, XbaI, ApaI), un promotor CMV MIE, un gen de neomicina fosfotransferasa, un IRES, una eGFP y un sitio Lox511. Se aislaron las células que expresaban eGFP pero no DsRed por citometría de flujo como una combinación. Se aisló el ADN genómico, se digirió con endonucleasa de restricción XbaI, y se auto-ligó para crear pRG 1106. Se rescataron secuencias genómicas 3' al casete de expresión por transfección de células RGC21 con plásmidos pTE495 circulares, un vector que tiene, en dirección 5' a 3', un sitio LoxP, un promotor CMV MIE, un gen de neomicina fosfotransferasa, un IRES, una eGFP, un origen bacteriano de replicación, un sitio de clonación múltiple (EcoRI, BglII, XbaI, ApaI), un gen de resistencia a ampicilina, y un sitio Lox511. Se aislaron las células que expresaban eGFP pero no DsRed por citometría de flujo como una combinación. Se aisló el ADN genómico, se digirió con endonucleasa de restricción MfeI, y se autoligó para crear pRG1099.

Ejemplo 7. Construcción plasmídica para análisis de EESYR

Se escindieron secuencias de EESYR de pRG1106 o pRG1099 como fragmentos Agel y se insertaron en los sitios Agel y NgoMIV de pTE575, un plásmido que expresa FCFP2, para producir el plásmido pTE809. El plásmido pTE575 tiene, en dirección 5' a 3', un promotor tardío de SV40, un gen de resistencia a higromicina, un promotor CMV MIE, y un gen que codifica la proteína FCFP2. En células CHO transfectadas de forma estable, pTE809 y pTE575 produjeron un 97,22% y un 38,57% de células que expresan niveles detectables de proteína FCFP2, respectivamente. La fluorescencia media de FcFP2 detectada por un anticuerpo conjugado con FITC fue 482,54 y 279,75 para cultivos transfectados con pTE809 y pTE575, respectivamente. Por tanto, la inclusión de EESYR en vectores de expresión aumentaba la expresión de proteína FCFP2 en transfección estable.

Ejemplo 8. Rescate y subclonación de secuencias de EESYR

Se seleccionó una línea celular CHO (denominada RGC21) que expresaba elevados niveles de un gen indicador, DsRed, para el aislamiento de secuencias de EESYR, ya que el análisis de transferencia de Southern indicaba que la expresión elevada de expresión de DsRed observada para esta línea celular está dirigida por una única integración de un casete de expresión que codifica DsRed. Se rescataron secuencias genómicas 5' al casete de expresión por transfección de células RGC21 con plásmidos pTE494 linealizados, un vector que tiene, en dirección 5' a 3', un sitio LoxP, un gen de resistencia a ampicilina, un origen bacteriano de replicación, un promotor CMV MIE, un gen de neomicina fosfotransferasa, un IRES, una eGFP y un sitio Lox511. Las células que expresaban eGFP pero no DsRed se aislaron por citometría de flujo como una combinación. Se aisló el ADN genómico, se digirió con endonucleasa de restricción XbaI, y se autoligó para crear pRG1106. Se rescataron secuencias genómicas 3' al casete de expresión por transfección de células RGC21 con plásmidos pTE495 circulares, un vector que tiene, en

dirección 5' a 3', un sitio LoxP, un promotor CMV MIE, un gen de neomicina fosfotransferasa, un IRES, una eGFP, un origen bacteriano de replicación, un gen de resistencia a ampicilina, y un sitio Lox511. Se aislaron las células que expresaban eGFP pero no DsRed por citometría de flujo como una combinación. Se aisló el ADN genómico, se digirió con endonucleasa de restricción MfeI, y se autoligó para crear pRG1099.

5

Ejemplo 9. Construcción plasmídica para el análisis de EESYR

Se escindieron las secuencias de EESYR de pRG1106 o pRG1099 como fragmentos Agel y se insertaron en los sitios Agel y NgoMIV de pTE575, un plásmido que expresa FCFP2, para producir el plásmido pTE809. El plásmido pTE575 tiene, en dirección 5' a 3', un promotor tardío SV40, un gen de resistencia a higromicina, un promotor CMV MIE, y un gen que codifica la proteína FCFP2. En células CHO transfectadas de forma estable, pTE809 y pTE575 produjeron un 97,22% y un 38,57% de células que expresaban niveles detectables de proteína FCFP2, respectivamente. La fluorescencia media de FcFP2 detectada por un anticuerpo conjugado con FITC fue 482,54 y 279,75 para cultivos transfectados con pTE809 y pTE575, respectivamente. Por tanto, la inclusión de EESYR en vectores de expresión aumentaba la expresión de proteína FCFP2 en transfección estable.

10

15

LISTADO DE SECUENCIAS

20

<110> Regeneron Pharmaceuticals, Inc.

<120> Regiones de expresión y estabilidad potenciadas

<130> 3051A

25

<140> A asignar

<141>04-06-2008

<150> 60/933.213

<151> 04-06-2007

30

<160> 8

<170> FastSEQ para Windows Versión 4.0

35

<210> 1

<211> 6473

<212> ADN

<213> *Cricetulus griseus*

40

<400> 1

ES 2 522 615 T3

tctagaaaca	aaaccaaaaa	tattaagtca	ggcttggett	caggtgctgg	ggtggagtgc	60
tgacaaaaat	acacaaattc	ctggctttct	aaggcttttt	cggggattca	ggtattgggt	120
gatggtagaa	taaaaatctg	aaacataggt	gatgtatctg	ccatactgca	tgggtgtgta	180
tgtgtgtgta	tgtgtgtctg	tgtgtgtgcc	cagacagaaa	taccatgaag	gaaaaaaaca	240
cttcaaagac	aggagagaag	agtgacctgg	gaaggactcc	ccaatgagat	gagaactgag	300
cacatgccag	aggaggtgag	gactgaacca	tcaacacaaa	gtggtgaata	gtcctgcaga	360
cacagagagg	gccagaagca	ctcagaactc	caggggggtca	ggagtgggtc	tctggagggt	420
tctgceettg	gaggttctctg	aggaggaggc	ttccatattg	aaaatgtagt	tagtggccgt	480
ttccattagt	acagtgacta	gagagagctg	agggaccact	ggactgaggc	ctagatgctc	540
agtcagatgg	ccatgaaagc	ctagacaagc	acttccgggt	ggaaaggaaa	cagcaggtgt	600
gaggggtcag	gggcaagtta	gtgggagagg	tcttccagat	gaagtagcag	gaacgggagc	660
gcaactggatg	gccccacttg	tcaaccagca	aaagcttggg	tcttgttcta	agaggccagg	720
gacatgacaa	gggtgatctc	ggtttttaa	agctttgtg	ttacctaate	acttctatta	780
gtcagatact	tgtaacaca	aatgagtact	tggcctgtat	tttagaaact	tctgggatcc	840
tgaaaaaaca	caatgacatt	ctggctgcaa	cacctggaga	ctcccagcca	ggccctggac	900
ccgggtccat	tcatgcaaat	actcagggac	agattcttca	ctaggtactg	atgagctgtc	960
ttggatgcaa	atgtggcctc	ttcattttac	tacaagtcac	catgagtcat	gaggtgctgt	1020
ttgcacagtg	tgactaagtg	atggagtggt	gactgcagcc	attcccggcc	ccagcttgtg	1080
agagagatcc	ttttaaattg	aaagtaagct	caaagttacc	acgaagccac	acatgtataa	1140
actgtgtgaa	taatctgtgc	acatacacaa	accatgtgaa	taatctgtgt	acatgtataa	1200
actgtgtgaa	taatctgtgt	gcagcctttc	cttacctaact	accttccagt	gatcaggttt	1260
ggactgcctg	tgtgctactg	gacctgaat	gtccccaccg	ctgtccctg	tcttttacga	1320
ttctgacatt	tttaataaat	tcagcggctt	cccctctgct	ctgtgcctag	ctataccttg	1380
gtactctgca	ttttggtttc	tgtgacattt	ctctgtgact	ctgctacatt	ctcagatgac	1440
atgtgacaca	gaaggtgttc	cctctggaga	catgtgatgt	ccctgtcatt	agtggaatca	1500
gatgccccca	aactgttgtc	cagtgtttgg	gaaagtgaca	cgtgaaggag	gatcaggaaa	1560
agaggggtgg	aaatcaagat	gtgtctgagt	atctcatgtc	cctgagtgg	ccaggctgct	1620
gacttcactc	ccccaaagtga	gggaggccat	ggtgagtaca	cacacctcac	acatactata	1680
tccaacacac	acacacacac	acacacacac	acgcacgcac	gcacgcacgc	acgcacacac	1740
gcacacacac	gaactacatt	tcacaaacca	catacgcata	ttacacccca	aacgtatcac	1800
ctatacatac	cacacataca	cacccctcca	cacatcacac	acataccaca	cccacacaca	1860
gcacacacat	acataggcac	acattcacac	accacacata	tacatttgtg	tatgcataca	1920
tgacatacaca	cacaggcaca	cagacaccac	acacatgcat	tgtgtacgca	cacatgcata	1980
cacacacata	ggcacacatt	gagcacacac	atacatttgt	gtacgcacac	tacatagaca	2040
tatatgcatt	tgtatatgca	cacatgcatg	cacacataca	taggcacaca	tagagcacac	2100
acatacattt	gtgtatgcac	acatgcacac	accaatcaca	tgggaagact	caggttcttc	2160

ES 2 522 615 T3

actaaggttc acatgaactt agcagttcct ggttatctcg tgaacttgg aagattgctg 2220
 tggagaagag gaagcgttgg cttgagccct ggcagcaatt aaccccgcce agaagaagta 2280
 ggtttaaaaa tgagagggtc tcaatgtgga acccgcaggg cggcagttca gagaagagac 2340
 ctaccaaacg caactgagag caaaggcaga gggatgaacc tgggatgtag tttgaacctc 2400
 tgtaccagct gggcttcctg ctattttgtt atatctttat taaatattct tttagtttta 2460
 tgtgctgtaa laccttgctt gcataaatgt atgggcactg tatgtgttct tggtgccggt 2520
 ggaggccagg agagggcatg gatcctccgg agctggcgtt tgagacagtt gtgaccaca 2580
 gtgtggggtc tgggaactgg gtcttagtgt tccgcaagtg cagctggggc tcttaacctc 2640
 tgagccatcc ctccagcttc aagaaactta tttcttagg acatggggga agggatccag 2700
 ggttttaggc ttgtttgttc agcaaact ctttctgtgt attttgaatt ttattttatt 2760
 ttactttttt gggatagaat cacattctgc agctcaggct gggcctgaac tcatcaaaat 2820
 cctcctgtct cagcttacca ggtgataaga ttactgatgt gagcctggtt ttgacaagca 2880
 ctttagagtc ccagccctt ctggacactt gttccaagta taatataat atataatata 2940
 atataatata atataatata atataattgt tgtgtgtgtt tgtgtgtgta tgagacactt 3000
 gctctaaggg tatcatatata atccttgatt tgcctttaat ttatttttta attaaaaatg 3060
 attagctaca tgtcacctgt atgogtctgt atcatctata tatccttctc tccttctctc 3120
 tctttctctc ttcttcttct cacccccagg catctatttt caaatccttg tgcagaggag 3180
 atgccaaagag tctcgttggg ggagatggtg agggggcgat acaggggaag agcaggagga 3240
 aggggggaca gactgggtgt ggtctttgga gagctcagga gaatagcagc gatcttccct 3300
 gtccctggtg tcacctctta cagccaacac cattttgggg cctggcagaa gatttgcac 3360
 gctggtcgca ggtctgccac acaacccccaa tctggcccca agaaaaggca cctgtgtgtg 3420
 actctggggg taaaggcgtt gcttggctgt ctccagctgg acttgaact cccgtttaat 3480
 aaagagttct gcaaaataat acccgcagag tcacagtgcc aggttccctg gcttctctga 3540
 agcggccaggc acgggttccc taggaaatgg ggccttgctt gccaaactcc cacggcttgc 3600
 cctgcaaacg gctgaatga tctggcactc tgcgttgcca ctgggatgaa atggaaaaaa 3660
 gaaaaagaag aagtgtctct ggaagcgggc gcgctcacac aaacccgcaa cgatttcta 3720
 aacactctcc attgagaatc tggagtgggg ttgccctcta ctggggagct gaagacagct 3780
 agtggggggc gggggaggac cgtgctagca tcttccacg gtgctcgtg gctgtggtgc 3840
 atgccgggaa ccgaaacgcg gaactaaagt caagtcttgc tttggtggaa ctgacaatca 3900
 acgaaatcac ttcgattggt ttctctttt tactggaatt cttggatttg atagatgggg 3960
 gaggatcaga gggggagggg aggggggggg agcaggaggg aggaggggag gaggggagga 4020
 ggggaggagg ggaaggaggg aagggatgga ggaataact aacttttcta attcaacatg 4080
 acaagatc acgagaagtg caccgctagt gaccgggagg agaatgcc tattgggcat 4140
 tatattccct gtcgtctaat ggaatcaaac tcttggttcc agcaccaagg attctgagcc 4200
 tatcctatcc aagacagtaa ctacagccca cacygaagag gctatacaac tgaagaata 4260
 aaattttcac ttattttcat ttctgtgact gcatgttccac atgtagagag ccacctgtgt 4320
 ctaggggctg atgtgctggg cagttagagt ctgagcccgt taactggaac aaccagaac 4380
 tcccaccaca gtttagagctt gctgagagag ggagccctt ggtgagattt gctgtgtat 4440
 ttatttagag acagggctct atactgtagt ccaagctagc ctccagctca cagaaattct 4500
 cctgttccgg ttccaaagt actggagttt tgagtgtgtg ttaattgaac gctaagaatt 4560
 tgcctgattg agaaaacctc aagtgggttt ggctaatccc cagcaccoca gaggctgagg 4620
 caggaggaat gagagaatcc aaggtttgcc agagccacag ggtgagctca atgtggagac 4680
 tgtgaggggt agctcaatgt ggagactgtg agggtgagct caatgtggag actgtgaggg 4740
 tgagctcaat gtggagactg tgaggggtgag ctcaatgtgg agactgtgag ggtgagctca 4800
 atgtggagac ctgtatcaag ataataatag tagtagtaac aatgcaggcg aggggtgtgt 4860
 tgagtgttag agcagttagt tgatttgaca tgcctgaggt ctcccgttcc atctgtggcc 4920
 ctgcaacagg aagggaggga ggaagggggg gaacgagaga gaggaaagag agacagaagc 4980
 taagataggg aatgagagag gaaggaagaa acgggaagaa attcagactc ctctctgagt 5040
 tccgccaacg cctagtgaca tctgtgacac accctaaggt ggcctttgtg tggcactggc 5100
 ttgggtggtc gggaaaggca ttttcagctt gttgcagaac tggcacagta gcatgctggg 5160
 tccgtgaaag tttctgcccg ttaacaagaa gtctctacta cttgtgacct caccagtga 5220
 aatttcttta atgtctctct ggtgttctgg gttttgcatt tttgtttcta aggatacatt 5280
 cctgggtgat gtcataaggt ccccaaagac acagtggggc tgtgttggat tgggaaagat 5340
 gatttatctg ggggtgcaaa aggaaaagaa gggaaacagg cacttgggaa aatgtcctcc 5400
 cgcccaccgg aattttggct tggcaaccgt ggtggaggag caagaacac gtggaogttt 5460
 gaggagccat ggggtctag gaggacagga agcagaagga gagagctggg ctgacagcct 5520
 gcaggcattg cacagtttca gaaggagatt acagcatgac tgagttttta gggatccaac 5580
 agggacctgg tlagagatcc tgtggctct gaggcaactt gacctcagcc agatggtatt 5640
 tgataaacct gctcttagag gaaaaacaga catagcaaac agagccactg ttagtatga 5700
 aactctcact ttgctgaggt catgtggggc catgcccagg ggtcaggctg aactcaact 5760
 caaaaacaag tgagaaattg aagacaatcc gtggtggcag ctactggaag ggcaccaca 5820

ES 2 522 615 T3

```

tccccagaaa gagtggagct gctaaaaagc catttgtgat aggcacagtt atcttgaatg 5880
catggagcag agattacgga aaaatcgaga atgttaatga ggcaacattc gagttgagtc 5940
atcagtggtg ggaaaccag acgcttccat ccctaataag gaacatcttg ctctcagtc 6000
aaatggaaat aaaaattggg gcttgaattt ggcaaatgat tcagaactct gtgtaggtat 6060
tttcacacgc acagtggata attttcatgt tggagtttat ttgtgctaaa aggcagaaaa 6120
gggtaaaaag cacatcttaa gagttatgag gttctacgaa taaaaataat gttacttaca 6180
gctatctctt aattagtacc ccttccacc tgtggttaatt tcctgagata gtcagtgagg 6240
aaaagatctc tcttctctt ctttctcccc ctccccctct ctccccctct cctccccctc 6300
ctccccctc tccccctct cccccctct tcttctttg ctcttctctc tctgctctct 6360
tctccccctc ttcttcattt attctaagta gcttttaaca gcacaccaat tacctgtgta 6420
taacgggaaa acacaggctc aagcagctta gagaagattg atctgtgttc act 6473

```

<210> 2

<211> 7045

5 <212> ADN

<213> *Cricetulus griseus*

<400> 2

```

actagcgtgc aattcagagg tgggtgaaga taaaaggcaa acatttgagg ccatttccctt 60
atttggcacg gcacttagga agtggacat gcctaatact ctggtttgta ccaccttcc 120
ctataatgga ctggttgagg agctcctggg caaccgattc tggcatctca ttggctcagag 180
gctgtttaa tgggtactctt atttgcaaag aaggctgtaa ctgttagctt taaaagcctc 240
tcctcaagaa agaagggaga aaggatatgg ctagacatat ctaatagact taaccactgt 300
gaaaagcctt agtatgaatc agatagaacc tatttttaac tcagttttga aaaaaataat 360
ctttatattt atttgtgtgt gtgtgtgtgt gtgtgtgtgt gtgtgtgtgt gtgtgtgtgt 420
gaaccacatg tagcaggtgc tggaggaggc cagaagaggg caccagatct cctggaactg 480
acaccacaca tggttatgag ctgcctgatg tgggtgctgg gaactgaact ctctgttct 540
gcaagagcag caactgttct ctttaactgat gagccatctc tccagcccc cccataattt 600
taattgttca ttttagtaaa ttttattcat aatcaattat cacagtataa aacaatgatt 660
ttatataat catatacata tcaaggatga cagttagggg gatatgtgtg tgtgtgtgtg 720
tgtgtgtgtg tgtgtgtgtg tgtgttattt gtgtgtgtgc tttttaagaa ggtgccatag 780
tactgcatt tctctgaagg atttcaaagg aatgagacat gtctgtctgc caggaacct 840
atcttctctt ttgggaatct gacccaaag aggtattctg aggaactgaa tgaagagctc 900
aagtagcagt gtcttaaac caaatgtgct gtctagagaa agtcaactgc atcagtgage 960
tgaggagaga tttactgagc ggaagacaag cgctctttga ttttaagtggc tccaacagtc 1020
acgctgtgg agtggagcct gtgctcaggt ctgaggcagt ctttgcctagc cagctgtgat 1080
gagcagtgaa gaaaggggg agatggaggc aggtgggag cagggctatg gttcagacta 1140
ggtatcgtga gcacaccagc tggttgactt gtggtctgtg ggtcaggcgt tgtaaacgcc 1200
ctcaggtca ggcagtcaca ttgcttgaag ctgaatgggt gaggcaacac agagagtgca 1260
aagaaggcaa agtaccacct ctccccgac ccaggtcact tctgggttat agctgagact 1320
ccggacagca tgaaccagc tggttagagc ttcagggaaa acttgatgtc tgcattgtgc 1380
tatgaaatgt gattcgggtac atctggagaa aatttataat gctggctcag tcaagcactg 1440
aacaaggta ccttggcttt gggagctaca tgacattgac ttgtagycag accttttttt 1500
ttctgcccgc caattcccag ataaccaata tggaggctca atattaatta taaatgctcg 1560
gctgatagct caggcttgtt actagctaac tcttccaact taaatgaacc ctttctatt 1620
atctacatc tgcacagtgat ctttacctgt tacttctctt tctctctct tgtctgactc 1680
tgccctctct cttcccagag tcttagtct ggttctctct cctaacctta tctgcccag 1740
ctgctgacca agcatttata attaatatta agtctcccag tgagactctc atccagggag 1800
gacttgggtg ctccccctc ctcatggca tccgtgtctt cctcttccct cgttcccc 1860
tctcttctct gctcttctc ctccacctc ccttctcatag tattgatggc aagggtgttc 1920
tagaatggag gagtgcccat aggcattgca agaaaccagt taggatgctc tgtgaggggt 1980
tghtaatcata agcagatggac acaattcaag ccacagagtg aagacggaag gatgcactgt 2040
gctctagagc aacttctggg gcagaatcac aggtgagtt tctgacttga gggcgaagag 2100
gccacgagga agggagtgag tttgtctgag ctagaagctc cggcccacct tttggtagca 2160
gacctgcccc caagcattgt ttgttaatca tgtgggactc gattttctc taaatctatg 2220
ttcaactctt aagaaaatgt gaattctcac attaaaattt agatatacgt cttttgggtg 2280
ggggggtgta aaaaatctc aagaatatgg atttctgggg gccggagaga tggctcagag 2340
gttaagagaa ctgggtgtct ttctagacat tctgagttca attcccagca accacatggt 2400
ggctcacaac catctgtaat ggcacctggt gccatcttct gacatgcatg gatcacatgca 2460
ggcagaaagc tgtatacata gtaaattgat aaatcttttt ttaaaaagag tatggattct 2520
gccgggtgtt ggtggcgcac gcctttaatc ccagcactct ggaggcagag gcaggtggat 2580

```

ES 2 522 615 T3

ctctgtgagt tccgagaccag cctgggtctat aagagctagt tccaggacag cctccaaagc 2640
cacagagaaa cctgtctcog aaaaacccaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 2700
gtatggatcc taagaaagcc gtaacagctg gagctgtgta cggagttcag cgtggtacta 2760
gaagaacaga cattcatgat gaaacacccc aggattttta cttagtatct agtttccatt 2820
gttgttttga gaccggctct tatgctctcc aggcctggcct caaactgctg atcttccocg 2880
ctctacctct caagtctctgg gactacttgg ctcataaaac agtttttgtc gggctccctg 2940
aagttatggt tgtacaaacc gtgggggtca atatactcac ttgggcagag agagaaggtc 3000
tgaatcccag acaatgactg catctcagga cagttgggaa gaggacaatg gcagaaggac 3060
ttagaaaaga tagactggag ggtggaaaag cagcaggaac agagaaacaa aacaggaagc 3120
ttgctatcca gggccactct ggagtcctgt ggcaagatgg aagcgggcta ggggaatata 3180
tttgtgctac tgtgtgtgtg tgtgtgtgtg tgtgtgtgtg tgtgtgtgtg caatgcctat 3240
caatgttgaa ggggaaatat gtataccaca ttgattctgg gagcaattct cagtatctgg 3300
cctagagaaa ggaatggccc ctgcagaata gacagagtga atggtgcctt ttatcatttg 3360
ctaaagtgaa ggagaaataa acatccttcc atagagtttc aggtaaatga accccacagt 3420
tcatctgtgc cgtggtggag gectggccaa cagtaaaaaa gattagacac ggacaaagtc 3480
tgaaggaaac acctcgaata ggaagaggag agccacctca ttctgtaact ttctcaagg 3540
ggaagatggt ccaagagtgg gaataaatgg tcaaaggggg gatttttaat taggaaaacg 3600
atcttctgta tcaactgtga aactggaggt tgatttgggg cataggacaa tagattttag 3660
gctttgcaaa aagctgtttc aaagcagaga aatggaatag agacaattat gtagcagga 3720
gggagggtgg ggcgaagatg gagacagaga agtgaaagct gactttaggg aagaggaaaca 3780
tagaccacag gggcggggcg gggggcaggg gcggggggcg gggctcaaag gaggcagtg 3840
gaacggtgct agtgttcgca gcgtaagcgt gaatgtgcaa gcgtctttgt ggtgtgtgac 3900
caggagtggc gtggctggct tgtgtgctgc ttgtaatccc agtctttgag gtttcacac 3960
tgttccacag tgggtgtgat tttccctcgg agagcatgag ggctctgctt tccccacatc 4020
ctccccagcg ttcgttggtg tttgtttcca agatgttagt gggtgagaca aagcctctct 4080
gttgatttgc ctttaacagv tgacaaaaaa agctcaacca ggagacattt ttgccttct 4140
ggaaggtaat gtcctcatgt agagcaatgg gacctctc taaggtgagg ctactcttgc 4200
agtttgcacc cagctcttct gatgcaggaa ggaagttggt gggcaagcaa gactgtttgc 4260
ttcttgcgat ggacacattc tgcacacaaa ggctcaggag gggagaaggc tgtttgatgt 4320
ttagcactca ggaaggcccc tgatgcactc gtgattagct gtctccatct gtggagcaga 4380
cacggactaa ctaaaaacca gtgttttttaa attgtcaagc ctttaagggtg aggaaattga 4440
cttattgtgc tgggccatac gtagagcaag tgcctcgc attgcaaatc ctccattcc aggtctcag 4500
gtttctaggg accagaatgg cctagaacta tctcacaate ctccattcc aggtctcag 4560
tgcataagat aaccactata ccagcctgcc agcctgcta cctgcctcc taaattttaa 4620
atcatgggga gtaggggaga atacacttat cttagttagg gtttctattg ctgtgaagag 4680
acaccatgag catggcaact cttataaagg aaaacattta gttgggtggc agtttcagag 4740
gttttagtac attgtcatca tggctgggaa catgatggca tgcagacaga catggtgctg 4800
gagaaaggga tgagagtcct acatcttgca ggcaacagga cctcagctga gacactggct 4860
ggtaccctga gcataggaaa cctcacagcc caccctcaca gtgacatatt tccctcaaca 4920
aagccatacc tccataatag gccactccct atgagatgac agggccaatt acattcaaac 4980
tgcataaaca ctttaaagta ttttattttt attattgtaa attatgtatg tagctgggtg 5040
gtggcagccg aggtgcacgc ctttaatccc agcacttggg aggcagagggc agatggatct 5100
ctgtgagttc aagaccagcc tggctataaa gagctagtgg caaggaagga tatacaaaaga 5160
acagttctag gatagccttc aaagccacag agaagtgctg tcttgaaaac caaaaattgt 5220
gctgggacct gtctctgctt tgggtgcttc ccactcccc agagctggac tcttggtea 5280
cactgaatca gctgcaaaat aaactcctgg atctctctct tgtaacagga gcccgagtc 5340
aggcgeccac ttgtcttctc gcaggattgc catagacttt ttctgtgtgc ccaccattcc 5400
agactgaagt agagatggca gtggcagaga ctgggaaggc tgcaacgaaa acaggaagtt 5460
attgcacctt gggaatagtc tggaaatgaa gcttcaaaac ttgcttcatg ttcagttgta 5520
cacagactca ctcccaggtt gactcacacg tgtaaatatt cctgactatg tctgcaactgc 5580
ttttatctga tgcctccttc ccaaaatgcc aagtgtaaaa ggtgagggaa tcaccttgg 5640
attcagagcc cagggtctgc ctccctaacc tggacttgtc tttctcgggc agcctctgac 5700
acccctcccc ccattttctc tatecagaagg tctgagcaga gttggggcac gctcatgtcc 5760
tgatacactc cttgtcttcc tgaagatcta actctgacc cagaaagatg gctaagggtg 5820
tgaagtgttt gacatgaaga cttgggtctta agaactggag caggggaaaa aagtccgatg 5880
tggcagcatg tacccgaaat ccagaaactg gggaggtaga gacggatgag tgcccggggc 5940
tagctggctg ctagctgaat ctgcaaatc tgccaaatc caactcctat tgaaaaacct 6000
ttaccaaaa acaaaaacaaa caaataataa caacaacaac aacaacaac tgaacctac 6060
aaggtggggc gctcttggct cttgaggaat gactcaccca aacccaaagc ttgccacagc 6120
tgttctctgg cctaaatggg gtgggggtgg ggcagagaca gagacagaga gagacatgac 6180
ttcctgggct gggctgtgtg ctctaggcca ccaggaactt tctctgtctt ctctctgtct 6240

ES 2 522 615 T3

```

ggcaccagcca gagcaccagc acccagcagc tgcacacacc tccctcogtg cttcttgagc 6300
aaacacaggt gccttggctc gtctattgaa ccggagtaag ttcttgcaga tgtatgcatg 6360
gaaacaacat tgcctgggtt ttatttctac tgttgatgata aaaaccgggg aactccagga 6420
agcagctgag gcagaggcaa atgcaaggaa tgcctgctcc tagcttgctc cccatggctt 6480
gccgggctg cttctgcaa gcccttctct cccattggc atgctgaca tgaacagegt 6540
ttgaaatgct ctcaaagtgc actttcaaag aaggcttctc tgatcttget aactaaatca 6600
gaccatggtt caccgtgcat tatctttctg ctgtctgtct gtctgtctgt ctgtctatct 6660
gtctatcctc tatcaatcat ctatctatct atcttctatt tatctaccta tcattcaatc 6720
atctatcttc taactagtta tcatttattt atttgtttac ttactttttt ttttgagac 6780
agtatttctc tgagtgcagc ccttggctgt cctggaacc attctgtaac caggctgtcc 6840
tcaaactcac agagatccaa ctgctctgc ctctctggtg ctggggttaa agacgtgcac 6900
caccaacgcc ccgctctatc atctatttat gtacttatta ttcagtcatt atctatctc 6960
taactatcca tcatctgtct atccatcctc tatctatcta tctatctatc 7020
tctatcatcc atctataatc aattg 7045

```

<210> 3

<211> 6473

5 <212> ADN

<213> *Cricetulus griseus*

<400> 3

```

agtgaacaca gatcaatctt ctctaagctg cttgagcctg tgttttcccg ttatacacag 60
gtaattgggtg tgcctgtaaa agctacttag aataaatgaa gaagaaaggg agaaggaggc 120
agaggagaag gagcaaagaa agaaggaaag ggggagggag ggagaggagg gaggggagga 180
gggagggagg gagaggagg gagggggaga aagaagagaa ggagagatct tttcccact 240
gactatctca ggaaattacc acagggtggaa gggggtaacta attaaggaat agctgtaagt 300
aacattatct ttatctgtag aacctcataa ctcttaagat gtgcttttta ccttttctg 360
ccttttagca caaataaaact ccaacatgaa aattatccac tgtgcgtgtg aaaataccta 420
cacagatgct tgaatcattt gccaaattca agcccaatt tttatttcca ttttgactga 480
gagcaagatg ttccttttag gggatggaag cgtctgggtt tcccacactg aatgactcaa 540
ctcgaatggt gcctcattaa cattctcgat tttccgtaa tctctgctcc atgcattcaa 600
gataactgtg cctatcacia atggcttttt agcagctcca ctctttctgg ggatgtggtg 660
gcccttccag tagctgccac cacggattgt ctccaatttc tcaactggtt ttgagttgag 720
tgtcagcctg acccctgggc atggccgcac atgactcagg caaagtgaga gtttcatcac 780
taaactgtgg tctggttget atgtctggtt tcccttaag agcaggttat tcaaatacca 840
tctggctgag gtcagttgct ctcagagccc acagaatctc taccaggtc cctgttggat 900
cctaaaaaac tcagtcctgc tgaatctcc tctgaaact gtgcaatgcc tgcaggctgt 960
cagcccagct ctctcctctc gcttctgtc ctctaggac cccatgctc ctcaaagctc 1020
cacgtgtttc ttgctcctcc accacgggtg ccaagccaaa attcgggtgg gcgggaggac 1080
attttcccaa gtgctggtt ccttctttt ccttttgaca cccagataa atcatcttc 1140
ccaatccaac acagcccac tgtgtctttg gggactcat gacatcacc aggaatgat 1200
ccttagaaac aaaaatgcaa aaccocagaac accaggagac aattaagaa attttactg 1260
gtgaggtcac aagtagtaga gacttcttgt taacgggcag aaactttcac ggaccagca 1320
tgctactgtg gcagttctgc aacaagctga aaatgcctt cccgaccacc caagccagt 1380
ccacacaaag gccaccttag ggtgtgcaca ggtgtcact aggcgttggc ggaactcagg 1440
aaggagtctg aatttcttcc cgtttcttcc tctctctc attcctatc ttagctctg 1500
tctctcttc ctctctctc tccccctc tctcctcc ctctctgtg cagggccaca 1560
gatggaccyg gagacctcaa gcatgtcaaa tcaactaact gctctaccac tcaaccacac 1620
cctcgcctgc attgttacta ctactattat tatcttgata caggctcca cattgagctc 1680
accctcacag tctccacatt gagetcacc tcacagtctc cacattgagc tcaccctcac 1740
agtctccaca ttgagctcac cctcacagtc tccacattga gctcaccctc acagtctcca 1800
cattgagctc atcctgtggc tctggcaaac ctgaaattct ctatctctc ctgctcagc 1860
ctctggggtc gtggggatta gccaaacca ettgaggtt tcttcaatca gcaaatctt 1920
agcgttcaat taacacacac tcataactcc agtactttgg aaaccggaac aggagaattt 1980
ctgtgagctg gaggctagct tggactacag tatgagacc tgtctctaaa taaatacaca 2040
aagaaatctc accaagggcc tccctctctc agcaagctct aactgtggtg ggagttctg 2100
gttgttccag ttaacgggct cagaactcta ctgccagca catcagcccc tagacacag 2160
tggtctctca catgtgaaca tgcagtcaca gaaatgaaat aaagtgaaaa ttttattct 2220
tcagttgtat gcctcttcc gtgtgggctg tagttactgt cttgaaatagg ataggctcag 2280
aatccttggg agctggaacca agagtttgat tccatagac gacagggaa ataatgcca 2340
atagggcatt cctcctccc gtcactagcg gtgcactttc tccgaatctt tgtcatgtg 2400

```

10

ES 2 522 615 T3

aattagaaaa gttagtattt teetccatcc ctteccctcc tcccctctcc cccctctccc 2460
ctctccccc cctccctccg tctcccgcgc cctcccccct cccctctgatc ctcccccatc 2520
tatcaaatcc aagaattcca gtaaaaagag gaaaaacaate gaagtgattt cgttgattgt 2580
cagttccacc aaagcaagac ttgacttttag ttccgcggtt cgggtcccggt catgcaccac 2640
agccagcgag caccgtggaa ggatgctage acggctctcc ccccgcccc actagctgtc 2700
ttcagctccc cagtagaggg caaccgcact ccagattctc aatggagagt gtttacacaa 2760
tcgttgccggg ttgtgtgag cgcgcccgcet tccagagaca ctctctcttt ttcttttttc 2820
catttcattc cagtggcaac gcagagtgcc agatcattca ggccgtttgc agggcaagcc 2880
gtgggagctt ggcaagcaag gccccatttc ctagggaacc cgtgcctggc gcttcaggaa 2940
agcaagggaa cctggcactg tgactctgcg ggtattatatt tgcagaactc tttattaaac 3000
gggagtttca agtccagctg gagacgacca ggacagcct ttaaccccag agtcacacac 3060
aggtgccttt tcttggggcc agattgggggt tgtgtggcag acctgcgacc agcttgacaa 3120
ctctctctcc aggccacaaa atgggtgttg ctgtaagagg tgacaccagg gacagggaa 3180
atcgctgcta ttctcctgag ctctccaaa acccacacca gtctgtcccc ctctctctct 3240
gctcttcccc tgtatcggcc cctcaccatc tcccccaacg agactcttgg catctctctg 3300
gcacaaggat ttgaaaatag atgcttgggg ctacaggtg atagagagaa agagagagaa 3360
ggaagggagg atatatagat gatacagacg ctagacaggtg acatgtagct aatcattttt 3420
aattaaaaaa taattaaaa gcaaatcaag gatatatatg atacccttag agcaagtgtc 3480
tcatacacac acaaacacac acacacaata tatatatata tatatatata tatatatata 3540
tatatatata ttatacttgg aacaagtgtc cagaagggtt ggggactcta aagtcttgt 3600
caaagccagg ctcacatcag taatcttacc acctggtaga ctgagacagg aggattttga 3660
tgagttcagg cccagcctga gctgcagaat gtgattctat cccaaaaaag taaaataaaa 3720
taaaattcaa aatacacgaa aagagtattt gctgaacaaa caagcctaaa gccctggatc 3780
ccttccccca tgtcctaaga aaataagttt cttgaagctg gagggatggc tcagaggtta 3840
agagccccag ctgcacttgc ggaacactaa gaccagttc ccagacccca cactgtgggt 3900
cacaactgtc tcaaacgcca gctccggagg atccatgccc tctcctggcc tccaccggca 3960
ccaagaacac atacagtgcc catacattta tgcaagcaag gtattcacgc acataaaact 4020
aaaagaatat ttaataaaga tataacaaaa tagcatgaag cccagctggc acagaggttc 4080
aaactacatc ccaggttcat cctctgctt ttgctctcag ttgcttggg taggtctctt 4140
ctctgaactg gcgcccctg ggttccacat tgagaccctc tcatttttaa acctactctt 4200
tctggcgggg ttaatttgc gccagggctc aagccaacgc ttcctctctt ccacagcat 4260
cttcccaagt ctacagagata accaggaact gctaagttca tgtgaacctt agtgaagaac 4320
ctgagttctt ccatgtgatt ggtgtgtgca tgtgtgcata cacaaatgta tgtgtgtgct 4380
ctatgtgtgc ctatgtatgt gtgcatgcat gtgtgcatat acaaatgcat atatgtctat 4440
gtaggtgtgc tacacaaatg tatgtgtgtg ctcaatgtgt gcctatgtgt gtgtatgcat 4500
gtgtgctgac acaatgcatg tgtgtgtgt ctgtgtgctt gtgtgtgtat gcatgtatgc 4560
atacacaatg gtatgtgtgt ggtgtgtgaa tgtgtgccta tgtatgtgtg tgctgtgtgt 4620
gggtgtggtg tgtgtgtgat gtgtggaggg gtgtgtatgt gtggatgta taggtgatac 4680
gtttgggggt taaatgcgt atgtggtttg tgaatgtag ttcgtgtgtg tgcaggtgtg 4740
cgtgcgtgcg tgcgtgctg cgtgtgtgtg tgtgtgtgtg tgtgtgtgtt ggatatagta 4800
tgtgtgaggt gtgtgtactc acctggcct cctcacttg ggggagtga gtcagcagcc 4860
tggaccactc agggacatga gatactcaga cacatcttga tttccacccc tctttctctg 4920
atctccttc acctgtcact tccccaaaca ctggacaaca gtttgggggc atctgattcc 4980
actaatgaca gggacatcac atgtctccag agggaacacc ttctgtgtca catgtcatct 5040
gagaatgtag cagagtcaca gagaaatgtc acagaaacca aaatgcagag taccaaggtta 5100
tagctaggca cagagcagag ggaagccgc tgaatttatt aaaaatgtca gaatcgtaaa 5160
agacagggga cagcgggtgg gacattcagg gtccagtagc acacaggcag tccaaaactg 5220
atcactggaa ggtagttagt aaggaaaggg tgcacacaga ttattcacac agtttataca 5280
tgtacacaga ttattcacat ggtttgtgta tgtgcacaga ttattcacac agtttataca 5340
tgtgtggctt cgtggttaact ttgagcttac tttcaattta aaaggatctc tctcacaagc 5400
tggggccggg aatggctgca gtcaacactc catcacttag tcacactgtg caaacagcac 5460
ctcctgactc atggtgactt gtagtaaaat gaagaggcca catttgcatc caagacagct 5520
catcagtaacc tagtgaagaa totgtccctg agtatttga tgaatggacc cgggtccagg 5580
gcttggctgg gagtctccag gtgttgacgc cagaatgtca ttgtgttttt caaggatccc 5640
agaagtttct aaaatacagg ccaagtactc atttgtgta caaagtatct gactaataga 5700
agtgattagg taacacaaag ccttttaaaa accgagatca cccttgcctat gtcctggcc 5760
tcttagaaca agatccaagc ttttgcgtgt tgacaagtgg ggccatccag tgcgtctccg 5820
ttcctgctac ttcatctgga agacctctcc cactaacttg cccctgaccc ctcacacctg 5880
ctgtttcctt tccaccggga agtgcttgtc taggctttca tggccatctg actgagcatc 5940
taggcctcag tccagtgtc cctcagctct ctctagtcac tgtactaatg gaaacggcca 6000
ctaactacat tttcaatatg gaagcctcct cctcagggaac ctccaagggc agaagcctcc 6060

ES 2 522 615 T3

agagaaccac	tcoctgacccc	ctggagttct	gagtgcttct	ggccctctct	gtgtctgcag	6120
gactattcac	cacttgtgtt	gaatggttca	gtccctacct	cctctggcat	gtgctcagtt	6180
ctcatctcat	tggggagtc	ttcccaggtc	actctctct	cctgtctttg	aagtgttttt	6240
ttccttcatg	gtatttctgt	ctgggcacac	acacagacac	acatacacac	acatacacac	6300
ccatgcagta	tggcagatac	atcacctatg	tttcagattt	ttattctacc	atcacccaat	6360
acctgaatcc	ccgaaaaagc	cttagaaaagc	caggaatttg	tgtatttttg	tcagcactcc	6420
accccagcac	ctgaagccaa	gcctgactta	atatttttgg	ttttgtttct	aga	6473

<210> 4

<211> 7045

5 <212> ADN

<213> *Cricetulus griseus*

<400> 4

caattgatta	tagatggatg	atagatagat	agatagatag	atagatagat	agatagatga	60
tggatagaca	gatgatggat	agtttagagga	tagataatga	ctgaataata	agtacataaa	120
tagatgatag	agcggggcgt	tgggtggtgca	cgtctttaac	cccagcacca	gagaggcaga	180
ggcagttgga	tctctgtgag	tttgaggaca	gcctggttac	agaatgggtt	ccaggacagc	240
caaggctgtc	actcagagaa	atactgtctc	aaataaaaaa	agtaagtaaa	caaataaata	300
aatgataact	agttagaaga	tagatgattg	aatgataggt	agataaatag	aagatagata	360
gatagatgat	tgatagatga	tagacagata	gacagacaga	cagacagaca	gacagcagaa	420
agataatgca	cggtgaaaca	tggctctgatt	tagttagcaa	gatcagagaa	gccttctttg	480
aaagtgacat	ttgagagcat	ttcaaagcgt	gttcatgtca	ggcatgccaa	tggggagaga	540
agggcttgca	gaaagcaggc	ccggcaagcc	atggggagca	agctaggagg	cagcattcct	600
tgcatttgcc	tctgcctcag	ctgcttcctg	gagttccccg	gtttttatca	caacagtaga	660
aataaaaacca	ggacaatggt	gtttccatgc	atacatctgc	aagaacttac	tccggttcaa	720
tagacagacc	aaggcacctg	tgtttgctca	agaagcacgg	agggagggtg	gtgcacctgc	780
tgggtgctgg	tgctctggct	gtgccagaca	gagagcaaga	caggaaagtt	cctgggtggc	840
tagagcacac	agcccagccc	aggaagtcct	gtctctctct	gtctctgtct	ctgccccacc	900
cccccccat	ttaggccaga	gaacagctgt	ggcaagcttt	gggtttgggt	gagtcattcc	960
tcaagagcca	agagcccgcc	accttgtatg	gggtagtttg	ttggtgttgt	tgttgttatt	1020
atgtgtttgt	ttgtttgttt	ggtaaaggtt	ttcaatagg	agttggaatt	tggcaattca	1080
gctaggetgg	ctgagcagcc	agctagcccc	gggcaactcat	ccgtctctac	ctccccagtt	1140
ctgggatttc	gggtacatgc	tgccacatcc	gacttttttc	ccctgctcca	gttcttaaga	1200
ccaagtcttc	atgtcaaaca	cttcaccacc	ttagccatct	ttctgggtca	caagttagat	1260
cttcaggaag	acaaggagtg	tatcaggaca	ttagcctgcc	ccaactctgc	tcagaccttc	1320
tgatagagaa	aatgggggga	ggggtgtcag	aggtgcccgg	agaaagacaa	gtccaggtta	1380
aggaggacga	ccctgggctc	tgaatccaag	ggtgattccc	tcaccttgta	cacttggcat	1440
tttgggaagg	aagcatcaga	taaaagcagt	gcagacatag	tcaggaatat	ttacacgtgt	1500
gagtcaacct	gggagtgagt	ctgtgtacaa	ctgaacatga	agcaagtttt	gaagcttcat	1560
ttccagacta	ttcccagggt	gcaataaact	cctgttttcg	ttgcagcctt	cccagtctct	1620
gccactgcc	tctctacttc	agtctggaat	ggtgggcaca	cagaaaaagt	ctatggcaat	1680
cctgcgagaa	gacaagtggg	cgctgactt	cgggctcctg	ttacaagaga	ggaatccagg	1740
agtttatatt	gcagctgatt	cagtggtgac	caagactcca	gctctggggg	agtgggaagc	1800
aaccaaaagca	gagacaggtc	ccagcacaat	ttttgggttt	caagacagca	cttctctgtg	1860
gctttgaagg	ctatcctaga	actgttcttt	gtatatacctt	ccttgcaact	agctcttata	1920
gaccaggctg	gtcttgaact	cacagagatc	catctgcctc	tgcctcccaa	gtgetgggat	1980
taaaggcgtg	cacctcggct	gccaccacc	agctacatac	ataatttaca	ataataaaaa	2040
taaaataactt	taaagtgtta	tagcagtttg	aatgtaattg	gccctgtcat	ctcataggga	2100
gtggcactat	taggaggtat	ggctttgttg	aaggaaatat	gtcactgtga	gggtgggctg	2160
tgaggtttcc	tatgctcagg	gtaccagcca	gtgtctcagc	tgaggtcctg	ttgectgcaa	2220
gatgtaggac	tctcatccct	ttctccagca	ccatgtctgt	ctgcatgcca	tcatgttccc	2280
agccatgatg	acaatgtact	aaaacctctg	aaactgccac	ccaactaaat	gttttctttt	2340
ataagagttg	ccatgctcat	ggtgtctctt	cacagcaata	gaaacctaa	ctaagataag	2400
tgtattctcc	cctactcccc	atgatttaaa	atttaggaag	gcaggtaggc	aggcaggcag	2460
gctggtatag	tggttcatte	tagcacctga	gacctggaat	gggaggattg	tgagttagtt	2520
ctaggccatt	ctggtgccta	gaaaccagag	ccgggggttg	gcccaatgca	gagcacttgc	2580
tctacgtatg	gcccagcaca	ataagtcaat	ttcctcacct	taaaggcttg	acaatttaaa	2640
aacactgggt	tttagttagt	ccgtgtctgc	tccacagatg	gagacagcta	atcacagatg	2700
catcaggggc	cttccctgagt	gctaaacatc	aaacagcctt	ctccccctct	gagcctttgt	2760
gtgcagaatg	tgtccatcgc	aagaagcaaa	cagctctgct	tgcccaccaa	cttccctcct	2820

10

ES 2 522 615 T3

gcatcagaag agctgggtgc aaactgcaag agtagcctca ccttagagat ggggccatt 2880
 gctctacatg ggagcattac cttccaagaa ggcaaaaatg tctcctgggt gagctttttt 2940
 tgtcacctgt taaaggcaaa tcaacagaga ggctttgtct caccocactaa catcttgaa 3000
 acaaatacca acgaacgctg gggaggatgt ggggaaagca gagccctcat gctctccgag 3060
 ggaaaatcac acccactgtg gaacagtgtg gaaacctcaa agactgggat tacaagcagc 3120
 acacaagcca gccacgctac tctgggtcac acaccacaaa gacgcttgca cattcacgct 3180
 tacgctgcca acactagcaa cgttcccact gcctcctttg agccccgcc cccgccctg 3240
 cccccggccc cgccccgtg gtctatgttc ctcttcctta aagtcagctt ccacttctct 3300
 gtctccatct tgcccccacc ctccctcctc gctacataat tgtctctatt ccatttctct 3360
 gctttgaaac agctttttgc aaagcatcaa atctattgtc ctatgcccc aatcaacctc 3420
 cagtttcaca agtgatacag gaaatcgttt tctaatttaa aaatcccccc tttgaccatt 3480
 tattcccact cttggaacat ctcccccttg aggaaagtta cagaatgagg tggctctct 3540
 ctctctatc gaggtgtttc cttcagactt tgtcctgtc taatctttt aactgttggc 3600
 caggcctcca ccacggcaca gatgaactgt ggggttcatt tacctgaaac tctatggaag 3660
 gatgtttatt tctccttcac tttagcaaat gataaagggc accattcact ctgtctatc 3720
 tgcagggggc attcctttct ctaggccaga tactgagaat tgcctccaga atcaatgtgg 3780
 tatacatatt tccccctcaa cattgatagg cattgatcac acacacacac acacacacac 3840
 acacacacac acacagtgc acaaatgtat tcccctagcc cgcttccatc ttgccacagg 3900
 actccagagt ggccctggat agcaagcttc ctgtttgtt tctctgttcc tgcctgtttt 3960
 ccacctcca gttctatctt totaagtctc tctgcaattg tctctctctg cccaagtggag tatattgacc 4020
 agatgcagtc attgtctggg attcagacct tctctctctg cccaagtggag tatattgacc 4080
 cccacggttt gtacaacccat aacttcaggg agccccacaa aaactgtttt atgagccaag 4140
 tagtcccagg acttgagagg tagaggcggg aagatcagca gtttgaggcc agcctggaga 4200
 gcataagagc cgggtctcaaa acaacaatgg aaactagata ctaagtaaaa atcctggggg 4260
 gtttcatcat gaatgtctgt tcttctagta ccacgctgaa ctccgtacac agctccagct 4320
 gttacggctt tcttagaate catactcttt ttttttttt ttttttttt tttttttttg 4380
 tttttcgaga cagggtttct ctgtggcttt ggagctgtc ctggaactag ctcttataga 4440
 ccaggctggt ctggaactca cagagatcca cctgctctg cctccagagt gctgggatta 4500
 aaggcgtgag ccaccaacac ccggcagaat ccatactctt tttaaaaaaa gatttatcaa 4560
 tttactatgt atacagcttt ctgctgcat gtatccatgc atgtcagaag atggcaccag 4620
 gtcgcattac agatggtgtg gagccacct gtggtgtctg ggaattgaaac tcagaatgtc 4680
 tagaagagca accagttctc ttaacctctg agccatctct ccggccccc gaaatccata 4740
 ttcttgagga ttttttacac cccccccacc aaaagacgta tatctaaatt ttaatgtgag 4800
 aattcacatt ttcttaagag ttgaacatag atttagagga aaatcagatc ccacatgatt 4860
 aacaaagcat gcttgtgggc aggtctgcta ccaagaggtg ggccgtagct tctagctcag 4920
 acaaaactcac tcccctcctc gtggcctctt cgcctcaag tcagaaactc acctgtgat 4980
 tctgccccag aagttgctct agagcacagt gcactcctcc gtcttcactc tgggttga 5040
 attgtgtcca tgccttatga ttacaacccc tcacagagca tectaactgg tttctttgca 5100
 tgcctatggg cactcctcca ttctagaaca ccttgccat caatactatg aaaggagggg 5160
 tggaggagga agagcaggaa gaggaggggg aagcgagggg agaggaagac acgggatggca 5220
 atgaggaggg gggagcacc caggctctcc tggatgagag tctcactggg agacttaata 5280
 ttaattataa atgcttggtc agcagctggg caggataagg ttaggcagga gaaccagact 5340
 aaggactctg ggaagcagaa gggcagagtc agacaaggag aggaaacagg aagtacaagg 5400
 taaagtcacg tggcagaatg tagataatag aaatgggttc atttaagttg gaagagttag 5460
 ctagtacaaa gctgagcta tcagccgagc atttataatt aatattgagc ctccatattg 5520
 gttatctggg aatggcggg cagaaaaaaa aaagtctgcc tacaagtcaa tgtcatgtag 5580
 ctcccaaagc caaggtacct ttgttcagtg cttgactgag ccagcattat aaattttctc 5640
 cagatgtacc gaatcacatt tcatagcaac atgcagacat caagttttcc ctgaagctct 5700
 aaccagctgg ttgcatgctg tccggagtct cagctataac ccagaagtga cctgggtcgg 5760
 ggaagaggtg gtactttgcc ttctttgca tctctgtgtt gcctcacoca ttcagcttca 5820
 agcaatgtga ctgctgacc ctgagggcgt ttacaacgcc tgacccacag accacaagtc 5880
 aaccagctgg tgtgtcaag atacctagtc tgaacctag cctgctccc acctgctc 5940
 catctccacc ctttcttcac tgcctatcac agctggctag caaagactgc ctccagctg 6000
 agcacaggct ccaactccaca gccgtgactg tctcgaccac ttaaatacaa gagcgttgt 6060
 ctccgctca gtaaatctct cctcagctca ctgatgactg tgactttctc tagacagcac 6120
 atttgggttt aagacactgc tacttgact cttcattcag ttctcagaa tacctcattt 6180
 gggtcagatt cccaaagagg aagataggg tctggcaga cagacatgtc tcaattcctt 6240
 gaaatectc agagaaatgc agtgactatg gcacctctt aaaaagcaca cacacaaata 6300
 acacacacac acacacacac acacacacac acacacacac atatccccct cactgtcatc 6360
 cttgatatgt atatgatata tataaaatca ttgtttata ctgtgataat tgattatgaa 6420
 taaaatttac taaaatgaac aattaaaatt atgggggggg ctggagagat ggctcatcag 6480

ES 2 522 615 T3

ttaagagaac agttgctgct cttgcagaac acgagagttc agttcccagc acccacatca 6540
 ggagctcat aaccatgtgt ggtgtcagtt ccaggagatc tgggtccctc ttctggcctc 6600
 ctccagcacc tgctacatgt ggttcacaca cacacacaca cacacacaca 6660
 cacacacaca caaataaata taaagattat ttttttcaaa actgagttaa aaataggttc 6720
 tatctgattc atactaaggc ttttcacagt ggttaagtct attagatatg tctagccata 6780
 tcctttctcc cttctttctt gaggagagggc ttttaaagct acaagttaca gccttctttg 6840
 caaataagag taccatttaa caggcctctg accaatgaga tgccagaatc ggttgcccag 6900
 gagcttccc aacagtcocat tatagggaaa ggtggtacaa accagtagat taggcattgt 6960
 ccacttctta agtgccgtgc caaataagga aatggcctca aatgtttgcc ttttatcttc 7020
 acccacctct gaattgcaag ctagt 7045

<210> 5

<211> 13515

5 <212> ADN

<213> *Cricetulus griseus*

<400> 5

tctagaaaca aaacaaaaa tattaagtca ggcttggctt caggtgctgg ggtggagtgc 60
 tgacaaaaat acacaaatc ctggctttct aaggcttttt cggggattca ggtattgggt 120
 gatggtagaa taaaaatctg aaacataggt gatgtatctg ccatactgca tgggtgtgta 180
 tgtgtgtgta tgtgtgtctg tgtgtgtgcc cagacagaaa taccatgaag gaaaaaaaca 240
 cttcaaagac aggagagaag agtgacctgg gaaggactcc ccaatgagat gagaactgag 300
 cacatgccag aggaggtgag gactgaacca ttcaacacaa gtgggtgaata gtctgcaga 360
 cacagagagg gccagaagca ctcagaactc cagggggtea ggagtggtc tctggaggct 420
 tctgcccttg gaggttctct aggaggaggc ttccatattg aaaatgtagt tagtggcctg 480
 ttccattagt acagtgacta gagagagctg agggaccact ggactgaggc ctagatgctc 540
 agtcagatgg ccatgaaagc ctagacaagc acttccgggt ggaaaggaaa cagcaggtgt 600
 gaggggtcag gggcaagtta gtgggagagg tcttccagat gaagttagcag gaacggagac 660
 gcaactggat gccccaactg tcaaccagca aaagcttggg tcttgttcta agaggccagg 720
 gacatgacaa ggtgatctc ggtttttaa aggcttggg ttacctaatc acttctatta 780
 gtcagatact ttgtaacaca aatgagtact tggcctgtat tttagaaact tctgggattc 840
 tgaaaaaaca caatgacatt ctggctgcaa cacctggaga ctcccagcca ggccctggac 900
 cggggtccat tcatgcaaat actcagggac agattcttca ctaggtagct atgagctgtc 960
 ttggatgcaa atgtggcctc ttcattttac tacaagtca catgagtcag gaggtgctgt 1020
 ttgcacagtg tgactaagtg atggagtgt gactgcagcc attcccggcc ccagcttgtg 1080
 agagagatcc ttttaaattg aaagtaagct caaagttacc acgaagccac acatgtataa 1140
 actgtgtgaa taatctgtgc acatacacia accatgtgaa taatctgtgt acatgtataa 1200
 actgtgtgaa taatctgtgc acatacacia accatgtgaa taatctgtgt acatgtataa 1260
 ggactgectg tgtgctactg gaccctgaat gtccccaccg ctgtcccctg tcttttacga 1320
 ttctgacatt ttttaataaat tcagcggctt cccctctgct ctgtgcctag ctatacctg 1380
 gtactctgca ttttggtttc tgtgacattt ctctgtgact ctgctacatt ctcatatgac 1440
 atgtgacaca gaaggtgttc cctctggaga catgtgatgt cctgtcatt agtggaatca 1500
 gatgccccca aactgttgc cagtgtttgg gaaagtgaca cgtgaaggag gatcaggaaa 1560
 agaggggtgg aactcaagat gtgtctgagt atctcatgct cctgagtggt ccaggctgct 1620
 gacttcactc cccaagtga gggaggccat ggtgagtaca cacacctcac acatactata 1680
 tccaacacac acacacacac acacacacac acgcaagcac gcacgcagc acgcacacat 1740
 gcacacacac gaactacatt tcacaaacca catacgcata ttacaccca aacgtatcac 1800
 ctatacatac cacacataca caccctcca cacatcacac acataccaca cccacacaca 1860
 gcacacacat acataggcac acattcacac accacacata tacatttgtg tatgcataca 1920
 tgcatacaca cacaggcaca cagacaccac acacatgcat tgtgtacgca cacatgcata 1980
 cacacacata ggcacacatt gagcacacac atacatttgt gtacgcacac tacatagaca 2040
 tatatgcatt tgtatagca cacatgcat cacacata taggcacaca tagagcacac 2100
 acatacattt gtatagca acatgcacac accaatcaca tgggaagact caggttcttc 2160
 actaaggttc acatgaactt agcagttcct ggttatctcg tgaaacttgg aagattgctg 2220
 tggagaagag gaagcgttgg cttgagccct ggcagcaatt aaccccgcctc agaagaagta 2280
 ggtttaaaaa tgagagggtc tcaatgtgga acccgcaggg cgccagttca gagaagagac 2340
 ctacccaagc caactgagag caaaggcaga gggatgaacc tgggatgtag tttgaacctc 2400
 tgtaccagct gggtctcatg ctattttgtt atatctttat taaatattct tttagtttta 2460
 tgtgcgtgaa taccttgett gcataaatgt atgggcactg tatgtgttct tgggtcccgg 2520
 ggaggccagg agagggcatg gatcctccgg agctggcgtt tgagacagtt gtgaccaca 2580
 gtgtggggtc tgggaaactg gtcttagtgt tccgcaagtg cagctggggc tcttaacctc 2640

ES 2 522 615 T3

tgagccatcc	ctccagcttc	aagaaactta	ttttcttagg	acatggggga	agggatccag	2700
ggctttággc	ttgtttggtc	agcaaaact	ctttctogtg	atthtgaatt	ttatthtatt	2760
ttactttttt	gggatagaat	cacattctgc	agctcaggct	gggcctgaac	tcataaaaat	2820
cctcctgtct	cagtctacca	ggtgataaga	ttactgatgt	gagcctggct	ttgacaagca	2880
cttttagagtc	cccagccctt	ctggacactt	gttccaagta	taatataat	atatataat	2940
atatataat	atatataat	atatattgtg	tgtgtgtgtt	tgtgtgtgta	tgagacactt	3000
gctctaaggg	tatcatatat	atccttgatt	tgcttttaat	ttatthttha	atthaaaaatg	3060
attagctaca	tgtcacctgt	atgctgtctg	atcatctata	tatecttctc	tccttctctc	3120
tctttctctc	ttcttctctc	caccccccaag	catctattht	caaatecttg	tgccgaggag	3180
atgccaagag	tctcgttggg	ggagatggtg	agggggcgat	acaggggaag	agcaggagga	3240
aagggggaca	cactggtgtg	ggtctttgga	gagctcagga	gaatagcagc	gatcttccct	3300
gtccctggtg	tcacctctta	cagccaacac	catthtgtgg	cctggcagaa	gagttgtcaa	3360
gctggtcgca	ggtctgccac	acaaccccaa	tctggcccca	agaaaaggca	cctgtgtgtg	3420
actctggggg	taaaggcctg	gcctggtcgt	ctccagctgg	acttgaaact	cccgtttaat	3480
aaagagttct	gcaaaaataat	accgcagag	tcacagtgcc	aggttcccct	gcttctctga	3540
agcgcacagg	acgggttccc	taggaaatgg	ggccttgctt	gccaagctcc	cacggcttgc	3600
cctgcaaacg	gcctgaatga	tctggcactc	tgcgttgcca	ctgggatgaa	atggaaaaaa	3660
gaaaaagaag	aagtgtctct	ggaagcgggc	gcgtcacac	aaacccgcaa	cgatgtgta	3720
aacactctcc	attgagaatc	tggagtgcgg	ttgccctcta	ctggggagct	gaagacagct	3780
agtggggggc	gggggaggac	cgtgctagca	tccttccacg	gtgctcctg	gctgtggtgc	3840
atgcccggaa	ccgaaacggc	gaactaaagt	caagtcttgc	tttggggaa	ctgacaatca	3900
acgaaatcac	ttcgtattgt	ttcctctttt	tactggaatt	cttggatttg	atagatgggg	3960
gaggatcaga	gggggagggg	aggggcgggg	agacggaggg	aggaggggag	gaggggagga	4020
gggggagggg	ggaggagggg	aagggatgga	ggaaaatact	aacttttcta	attcaacatg	4080
acaaaagatt	ggagaaagtg	caccgctagt	gaccgggagg	aggaatgcc	tattgggcat	4140
tatattccct	gtcgtctaat	ggaatcaaac	tcttgggtcc	agcaccaagg	attctgagcc	4200
tatectattc	aagacagtaa	ctacagccca	cacggaaagag	gctatacaac	tgaagaaata	4260
aaatthtcc	tttatttcat	ttctgtgact	gcatgttcac	atgtagagag	ccacctgtgt	4320
ctaggggctg	atgtgctggg	cagttagagt	ctgagcccgt	taactggaac	aacccagaac	4380
tcccaccaca	gttagagctt	gctgagagag	ggaggccctt	ggtgagattt	ctttgtgtat	4440
ttatthtagag	acagggctct	atactgtagt	ccaagctagc	ctccagctca	cagaaattct	4500
cctgttccgg	tttccaaagt	actggagttt	tgagtgtgtg	ttaatgaaac	gctaagaatt	4560
tgctgattga	agaaaacctc	aagtgggttt	ggctaatccc	cacgacccca	gaggtgagg	4620
caggagggat	agagaatttc	aaggtttgcc	agagccacag	ggtgagctca	atgtggagac	4680
tgtgaggggt	agctcaatgt	ggagactgtg	agggctgagc	caatgtggag	actgtgaggg	4740
tgagctcaat	gtggagactg	tgaggggtgag	ctcaatgtgg	agactgtgag	ggtgagctca	4800
atgtggagac	ctgtatcaag	ataataatag	tagtagtaac	aatgcaggcg	aggggtggtt	4860
tgagtggtag	agcagttagt	tgatttgaca	tgcttgaggt	ctcccggctc	atctgtggcc	4920
ctgcaacagg	aagggagggg	ggaagggggg	gaacgagaga	gaggaagag	agacagaagc	4980
taagataggg	aatgagagag	gaaggaagaa	acgggaagaa	attcagactc	cttctgagt	5040
tccgccaca	cctagtgaca	tcctgtgcac	acctaaagt	ggcctttgtg	tggcactggc	5100
ttgggtggtc	gggaaaggca	tttccagctt	gttgcagaac	tgcacagta	gcatgctggg	5160
tccgtgaaag	tttctgcccg	ttacaagaa	gtctctacta	cttgtgaect	caccagtga	5220
aatthtctta	attgtctcct	ggtgtctctg	gttttgcatt	tttgttctta	aggatacatt	5280
cctgggtgat	gtcatgaagt	ccccaaagac	acagtggggc	tgtgttgat	tgggaaagat	5340
gatttatctg	gggtgtcaaa	aggaaaagaa	gggaaacagc	caattgqaa	aatgtccctc	5400
cgcccaccgg	aatthtggct	tggcaaccgt	ggtggaggag	caagaaacac	gtggagcttt	5460
gaggaggcat	gggtcctag	gaggacagga	agcagaagga	gagagctggg	ctgacagcct	5520
gcagycattg	cacagtttca	gaaggagatt	acagcatgac	tcaattttta	gggatccaac	5580
agggacctgg	gtagagattc	tgtgggctct	gaggcaactt	gaectcagcc	agatggattt	5640
tgaataacct	gctcttagag	ggaaaacaga	catagcaaac	agagccactg	ttagtgatga	5700
aactctcact	ttgcctgagt	catgtgcggc	catgccaggy	ggtcaggctg	acactcaact	5760
caaaaacaag	tgagaaattg	aagacaatcc	gtggtggcag	ctactggaag	ggccaccaca	5820
tccccagaaa	gagtggagct	gctaaaaagc	catthtgtgat	aggeacagtt	atcttgaatg	5880
catggagcag	agattacgga	aaaatcgaga	atgttaatga	ggcaacatc	gagttgagtc	5940
attcagtgat	ggaaacccag	acgcttccat	cccataaaag	gaacatcttg	ctctcagta	6000
aaatggaat	aaaaattggg	gcttgaattt	ggcaaatgat	tcagaactct	gtgtaggta	6060
tttcacacgc	acagtgagata	atthtcatgt	tggagtttat	ttgtgctaaa	aggcagaaaa	6120
gggtaaaaag	cacatcttaa	gagttatgag	gttctacgaa	taaaaataat	gttacttaca	6180
gctatctctt	aattagtagc	cccttccacc	tgtggtaatt	tcctgagata	gtcagtgagg	6240
aaaagatctc	tccttctctt	ctttctcccc	ctcccctct	ctcccctct	ccctccctcc	6300

ES 2 522 615 T3

cteectctcc tcectcctcc cccctttcct tctttctttg ctcectctcc tctgectcct 6360
 tctcccttcc ttcttcattt attctaagta gcttttaaca gcacaccaat tacctgtgta 6420
 taacgggaaa acacaggctc aagcagctta gagaagattg atctgtgttc actagcgtgc 6480
 aattcagagg tgggtgaaga taaaaggcaa acatttgagg ccatttcctt atttggcacg 6540
 gcacttagga agtggaaat gectaatact ctggtttgta ccacctttcc ctataatgga 6600
 ctglttggga agctcctggg caaccgatcc tggcatctca ttggtcagag gcctgttaaa 6660
 tggtagctct atttgcaaaag aaggctgtaa cttgtagctt taaaagcctc tcctcaagaa 6720
 agaagggaga aaggatatgg ctagacatat ctaatagact taaccactgt gaaaagcctt 6780
 agtatgaatc agatagaacc tatttttaac tcagttttga aaaaaataat ctttatattt 6840
 atttgtgtgt gtgtgtgtgt gtgtgtgtgt gtgtgtgtgt gtgtgtgtgt gaaccacatg 6900
 tagcaggtgc tggaggaggc cagaagaggg caccagatct cctggaactg acaccacaca 6960
 tggttatgag ctgctgatg tgggtgctgg gaactgaact ctctgttct gcaagagcag 7020
 caactgttct ctttaactgat gagccatctc tccagcccc cccataaatt taattgttca 7080
 ttttagttaa ttttattcat aatcaattat cacagtataa aacaatgatt ttatatatat 7140
 catatacata tcaaggatga cagtggggg gatattgtgt tgtgtgtgtg tgtgtgtgtg 7200
 tgtgtgtgtg tgtgttattt gtgtgtgtgc ttttaagaa ggtgccatag tcaactcatt 7260
 tctctgaagg atttcaaagg aatgagacat gtctgtctgc caggaacct atcttctct 7320
 ttgggaatct gacccaaag aggtattctg aggaactgaa tgaagagctc aagtagcagt 7380
 gtcttaaac caaatgtgct gtctagagaa agtcaacgct atcagtgagc tgaggagaga 7440
 tttactgagc ggaagacaag cgtctttga tttaaagtgc tgaacagtc acggctgtgg 7500
 agtggagcct gtgctcaggt ctgaggcagt ctttgctagc cagctgtgat gagcagtga 7560
 gaaaggggtg agatggaggc aggggtggag cagggctatg gttcagacta ggtatcgtga 7620
 gcacaccagc tgggtgactt gtggtctgtg ggtcaggcgt tgtaaacgcc ctgagggtca 7680
 ggcagtcaaca ttgcttgaag ctgaatgggt gaggcaacac agagagtga aagaaggcaa 7740
 agtaccacct ctccccgac ccaggtcaact tetgggttat agctgagact ccggacagca 7800
 tgaaccagc tgggttagagc ttcagggaaa acttgatgct tgcattgtgc tatgaaatgt 7860
 gattcggtac atctggagaa aatttataat gctggctcag tcaagcactg acaaaagga 7920
 ccttggcttt gggagctaca tgacattgac ttgtaggcag actttttttt ttctgcccgc 7980
 caattcccag ataaccaata tggaggctca atattaatta taaatgctcg gctgatagct 8040
 caggcttgtt actagctaac tcttccaact taaatgaacc ctttctatt atctacattc 8100
 tgccacgtga ctttaccctg tacttctctg ttctctcct tgtctgactc tgccctctctg 8160
 ctccccagag tcttagtctt ggttctcctg cctaaccctt tectgcccag ctgctgacca 8220
 agcatttata attaatatta agtctcccag tgagactctc atccaggag gacttgggtg 8280
 ctccccctc ctcaatgcca tccgtgtctt cctcttccct cgttcccc ctctctcct 8340
 gctcttctc ctccacctt cctttcatag tattgatggc aagggtgttc tagaatggag 8400
 gagtgcccat aggcattgcaa agaaccaggt taggatgctc tgtgaggggt tghtaatata 8460
 agcgtaggac acaattcaag ccacagagtg aagacggaag gatgcaactg gctctagagc 8520
 aacttctggg gcagaatcac agggtaggt tetgacttga gggcgaagag gccacgagga 8580
 agggagtgag ttgtctgag ctagaagcta cggcccacct cttggttagca gacctcccc 8640
 caagcatgct ttgttaataca tgtgggatct gattttctc taaatctatg ttcaactctt 8700
 aagaaaatgt gaattctcac attaaaattt agatatacgt cttttggtgg ggggggtgta 8760
 aaaaatctc aagaatatgg atttctgggg gccggagaga tggctcagag gttaaagaaa 8820
 ctggttctc ttctagacat tetgagttca attcccagca accacatggt ggtcacaac 8880
 catctgtaat gcgacctggt gccatcttct gacatgcatg gatacatgca ggcagaaagc 8940
 tgtatacata gtaaatgat aatcttttt ttaaaaagag tatggattct gccgggtgtt 9000
 ggtggcgcac gcctttaatc ccagcactct ggaggcagag gcaggtggat ctctgtgagt 9060
 tcgagaccag cctggtctat aagagctagt tccaggacag cctccaaagc cacagagaaa 9120
 cctgtctctg aaaaacccaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa gtatggattc 9180
 taagaaagcc gtaacagctg gagctgtgta cggagttcag cgtggtacta gaagaacaga 9240
 cattcatgat gaaacacccc aggatttta cttagtatct agtttccatt gttgttttga 9300
 gaccggtctc tatgctctcc aggtggcct caaactgctg atcttcccgc ctctacctct 9360
 caagtctgg gactacttgg ctcataaaac agtttttgtc gggtccctg aagttatggt 9420
 tgtacaaacc gtgggggtca atatactcac ttgggcagag agagaaggct tgaatcccag 9480
 acaatgactg catctcagga cagtggggaa gaggacaatg gcagaaggac ttgaaaaga 9540
 tagactggag ggtggaanaag cagcaggaac agaaaacaa aacaggaagc ttgctatcca 9600
 gggccactct ggagctcctgt ggcaagatgg aagcgggcta ggggaataca tttgtgctac 9660
 tgtgtgtgtg tgtgtgtgtg tgtgtgtgtg tgtgtgtgat caatgctat caatgtgaa 9720
 ggggaaatgt gtataccaca ttgattctgg gagcaattct cagtatctgg cctagagaaa 9780
 ggaatggccc ctgcagaata gacagagtga atggtgccc ttatcatttg ctaaaagtga 9840
 ggagaaataa acatccttcc atagagtttc aggtaaatga accccacagt tcatctgtgc 9900
 cgtggtggag gcctggccaa cagttaaaaa gattagacac ggacaaagtc tgaaggaac 9960

ES 2 522 615 T3

acctgaata ggaagaggag agccacctca ttctgttaact ttctcaagg ggaagatgtt 10020
 ccaagagtgg gaataaatgg tcaaaggggg gatttttaaat taggaaaaacg atttctctgta 10080
 tcacttgtga aactggaggt tgatttgggg cataggacaa tagatttgat gctttgcaa 10140
 aagctgtttc aaagcagaga aatggaatag agacaattat gtagcgagga gggagggtgg 10200
 ggcgaagatg gagacagaga agtggaagct gactttaggg aagaggaaca tagaccacag 10260
 gggcgggggc gggggcaggg gcggggggcg gggctcaaag gaggcagtgg gaacgttgc 10320
 agtgttcgca gcgtaagcgt gaatgtgcaa gcgtcttgt ggtgtgtgac caggagtagc 10380
 ttgggtggct tgtgtgctgc ttgtaatccc agtctttgag gtttccacac tgttccacag 10440
 tgggtgtgat tttccctcgg agagcatgag ggctctgctt tccccacac cccccagcg 10500
 ttogttggta tttgtttcca agatgttagt gggtgagaca aagcctctct gttgatattg 10560
 cttaaacagg tgacaaaaaa agctcaacca ggagacattt ttgccttctt ggaaggtaat 10620
 gctcccattg agagcaatgg gacccatctc taaggtgagg ctactcttgc agtttgacc 10680
 cagctcttct gatgcaggaa ggaagtgggt gggcaagcaa gactgattgc ttcttgcag 10740
 ggacacattc tgcacacaaa ggctcaggag gggagaaggc tgtttgatgt ttgacactca 10800
 ggaaggcccc tgatgcatct gtgattagct gtctccatct gtggagcaga cacggactaa 10860
 ctaaaaacca gtgtttttaa atgtcaagc cttaaggtg aggaaattga cttattgtgc 10920
 tgggccatcc gttagagcaag tgctctgcat tgggccaaac cccggctctg gtttctagc 10980
 accagaatgg cctagaacta actcacaatc ctccattcc aggtctcagg tgctagaatg 11040
 aaccactata ccagcctgoc tgcctgcta cctgcctcc taaattttaa atcatgggga 11100
 gtaggggaga atacacttat cttagttaggt gtttctattg ctgtgaagag acaccatgag 11160
 catggcaact cttataaagg aaaacattta gttgggtggc agtttcagag gttttagtag 11220
 atgtgcatca tggctgggaa catgatggca tgcagacaga catggtgctg gagaaaagga 11280
 tgagagtcct acatcttgca ggcaacagga cctcagctga gacactggct ggtaccctga 11340
 gcatagggaa cctcacagcc caccctcaca gtgacatatt tccctcaaca aagccatacc 11400
 tctaataagt gccactccct atgagatgac agggccaatt acattcaaac tgctataaca 11460
 ctttaaagta ttttattttt attattgtaa attatgtatg tagctgggtg gtggcagccg 11520
 aggtgcacgc ctttaatccc agcacttggg aggcagaggc agatggatct ctgtgagttc 11580
 aagaccagcc tggctataaa gagctagtty caaggaaagga tatacaaaaga acagttctag 11640
 gatagccttc aaagccacag agaagtgtct tcttgaaaac caaaaattgt gctgggacct 11700
 gtctctgctt tggttgcttc ccactccccc agagctggac tcttggtaaa cactgaatca 11760
 gctgcaaaat aaactcctgg atcctctct tgtaacagga gcccgaaagc aggcgccac 11820
 ttgtcttctc gcaggattgc catagacttt ttctgtgtgc ccaccattcc agactgaagt 11880
 agagatggca gtggcagaga ctgggaagyc tgcacgaaa acaggaagtt attgcacct 11940
 gggaatagtc tggaaatgaa gcttcaaaac ttgctcatg ttcagttgta cacagactca 12000
 ctcccaggtt gcaacacagc tghtaaatatt cctgactatg tctgactgc ttttatctga 12060
 tgcttcttcc ccaaaatgcc aagtgtacaa ggtgagggaa tcacccttgg attcagactc 12120
 cagggtcgtc ctccctaacc tggacttgtc ttctccggc agcctctgac acccctcccc 12180
 ccattttctc taccagaagg tctgagcaga gttgryggcac gctcatgtcc tgatacactc 12240
 cttgtcttcc tgaagatcta acttctgacc cagaaagatg gctaaggtgg tgaagtgtt 12300
 gacatgaaga cttggtctta agaactggag caggggaaaa aagtccgatg tggcagcatg 12360
 taccgaaat cccagaactg gggaggtaga gacggatgag tgcccggggc tagctggctg 12420
 ctccgcagc ctagtgaat tgccaaatc caactcctat tgaaaaacct ttaccaaca 12480
 acaaaacaaa caaataataa caacaacaac aacaacaac taccaccatac aaggtggggc 12540
 gctcttggct cttgaggaat gactcaccca aacccaaagc ttgccacagc tgttctctgg 12600
 ctaaatggg gtgggggtgg ggcagagaca gagacagaga gagacatgac ttctgggct 12660
 gggtgtgtg ctctaggcca ccaggaactt tctgtcttg ctctctgtct ggcacagcca 12720
 gagcaccagc acccagcagc tgcacacacc tccctcctg ctctctgagc aaacacaggt 12780
 gccttggct gtctattgaa ccggagttag ttcttgcaga tgtatgcatg gaaacaacat 12840
 tgtcctgggt ttatttctac tgttgtgata aaaaccggg aactccagga agcagctgag 12900
 gcagaggcaa atgcaaggaa tgctgcctcc tagcttgcct cccatggctt gccgggctg 12960
 ctttctgcaa gcccttctct ccccatggc atgctgaca tgaacagcgt ttgaaatgct 13020
 ctcaaagtgc actttcaaaag aaggtctctc gtacttctgt ctgtctgtct ctgtctatct gctatcactc 13140
 tatcaatcat ctatctatct atcttctatt tatctaccta tcattcaatc atctatctc 13200
 taactagtta tcaatttatt atttgtttac ttacttttt tatttgagac agtatttctc 13260
 tgagtgcag ccttggctgt cctggaaccc attctgtaac caggctgtcc tcaaacacac 13320
 agagatccaa ctgcctctgc ctctctgggt ctggggttaa agacgtgcac caccaacgcc 13380
 ccgctctatc atctatttat gtacttatta ttcagtcatt atctatctc taactatcca 13440
 tcatctgtct atccatcatc tatctatcta tctatctatc tatctatcta tctatcatcc 13500
 atctataatc aattg 13515

<210> 6
 <211> 14553
 <212> ADN
 <213> *Mus musculus*

5

<400> 6

ES 2 522 615 T3

cttgaagaac acatgttttc caagagggag caccatggt ggaatgacaa tqtatgtagt 60
 gctcctctcc tgtagggttag tgctcctttg ctataggttaa gtgctcctct cctataggtc 120
 agtgctcctc tcctataggt tagtgctcct ctctataggt ttagtgctcc tctcctacag 180
 gttagtgtc ctctgtctc ggtagtctc gctcctctat agtacctaga gagctagggc 240
 aatgggcta ggccgaagt gcagagacaa acagctatgg aagactgggt aagcacttcc 300
 aagctacgaa agagcagtgt gaagggctcag ggctgtgca gttagtaggg gagatcttcc 360
 agttgaagaa acagaagaac tgagagccac tgggtatcat cctcctgccc catgccttcc 420
 tggatactgc catgctccca ccttgatgat aatggaatga acctctgaac ctgtaagcca 480
 gccccaatga aatattgttt ttatgagagt tgccctgggtc atgctgtctg ttcacagcag 540
 taaaaccctg aataaaggcag aagttggtag cagtattgct gtgatagacc tgaccatgct 600
 ttcctttgaa agaatgtgga tttgggtgact ttggatttgc aacacagtgg aatgctttaa 660
 atggagatta atgggtcctc aattcctagt aggaatatgg aagactttgt tgcctggagt 720
 atttgaactg tgttgacctg gcttaagaga tttcaaagga gaagaatttc agaattgtgc 780
 ataaagacag tttttgtggg attttgggtg agaattgtgc tactttttgc ccttgtctga 840
 aaagtctgcc tgagactaaa gtgaagagaa tcagattaat tgcattgaca aggggaagtt 900
 gtggctgccc tatctggaaa cttacagcca gctccttggg cctcgggtga cttacgcaaa 960
 tactcagggg cagagatgct tgactctgta ctgatgagtt gtcttggatg caaatatggg 1020
 ctcttcattt gactacatgt cactgatgagt caggagctgc tctctccaga gtgtgacaaa 1080
 gcgaggggat gctgacggtg gctgttctag ctttgaaggt aagcctgcac ttatgctaaa 1140
 gtcacacata cagagccggg gtggagaacc tgtctgtgtg gagacacctt tcattactgt 1200
 tggcatccag cctctcaagg ttggactgccc tgtgtgctcc tggactctgg aggtcccact 1260
 gctctgtcct ctgctgctta tgatactgac atttttaaag aatccagtgg tccccctg 1320
 tactcgggtg ctacttctac ctggatgttc ctcatttatg ttctgtgaca cttctctgtg 1380
 actctgctgc attcctgggt gacatgtgga caccctgtcc ctttgagcac catgatgtca 1440
 ctgtcactag tggaaatcaga tgcctcaagt gttgtcctgt gtttgggaac gtgacaggca 1500
 gtacagaagc agaagaggaa ggggtgaaaac ggaaatgtca cagcagcctc tgatgtgtgc 1560
 ctcagtcacg catgctgctg attggaacta ctcagcatga gagagggcca tgggtgaatac 1620
 acaaccctat acacactgtg tccatttctc tctctctctt acacagagag agagggagga 1680
 gggggagggg gagggggagg gggaggggga gggagagggg gtgggagagg gagagggaga 1740
 gggagagggg gaggggaggg gagaggggaga gggagagttt aatgtctgtg aagagatacc 1800
 atgaccaaaag caactcttat aaaggacaac atttaattgg ggctggctta caggttcaga 1860
 aattcagttc atttctacca tgggtgggaag catgcaggta gatgtgggtgc tggaggaacc 1920
 aagagttcta tatcctgata tgaaggcagc caggagaaga ctgcctcttc tgcacagggc 1980
 agagcttgag catagaacat caaagcctt cccacactt cctccaacaa ggtcatacat 2040
 acttcaacaa agacacacct cctaacggtg cctcctctg tggaccaacc atttaaaccg 2100
 atgagtctat gaggggtcaaa gctcttcaaa ccaccacact catgtacaca cacacacaca 2160
 cacacacaca ctctcataca cacacacaca cactctaca cacacacaca cacacacaca 2220
 cacacacaca ccacacacac acacacacac agagttctat tttgcaactgt ttcactgtca 2280
 caaggttcta cttatctcag acacactgcc aggaattgtg tgggaagact ttcagttctc 2340
 ttgggttcac atggacttag cagttcttgg tgatcctgaa agatttctgc agaaagaagc 2400
 caaagtgttg agcccaaggc ctggccacac attagtctct tctagatgaa caggggttta 2460
 aaaataaggg ggcatacaagg tgaagccagc aggggctgac tttagagagga gaccaccca 2520
 agccaactgc tcgaagtcaa aagcagatgaa tcccataac cagctgtgcc cgggtgctgtc 2580
 ttgctacatc tttagtaaat gttcttttag ttgtatgctc atgaatattt' tgettgcata 2640
 tattttgtgta caccataggt gttcctaggg cctatggagg ccagaagagg gcatcagatc 2700
 ctttggaaact ggaattatag acacttgta cccatagagt agattgtggg aatgagcct 2760
 ttagtctctg agagcggcca gtgctcttaa ccttgggtcg tttctccagg tctttgagac 2820
 tttatttctc tggacatcag gacaggatcc agggctttga gcttgtttct tcagccagct 2880
 ttcttttcat gtatattaaa ttttatgtta ttttgccttc tttttcccca agacagaatc 2940
 aactctata tagctcaggc tgggtttgaa ttcagtttcc ctgtctcagt ctaccgggtg 3000
 atatgattac agatgtgagt ctgactttgg tatcaaagtc cccagcctt ctggatatgt 3060
 gttttaagga tatcagatat atccttgatt tgccttgaat tttcttttta gttacaacat 3120
 aattagttcc gtgtcacctg aatatgtgta tgtcacctac atagtcttcc tcttctctt 3180
 ctccctctc ccaccttccc aggtacctgt ctgtcttcat atccttgtgc tgagagtctt 3240
 gttgagggag atgatgaccg agacagagcc actggggaag ggagatgggc tagtgaggt 3300

ES 2 522 615 T3

cttcagagag gagctcgtga atattgtagc ccctttagtc cctggcatgt cctcttgtat 3360
agccaccgcc atgctgtggc ctggcagaag tgaataagtt gtccagctgt tgacaggcct 3420
gcoctccaga cccagctcga tcccaagaaa gggcatctgt gtctgtctct gaggccgtaa 3480
gtgctgcoct gttgtctcca gcttgacttg aactccctc ctttaataaga gtaccacaga 3540
acagggctct cagagtcctt gggccaggct cctgtgctgt cctggaatgc caggcgtgaa 3600
tttctctgtg agtaggactt tgetcgccaa gctcccacgg cttgcccttc agatagccag 3660
aattatctgg taccctgcat tgccgttcaa tacgcagagt atcaactgaa gcgcgcgcgc 3720
gcacacacac acacacacac acacacacac acacacacac acacgcccac tccatcttta 3780
aaccccaccc cccagcaacg gcggtgtaaa cactctccat caggaagctg aaacgcagt 3840
gcoctctgct gggagatga aggcagcttg ctggggcgga ggaccgtgct agcaacctc 3900
cctgggtgcac acgggctctg gtgcatgacg ggaacggaaa cgcggaacta aagtcagtcc 3960
tgcttttttt tttttttttt tttttttttt tttttttttt tttttttttt ggcgttgggtg 4020
gtggactgag tgacaatcag tgaatcaact taggttgttt ttctctctct cgttgggttt 4080
gatagaccgt gggagagggt cagaggagaa ggggagggat ggggagagag ggaggagga 4140
ggggcgggag gcggggggcg aggaaaacgt gctaaactct ccaatcctac aagacaagg 4200
tttgagaaa gccgcactga gtgaccacgc agaaggaatc caggaatgtc cgctggaatc 4260
tgactgttga tccagcgcc atgcagagaa tctaggctgg taggaacatt cttgtccta 4320
tccagcataa taactccaac caacacggaa aagaaaggct atacaagtga agaaaggca 4380
ttttcacttt catgactata caatcacttc caggtagtta cacgtgtcta gcacagcgt 4440
tctcaacctg ggggtcacga tccccactt ttctgcatat cagacatttt tacgttgtta 4500
ttcataacag tagcaaaatt gcagctatga agtaacaatg aatgcatttt atggtgctg 4560
tgtgtgtgtg tgggggggta tcaccttaac atttactgta agaaggttga gaatactgct 4620
ccagcagcta gtgtgttggc cttaggttct gggtatatta ttagcaatag ccaaccagaa 4680
tccccaccca ccacagcatt gaggccccat gcagggcttg ctgggagagg cactgataag 4740
acttctttat gtatttattt agagacgaat actcattagg taggccaagc tagcgtcaa 4800
ctcatggcaa ttctctcct ccagttcct aagtactgga ctcaggagtg tgttgccatc 4860
atatacagta aggatttatt gactgaagaa aatctcaagt ggctttgggtt aatccctact 4920
acgccagagg ctgaggcagg aggcgcgcaa ggtcaaggct tgctgggct acatatagag 4980
tgagctcaat ttgacactt ggtgcggtgt tagtagtaat agtaaatag aaggtgtggc 5040
tcaggtgggg ccggtgattg gacacacttg gggtctcctg gtccatctgc agctgtgcaa 5100
caggaagagc gggaatgag aggaaagaga gaaaagacag aatgagagag agggaggag 5160
agagaaaaag gaaaagagag aggaaaggaa aaaggaaaat gaggaaaagc agaaagaaga 5220
aatgagaaa agaaaaggga gaaagaaatg agagagagaa aagaaaagac agaactgcag 5280
agagggagga agagagaaaa aggaaaagag agagaaaagg gagaagaaag atgagaaaag 5340
cgagaaaaga gaaatgagaa agaggaaaag gagaaagaaa tgagagagag aaaagaaaag 5400
acagaatgcg agagagggag gaagagagaa aaaggaaaag agagaggag ggaaaaagga 5460
aaatgaggaa agcagagaa agaaatgag aaagagagaa aagagagaaa gggagaaaaga aatgagagag 5520
agaaaagaaa agacagaatg cgagagaggg aggaagagag aaaaaggaaa agagagagga 5580
agggaaaaag gaaaatgagg aaagcgagaa agaagaaatg agaagagga aagggagaaa 5640
gaaatgagag agagaaaaga aaagacagaa tgcgagagag ggaggagag agaaaaagga 5700
aaagagagag gaagggaaaa tggaaaatga gaaagcgag aaagaagaaa tgagaaagag 5760
gaaagggaga aagaaatgag cgagataaaa gacagaattt gagagagggg gaaagaaat 5820
ggaaaagaga ggaaaaggat gagaaaagag agaaaagaa agagatgaaa gagagaaaag 5880
agaaatgaaa tgagagagag agagagacac aaagagccag agagagagaa aaaaagggga 5940
aagagaaaag gaaagaggaa ggctcctctt ggacacatct tcctttatct tccctgggg 6000
accgccaaa cctggtggca tactgtacat tctgtacact gttcattcaa aacaggctct 6060
gtcttaaga tggtctgagc ggtcagaaaa gggtattgtt aacttgtttg caaaactgcc 6120
tcaggagagt gctgagtgcg tgaaagttag tgcctgtaa ggagaagtct ctactacttg 6180
tgactctacc atcgaaaatt tctttaattg tctcctggtg tctgggttt tgagttttg 6240
tttctaagga tacattcttg ggtgatgtca caaagtcccc aaagacacgg tggagctgtg 6300
ttagatgggg aaagacagtc tgetgaggat ttatctggaa ctgtcagaag gaaaagaagg 6360
taaatggggc acttgggaaa gtggcctcta gttgacttc tggcttagca aaggttgtgg 6420
ggagataagg catacacagt agttagcagg aggcaacagg gtccctgggag gacgcgaggc 6480
agaaggagag gctgggctga cagcatgcaa tcattgcata gtctccaaag gagattgcaa 6540
catggctgag ttttcagagg toctacagag ccggtggtag agattctgtg ggttctgaga 6600
caacttgact ttagccagat ggtatttgag taatctggga gagagaaaac agctacagca 6660
aacagggcca catttagtga cgaactctc aacttgactg ttgagtcatt tgcagtgggc 6720
cctgaggtca ggctggcct cagctcaaaa actaagcagg aactgaagca attactcaga 6780
taateccacg ccacagccac tggaaaaggc cacatcccca gagacagcac agcaggggtg 6840
gggggtggggc tatgagaaa tttagtgatt tagcagttat ctagaatgtg cggagcagag 6900
gaggttacac aaaaacctag aatgtcattc aatgtgggaa accgagaggc tcccaagccc 6960

ES 2 522 615 T3

taaaaggaac agtttgcttt cagccaaaat ggaaataaaa tttggggctt aaatctggca 7020
 aatgattcag accttctgtg taggtgtctt taaatgcaca gcagattgat tttcatgttg 7080
 gagtttattt gaactaaaag acagaaatgg tgaaaagcac acctgaagaa attgagatgc 7140
 tatgaataaa atcatttact tacagctatc acttaattag tacctccttc caccttgcgtg 7200
 atttattggg ctagtcaagg aagaaaagat ettccctcct ccttctctcc tccctccctc 7260
 cctctcctcc tccctcctcc tccctgaact tccctctctc cttttccctc cccccctct 7320
 tcttctcttc acccctctct cccctccctc cctctgtact cctcccttt cctcccaatc 7380
 tcttctcttc cccctctctc ctcttctctc cccctctctc tccctctctc tccctctctc 7440
 ctccctcctc cctctcctcc tccctctctc ctctcctc ttctcctcc tccctctctc 7500
 cctctcctct tccagcctc acctaccttc ccttctctct tcatttatte aaagtagctt 7560
 tgaacagcac tactcggttt agttgtgtat aaaaggaaaa tgcaggcca agcagcttgg 7620
 ggaagattgc tttttgctct ctggaggcag atgatgacag tcaagatca ttccttttgc 7680
 tccatgtcac aggaaggggg acatgccgaa tctaccagtt tgcagccacc tacacaggat 7740
 ccaccttcac ttctaaggaa atgtttggga agctacctac caaccacttc tggcatctca 7800
 tgggctagag gactttaaa tggcactctt atttgtttaa taaaggaggt tgtgacgtgt 7860
 agttttaaat cctctccaca caacaattgc tactctctga ccaaaaaaga agggagacag 7920
 gataccgcta ggtgtctagt agactttacc actttgaaaa gccttaatata aatcaggta 7980
 gatacatctt ttttaacttat tcttgtaaag acaaaaaaca aactttattt ttatttgtgt 8040
 gtatgcttgt gtgtgtgtgc ctgtgtgtat accacatgtc gctggtgccg gagaacacca 8100
 gaagagggga cctgatctcc tggagctaaa gctatccatg gttctgagct gcctgatgtg 8160
 ggtgtctggga acagaactct ggtcttctgc aagagcaaca agcctcctct taactacgaa 8220
 tctctctccc atcccccaa atacatttaa ttattcattt tagcagcttt atttcgtaac 8280
 tacttatcac agcataaaac aaggatttta tatatttac atgcaatcga ggataagagt 8340
 tgaggggaga tgcgtgtgct ccttctgggt tctctgtgct ttgaagaatg taagcagtgc 8400
 acaagggacc gaggcgtgcc tgtctgccag gagctgtctt ctcccttgg actctgagct 8460
 gagtgcagtg ctccgaagaa gtaaaagacg acctcatgaa gcaatgtctt caaccocaaac 8520
 atgctgtcca gacaaagtcc agcttcatta gtgctctgag gagagactta ctgagctca 8580
 ggaaagcccc cctcagcatg gcgaaagtcc actttgattg aagtgactcg aaagccatgg 8640
 cagtgcggcg gcggcgcgct ggagcttctg ctcgagctcg aagcggcatc tttgtcaggc 8700
 ggctgtgatt agcacgggga ggcaggactg gagtgaagga agagtgggg gcggggctta 8760
 gegtctgggt ctctcaagct gtagtccagc cctcaagatt tctaacctgc cttctgctt 8820
 cccagccagg cagtcaagtg gctccaagct gaagactgca aagtgcctct aaccttttgg 8880
 ttatagcgag gctgaagaca ccgtgctctt tcatgaaagc cggatgtctg aaatccgatt 8940
 tgataaatat ggataaaacg tataacgctc gatcaatcga atcgaaggag ctacagattg 9000
 gcaccacggc tttggggaca acagagtact gactcgttgg gaggacttgg atacttccc 9060
 tctcttcca tctcttccc tttctcact tctctctct tctctctca ttttctcct 9120
 ctctactgtt tcttactatt tttacaaaag attttattta ttattttatt tttatttata 9180
 tttatttatt tatttattta tttattaat gtatgcagat acactgtagc tgtttttaga 9240
 cacaccagaa gagggcgtca agttccatta gagatggttt cgagccacca tgtggttgc 9300
 ggggctctg gaaggaccgc cagtgtctt aacctctgag ccatttctcc agtacccttc 9360
 tcaccgttcc tcttcaatct tcttctctt ccttctccac tttctctgtc ttcttggttt 9420
 cattatcttt cctcctttct tctcttctc ccttctctc tctctcactg tagtttctc 9480
 tccctactct tttctgctt cctctctct cctctctcat tccccctct cttctctct 9540
 tetctctct cctctctct tctctctct cctctctct tetctctct cctctctcc 9600
 cctctcttc ctcttctct tctccacce ctctgtcac agtataaatg gcaagggtgt 9660
 tctagaatgg aggagtgtcc cctaggcact aacgaaagcc agttaggatg ctctgagacg 9720
 ggtacaatcc agggagggcc gtggggatgg aagggttctg ctggatcca ttctggagca 9780
 acccccaggc agaatcatga ggttggttcc ggattcgcag ggcacaatc agaagaggaa 9840
 ggtttcagga aggacagatt tgtctgagat aggagttaca tctgatgtct tggcagcaga 9900
 gccactgtac aagcgtgctt tattaaccac gtgggattaa atctctttt aaatttattt 9960
 tcaactctta aggaaacgtg aactttcaca ttcaaattda gacttgcage tcttatgggg 10020
 aaaaaagggt gatcttaaga atattaagca taggcggctg gagagatggc tcagcggtta 10080
 agagcactct ctgctctccc agaggctctg agttcaattc ctacgaacca cataatagtt 10140
 aacaacagtc tttaatgaat tctaatgccc tcttctggtg tgtctgaaga cagttacagt 10200
 gtactcatat aaataaaata aagaaattta aaaaaatgaa tattaggcat agattcctgg 10260
 atcctaagaa agccatcaga gctggagcca tgtgtgggat cctgcttggg gctggagggg 10320
 cagagttcat gccccgggg ttttactta ttatcacatt ttcatcgttg ttttgaaaca 10380
 gggctctgtg tggctcaggc tggcctttaa ctcatcttc agcctctacc tcacaggttc 10440
 tgggattact tggttcctaa aagtatctcc gtcaagctcc ctggtgtat ggtgtgcca 10500
 accaggagg ctgtataact cgtcaggta gaggagaag atccgaatct ctgacagggg 10560
 ctgctgcctc tcggggcaaa tggagtgaag gacagcggca gaaggattta ggaaagatgg 10620

ES 2 522 615 T3

acgggagagt ggaatgctg cagaagccag aaaacaaagc aggaagcctg ctgtccagtg 10680
 gggctcaaga gcggagggat gcgagggggc tgcgcaggaa catttagcgt ctgectctat 10740
 gggggtaggg gcggggtgcc agcacctagt cacctgaagg ggaatgctt gccaggggag 10800
 caggctctcag tagctgacct agagaaagga gcggccccta cagaggagac acgggtcact 10860
 gtttgttaaa gtgaaggaga aataaatatt ctttcaaaga atcttaggtg agcccagttc 10920
 atctgcgctg tggaggcctg ggaacagtt aaaaagacc tgacacacac ccaaggcaaa 10980
 caagcaacac acggctcctt ccgtaagggt ccatgattct ctgaagaatc agccccggaa 11040
 tcagcccccg aatcaggtag tccgtaaaca caaataga gggatagtca tttactctgc agaagtcag 11100
 cctgctggcg tctcccatta ccaaataga gggatagtca cgtgagctca cggctcgat 11160
 ttaaggcacg tggttttcca gggtagatga gotttggctt ctggaacctat tatggggcac 11220
 gaaggatgga gccaggattt tttttttttt ttttttttct tattagcaat tgatttgcct 11280
 gggettggtt ggacttgccc agttcttagg cccagttctc ttaactgccc atctgaagtc 11340
 tgtcatggag tcagcctagc cttctcactt ccttccagct cgaataggaa gaggagtgct 11400
 acaccagatg gtctgagagc agggataaat ggtgtgcctt tgtctttcag tatttcgtta 11460
 ttttaagtag gaagatgctt ttctgtatta cattgcttgt gaaaccggaa gttgattcgg 11520
 ggcacaggac aatggatttg gtgtttgca aggtgccttt cagaagagag aagagaggaa 11580
 ggggtggttag agtgaggagt ggggtgggac gggatggggg aagagaagga agggccagac 11640
 aggctaggtg gggctgagag gaggcggtgg gaacttcttg agttagcgcg gcagtaact 11700
 tggatgtgcg tgtatctttg tgatataatga cccggagccc tgtagctggc tccgatagta 11760
 ctgctaatgt cagtgtcggg gggggggggt cccatactgt tccacagggg ctgcacattc 11820
 ccctcgagag caggagggct cctctctcca tacatcctcg ccagcattcc ttgttgttc 11880
 tgtgatgaca ggggggtgga tgaaatctct ctgttggttt gagagaccgt gaagaagctc 11940
 aaccccagga cattttgcag tcttgggaagg cagtgcctcc atgtggagcc gtggagccca 12000
 tctctgagtc caggtcactc ttgcagttcg cactcagctc ttcagatgca ggagagcgt 12060
 tgggtgggaaa gcaagattgt ttgcttgttg agatagacac attctccaca caaaggctca 12120
 cgtggggcaa aggctgattg acgtacagcg ttcaggaacg cctgtggtag agctatgatt 12180
 agctgtctcc atctatgaag cagacaaaga gttataaaaa aaatcaatgt tttcaaattg 12240
 tcaaactttt aacccgacag caagcgtct gtccctgggc taatccctag ccctggttc 12300
 ttgagatggg gtcttttgtg cctagactg gcctagaact cacgatctta gtgtccagc 12360
 ctcccagctg ctgggatgag ccgctataac cagtctgcct gccttccctaa attttaagt 12420
 atgggaagtg ggggagaata cagtttaaaag tatgcagatc tgagagcagg aacctggcaa 12480
 agccaagggg ccggagttac aggcggtcaa catgggtgct gggaaactgac ccaggtcctt 12540
 gagaggagca gtgtgtactc ttgaccaaac aggtccgtct ctccagtcct cgtagtatta 12600
 aaaataggtg ctacgggcat ggtggtgcac acctttaatc ccagcactag ggaggcagag 12660
 gcagggtggt ttctgagttt gaggccagcc tgggtctaaa aatgagttcc aggacagcca 12720
 cggctataca gagaaacct gtcttgaaaa caaaaacaaca acaaaatagg tactacaaa 12780
 cgatgtaatt gtgctcaaac atgcaaacgg agggcactgt atgcataaga aagaaagaa 12840
 cggccacact ggttctatct ggggtgacag aaatcagtat ttttattttt cacattcatt 12900
 tttttgttgt tgttgttgac acagtgattt ttctatcaaa aacattattt cttttatagt 12960
 tcccctgagg agctgttttt aaagccgtgc tttgaaaaac cattgaagga gcagaggcag 13020
 ggagactcct gtgtggcagt cgggtgaaag ggcctctgac aggcaggctg gccctgact 13080
 tgggagctct tttccctccc tectgtgctc aaatagcaaa tgtcaggctt caatgtagct 13140
 agaaggttct agaatgatta agttccaag gctgaagagc ttccctgttt gcccttact 13200
 tcccgggaga ggtcgttgtg tgtccggag tctgcaaggt gcctttgggt atgocggtgg 13260
 ttcactctcg gagattccgc ctggaggacc caagttcaag cctgctga gctacagagt 13320
 gactttcagg tcttctgcgc aattcagtg gacccagctc acaataaaaa agtaaaaaga 13380
 aggctgtgga tggaaactcg tggtagagtt ctgggtttac tccctagagg aggggagaag 13440
 gaggagggag gaggaggaag aggaagaaag aagaagagaa gggagagga gaaggaagg 13500
 agggaaaggg ctgacaagaa gagagaagag ggagggaggg gagggaaagg aaggggaaag 13560
 gaagggaggg aaggggctga caagaagaga gaagaggag gaagggaggg aaaggaagg 13620
 ggaagaaga gaagggtaag aagaaactgt tccaatggtc tgggcccacag agtgatggcc 13680
 ttttgtggtg atcagctgta atccttgatt tgacacaacc tagaatctgg gaagcgagtt 13740
 tctgtgaagg agcattcaca ctggctggcc tgtgggctg catgtgggag actgtcataa 13800
 ttaggttcat taatacagga agtcccagcc cactacaaat ggcttogttc cataccaag 13860
 agatgctaac tgtagacggt tggagaaagc aagcaagctg tggatacccc acgctctttc 13920
 acctcggctc ctggggggtg ggtgcactgt gtctcttgggt attttaaagt cctgccttga 13980
 cgctccctgct tgacagact gtaactggaa ttgtgagctt tagtccctta gttttctacg 14040
 ttggtttttc tcaggatatt ttatcgagtt aacagaaaca agaccaggac acttgcctc 14100
 ctctgatcaa cactgaagag ttacaaaaca ggctgaggaa acaaaacttc ttctccctct 14160
 cccctctctg tccctccctt tccctctctc tccctccctt gccccctctc tccctgtctc 14220
 tgtctctctc tctgtctctg tctctgtctc tgtctctgct tctccctcc cctccctcc 14280

ES 2 522 615 T3

```
ctctgtctct gtctctgtct ctgtctctgt ctctgtctct gtctctgtcc cttctctctc 14340
tatctcctaa atggctggag gccatgctag ctcaatggtg aactttgaac acgtatcttag 14400
gaaatctttg ttcttaacag ttctgaagtg ctgaagtggg ggtttagtct ctgggctga 14460
caagctcact tctctcact ctgtcttaat gaccaaactc gccatttccc taaaacagca 14520
caggctccag ctccagggtg ctccggagcg gag 14553
```

<210> 7

<211> 34

5 <212> ADN

<213> Secuencia artificial

<220>

<223> Sintética

10

<400> 7

ataacttcgt ataatgtatg ctatacgaag ttgt 34

<210> 8

<211> 34

15

<212> ADN

<213> Secuencia artificial

<220>

<223> Sintética

20

<400> 8

ataacttcgt ataatgtata ctatacgaag ttag 34

REIVINDICACIONES

- 5 1. Una célula aislada que comprende una secuencia potenciadora de la expresión seleccionada entre las SEC ID N° 1, 3 ó 5, y al menos un sitio de reconocimiento de recombinasa localizado dentro de una secuencia de nucleótidos de las SEC ID N° 1, 3 ó 5.
2. La célula aislada de la reivindicación 1, donde el al menos un sitio de reconocimiento de recombinasa se selecciona entre el grupo que consiste en un sitio loxp, un sitio lox 511, un sitio lox 2272, y un sitio frt.
- 10 3. La célula aislada de la reivindicación 1, que comprende adicionalmente un primer gen de interés (GOI) que codifica una proteína, donde el primer GOI se integra en el al menos un sitio de reconocimiento de recombinasa; opcionalmente donde el GOI está unido de forma funcional a un promotor.
- 15 4. La célula aislada de la reivindicación 3, que comprende adicionalmente un segundo GOI que codifica una proteína, donde está presente un sitio de reconocimiento de recombinasa entre el primer GOI y el segundo GOI; que opcionalmente comprende adicionalmente un sitio de reconocimiento de recombinasa 5' al primer GOI y un sitio de reconocimiento de recombinasa 3' con respecto al segundo GOI.
- 20 5. La célula aislada de la reivindicación 1, donde la célula es una célula CHO, y donde la secuencia potenciadora de la expresión es la SEC ID N° 5.
- 25 6. La célula aislada de la reivindicación 5, donde el sitio de reconocimiento de recombinasa está localizado en una posición entre los nucleótidos 5.000-7.400 de la SEC ID N° 5; los nucleótidos 5.000-6.500 de la SEC ID N° 5; los nucleótidos 6.400-7.400 de la SEC ID N° 5; o los nucleótidos 6.400-6.500 de la SEC ID N° 5.
- 30 7. Una célula CHO que comprende una secuencia potenciadora de la expresión que es los nucleótidos 6.400-6.500 de la SEC ID N° 5, que comprende adicionalmente al menos un sitio de reconocimiento de recombinasa flanqueado 5' o 3' por un gen humano de interés (GOI), opcionalmente donde el GOI humano está flanqueado 5' por el primer sitio de reconocimiento de recombinasa, y dicha célula CHO comprende adicionalmente un segundo sitio de reconocimiento de recombinasa que flanquea 3' el GOI humano.
- 35 8. La célula CHO de la reivindicación 7, que comprende adicionalmente un segundo GOI humano que flanquea 3' el segundo sitio de reconocimiento de recombinasa; que opcionalmente comprende adicionalmente un tercer sitio de reconocimiento de recombinasa que flanquea 3' el segundo GOI humano.
9. La célula CHO de la reivindicación 8, que comprende adicionalmente al menos un gen marcador entre el segundo sitio de reconocimiento de recombinasa y el segundo gen de interés.
- 40 10. La célula CHO de la reivindicación 9, que comprende adicionalmente un promotor unido de forma funcional al primer GOI humano y un promotor unido de forma funcional al segundo GOI; opcionalmente donde los promotores son promotores eucariotas, y los promotores eucariotas están unidos de forma funcional a un operador procariota.
- 45 11. La célula CHO de la reivindicación 10, donde el segundo y el tercero sitios de reconocimiento de recombinasa están en una orientación opuesta al primer sitio de reconocimiento de recombinasa.
- 50 12. Una célula aislada que comprende una secuencia potenciadora de la expresión seleccionada entre las SEC ID N° 1, 3 ó 5 y al menos un GOI añadido de forma exógena dentro de la secuencia potenciadora de la expresión.
13. La célula aislada de acuerdo con la reivindicación 12, donde la célula es una célula CHO, la secuencia potenciadora de la expresión comprende la SEC ID N° 5, el al menos un GOI añadido de forma exógena es un gen humano, y el gen humano está unido de forma funcional a un promotor añadido de forma exógena.
- 55 14. Un método para preparar una proteína de interés, que comprende:
 - (a) proporcionar una célula hospedadora que comprende una secuencia potenciadora de la expresión seleccionada entre las SEC ID N° 1, 3 ó 5, o los nucleótidos 6.400-6.500 de la SEC ID N° 5,
 - (b) introducir en la célula hospedadora, dentro de la secuencia potenciadora de la expresión, un gen de interés (GOI) unido de forma funcional a un promotor;
 - (c) mantener la célula hospedadora de (a) en condiciones que permita que el GOI exprese una proteína de interés; y
 - (d) recuperar la proteína de interés.
- 60 65 15. La célula aislada, célula CHO o método de una cualquiera de las reivindicaciones precedentes, donde el primero y/o segundo GOI codifica una proteína seleccionada entre una cadena ligera de anticuerpo o fragmento específico de antígeno de la misma, una cadena pesada de anticuerpo o fragmento específico de antígeno de la misma, un ligando, y un receptor o fragmento específico de ligando del mismo.

16. Uso de la célula aislada de la reivindicación 1 o la célula CHO de la reivindicación 7 para preparar un anticuerpo o fragmento del mismo.

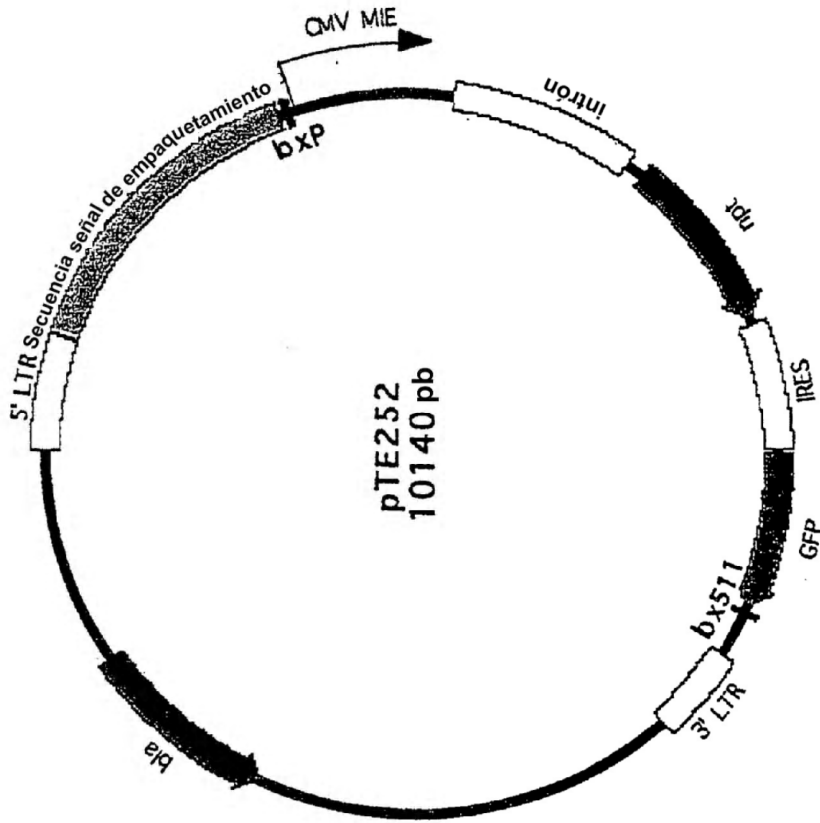


Fig. 1

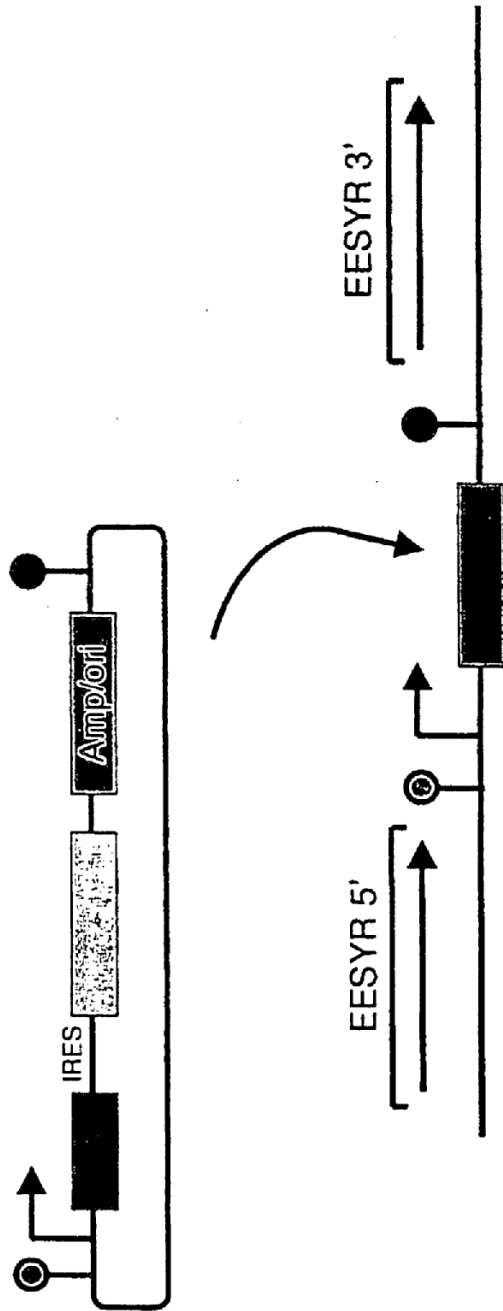


Fig. 2

Fig. 3A

EESYR	5032	TT-----CCTGAGTTCGCCAA-CGCCTAGTGACATCCTGTG-----	5067
Recorte de Ch3 de rata	5008	TTTATC-----TTTCCTGGGGACAGCCAA-AGCCTGGTGGCATCCTCTGTGAAGTGTTC	5061
Recorte de Ch2 de M. Mus	6021	TTTATC-----TTTCCTGGGGACAGCCAA-AGCCTGGTGGCATACTGTACATCTG---	6071
Recorte de Ch20 de H. Sap	5972	AGAAACACAACCTTCTCCAGGGAGAGTCAAGAGAGCAGTGTCTATTTCCATAAAAATGTCC	6031
		* * * * *	
EESYR	5068	-----CACACCCT-----AAGGTGGCCTTGTGTGGC-----ACTGGCTGGGGTG	5107
Recorte de Ch3 de rata	5062	ATCAAGACAGGCTCTGTCTTAAAGGTGGCCTTCGCATGGCGCTGGCACTGGCCGGGGCA	5121
Recorte de Ch2 de M. Mus	6072	-----TACACTGTTCATTCAAACAGGCTCTGTCTTAAA-----GATGGTCTGAGCG	6118
Recorte de Ch20 de H. Sap	6032	A--AGGCCAATCCAGGTCCTGGTGTGAGCCCTTGCCTGTGGC-----ACTGGCACCAGCA	6085
		* * * * *	
EESYR	5108	GTCGGGAAAGG-CATTTTCAGCTTGTG--GCAGAAGTGCACAGTAGCATGTGGGT--CCG	5164
Recorte de Ch3 de rata	5122	GTCAGGAGAGGGTATTTTTAGCTTCTTTGCAGAAATGGCCTCAGGAGCGTGTGAGT--CTG	5180
Recorte de Ch2 de M. Mus	6119	GTCAGAAAGGGTATTGTTACTTGTGTTGCAAACTGCCTCAGGAGAGTGTGAGT--GCG	6177
Recorte de Ch20 de H. Sap	6086	GCCAGAAAGGT-TACTGTGACTTGTTCAGTGTCTTCTGGGAGTGTCTGAAAGCTG	6144
		* * * * *	
EESYR	5165	TGAAAGTTCTGCCCGTTAACAGAAGTCTCTACTACTTGTGACCTCACCA-----GTGA	5219
Recorte de Ch3 de rata	5181	TGAAAGTTGCTGCCAGTTAAGGAGAAGTCTTACCACCTCGTACCTCACCA-----TTGA	5235
Recorte de Ch2 de M. Mus	6178	TGAAAGTTGCTGCCCGTTAAGGAGAAGTCTCTACTACTTGTGATCTCACCA-----TCGA	6232
Recorte de Ch20 de H. Sap	6145	GCAGTGTCTTCTGCCTCGTCAGAAGAGGCCATTCATTGTTGGGGTGTACCAGCAACTTAA	6204
		* * * * *	
EESYR	5220	AAATTTCTTTAATG-----TCTCCTGGTGTTC-TGGGTTTTG-----CATTTTTGT	5265
Recorte de Ch3 de rata	5236	AAATTTCTTTAATG-----TCTCCTCGTGTTC-TGGGTTTTG-----CAGTTTTGT	5281
Recorte de Ch2 de M. Mus	6233	AAATTTCTTTAATG-----TCTCCTGGTGTTC-TGGGTTTTG-----CAGTTTTGT	6278
Recorte de Ch20 de H. Sap	6205	AAAGTTTTTTTTTTCTTCCCTCTCCAGTGTAAATAGATTTTGAAAGAGGCATTTTTGT	6264
		* * * * *	
EESYR	5266	TTCTAAGGATACATTCCTGGGTGATGTCAAGTCCCAAAGACACAGTG--GGCTGTG	5324
Recorte de Ch3 de rata	5282	TCCTAAGGATACATTCCTGGGTGATGTCAAGTCCCAAAGACACAGTG--GGCTGTG	5340
Recorte de Ch2 de M. Mus	6279	TTCTAAGGATACATTCCTGGGTGATGTCAACAAGTCCCAAAGACACAGTG--GAGCTGTG	6337
Recorte de Ch20 de H. Sap	6265	TTCTAATTAACAATTCCTGGGTGATGTCTAATAAGCCCTCAAACACCCCGTGAAGGCGGCA	6324
		* * * * *	
EESYR	5325	TTGGATTGGGAAAGAT-----GATTTATCTGG-GCTGTCAAAGGAAAAGAAG	5371
Recorte de Ch3 de rata	5341	TTAGATCGGGACAGACAATATTGCTGAGGATTTATCTGAAGGTGTCAAAGGAGAAGAAG	5400
Recorte de Ch2 de M. Mus	6338	TTAGATGGGGAAGACAGT--CTGCTGAGGATTTATCTGGAAGTGTCAAAGGAAAAGAAG	6396
Recorte de Ch20 de H. Sap	6325	CTGAATGGGGAAGACAATA-TGCTAACAAGTTATCTGAGGGTGTCAAAGGAAAAGAAG	6383
		* * * * *	
EESYR	5372	GGAACACAGG-CACTTGGGA-AAATGTCTCCCGCCACCCGAATTTGGCTT-GGCAACC	5428
Recorte de Ch3 de rata	5401	GGAACACAGGCACTCAGGG-AAATGGCCTCTAGTCT----GAGTCTGGCTC-AGCAACA	5454
Recorte de Ch2 de M. Mus	6397	GTAATGGGGCACTTGGGA-RAAGTGGCCTCTAGTCT----GACTTCTGGCTT-AGCAAAG	6450
Recorte de Ch20 de H. Sap	6384	AAAAACAGTACAGTTGGGGAAATGTTTTCCAGCCT----GCTTCTGGTTTTAGCGAT	6439
		* * * * *	
EESYR	5429	GT-GGTGGAGGAGCAAGAAACCGTGGACGTTTGA-GGAGGCAT-GGGGCTCCTAGGAGGA	5485
Recorte de Ch3 de rata	5455	GA-GGTGGGGAGATAAGGCACACACAGTGGTTAGAAAGGATCATCAGGTTCTGGGAGGA	5513
Recorte de Ch2 de M. Mus	6451	GT-TGTGGGGAGATAAGGCATACACAGTAGTTAGCAGGAGGCAACAGGTCCTGGGAGGA	6509
Recorte de Ch20 de H. Sap	6440	GCATGGGAAGAGATAAGACACACATGGCTTTTATAAGGAGCCATCGGGATCTCAGGGGA	6499
		* * * * *	
EESYR	5486	CAGGAAGCAGAAGGAGAGAGCTGGGCTGACAGCCTGCAGGCATGACACAGTTTC--AGAA	5543
Recorte de Ch3 de rata	5514	CAGGAGGCAGGAG----AGGCAGGGCTGACAGTGTGCAATCATTTGTGTAGTCTC--CCAA	5567
Recorte de Ch2 de M. Mus	6510	CGCGAGGCAGAAGGAG-AGGCTGGGCTGACAGCATGCAATCATTTGATAGTCTC--CAA	6566
Recorte de Ch20 de H. Sap	6500	CATGAGGCAGGAGAAAAGAAATGGGGTGAAAGCATCCAATCATCATATTCACGGAGAA	6559
		* * * * *	
EESYR	5544	GGAGATTACAGCATGACTGAG-----TTTTTAGGGATCCAACAGGGACCTG--GGTAG	5594
Recorte de Ch3 de rata	5568	GGAGATTACAACATGGCTGAA-----TTTTCAGGGGTCCAACGGAGACTGT--AGTGG	5618
Recorte de Ch2 de M. Mus	6567	GGAGATTGCAACATGGCTGAG-----TTTTCAGAGGTCCACAGAGCCCGT--GGTAG	6617
Recorte de Ch20 de H. Sap	6560	AGAGATTACAATATAGCAGGAAAGCTTCCAGGGCTCCTACAGGGACCTTTGGGACAA	6619
		* * * * *	
EESYR	5595	AGATTCTGTGGGCTCTGAGGCAACTTGACCTCAGCCAGATGGTATTGTAATAACCTGCTC	5654
Recorte de Ch3 de rata	5619	AGATTCTGTGGGTTCTGAGACAACCTTGACTTCAAGCCAGATGGCATTGTAATAAC-----	5672
Recorte de Ch2 de M. Mus	6618	AGATTCTGTGGGTTCTGAGACAACCTTGACTTCAAGCCAGATGGTATTGTAATAACCTGCG-	6676
Recorte de Ch20 de H. Sap	6620	AGATTCAAGTGGGCTTGGGACAGCTTGACTTCAACTAGATGGTATTGTAATAACCTGCT-	6678
		* * * * *	

Fig. 3B

```

EESYR          5655 TTAGAGGGAAAACAGACATAGCAAACAGAGCCACGTTTAGTGATGAAACTCTCACTTTGC 5714
Recorte de Ch3 de rata 5673 -----AGCTCCAGCAAGCACAGCCACATTTAGGGATGAAACTCTCACTTTGA 5719
Recorte de Ch2 de M. Mus 6677 --AGAGAGAAAACAGCTACAGCAAACAGGGCCACATTTAGTGACGAAACTCTCACTTTGA 6734
Recorte de Ch20 de H. Sap 6679 --GGAGAGAAAACAGATATAGCACACACTGTCACATTTAGCGTTGAAACTCTCGGTTTGA 6736
                **      ****  **  *  ***  *****  *  *****  ****

EESYR          5715 CTG---AGTCATGTGCGGCCATGCCAGGGGTCAGGCTGACACTCAACTCAAAAACAAGT 5771
Recorte de Ch3 de rata 5720 CTGT-GAGTCACGTGTAGCTGTGTCCCGAGGTCAGGCTGGCCCTCAGCTCAAAAACAAGT 5778
Recorte de Ch2 de M. Mus 6735 CTGTTGAGTCATTTGCAG-TGGGCCCTGAGGTCAGGCTGGCCCTCAGCTCAAAAACAAGC 6793
Recorte de Ch20 de H. Sap 6737 CTAT-GAGTAATGTTTCAGCCATGCCCAAGGGTCAGGCCTACACTCA-CTCAGAAAACAAGT 6794
                **      ***  *  *  *  *  **  *****  *  ****  ****  *****

EESYR          5772 GAGAAATTGAAG-----ACAATCCGTGGTGGCAGCTACTGGAA-GGGCCACCAC 5819
Recorte de Ch3 de rata 5779 GAGGGATTGAAGCAATTACTCAGATAATTCACAGCCACAGCTACGGGGA-GGGCCGC--- 5834
Recorte de Ch2 de M. Mus 6794 GAGGAACCTGAAGCAATTACTCAGATAATCCACAGCCACAGCCACTGGAAAGGGCCAC--- 6850
Recorte de Ch20 de H. Sap 6795 GGGGAATTGAAGCAATTATTAGATAATCCATAGACATAGCTACTGGCCAGTGCTGC--- 6851
                *  *  *  *****  *  ***  *  *  ***  **  *  *  **  *

EESYR          5820 ATCCCCAGAAAGAGT-----GGAGCTGCTAAAAAGCCAT-----TTGTGATAGGCA 5865
Recorte de Ch3 de rata 5835 ATCCCCAGAAACATC-----GGGGTACTATAAAGCTA-----GTGGTGGTCA 5877
Recorte de Ch2 de M. Mus 6851 ATCCCCAGAGACAGCACAGCAGGGGTGGGGGTGGGGCTATGAGAAAAGTTAGTGATTGTAG 6910
Recorte de Ch20 de H. Sap 6852 -ACCCCTGATTTAGCCCA---GAAACAGTAATGCTATTATAAGCT-GTGTGGTATTGTG 6906
                ****  **  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *

EESYR          5866 CAGTTATCTTGAATGCATGGAGCAGAG---ATTACGAAAAATCGAGAATGTTAATGAGG 5922
Recorte de Ch3 de rata 5878 CAGTTATCTTGAATGTATGGAGCAGAGGAGATTACAGAAAAACCTAGAATGTTAATGAGG 5937
Recorte de Ch2 de M. Mus 6911 CAGTTATCTAGAATGTGCGGAGCAGAGGAGTTACACAAAAACCTAGAATG----- 6961
Recorte de Ch20 de H. Sap 6907 GAGCAATCTTGAATTTATTAAGGAARAGGAGATTATAGAAAAATCCAGAATGCCAATG-GG 6965
                **  *****  *****  **  *  **  ***  *****  *  *****
    
```

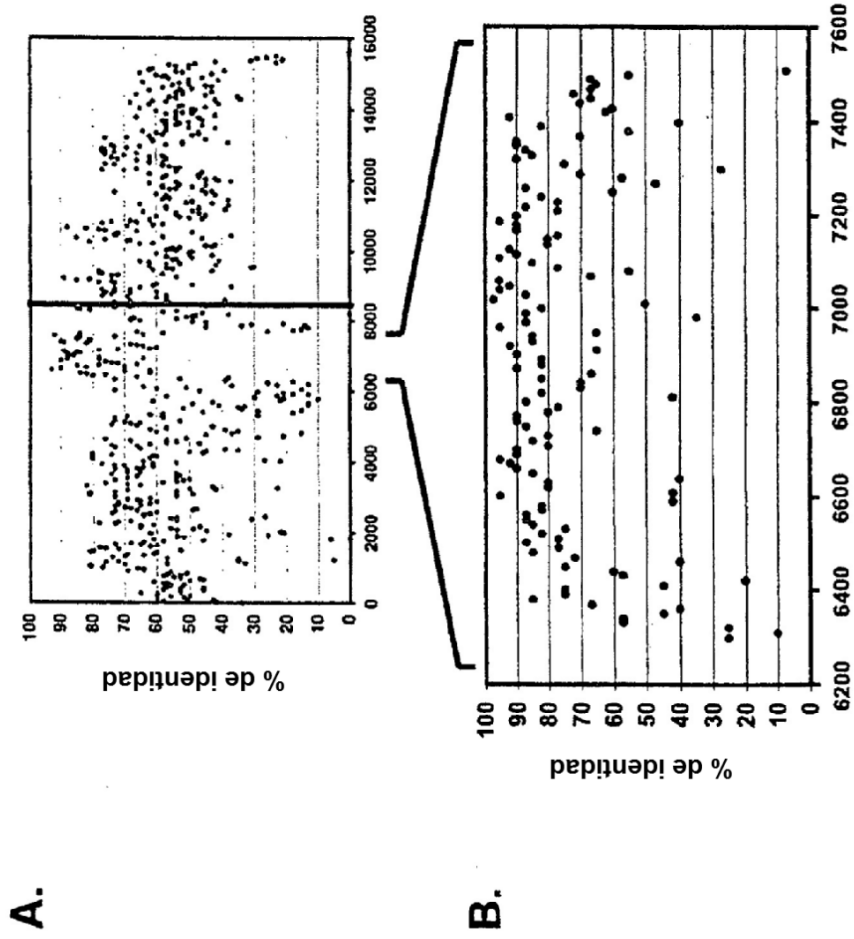


Fig. 4

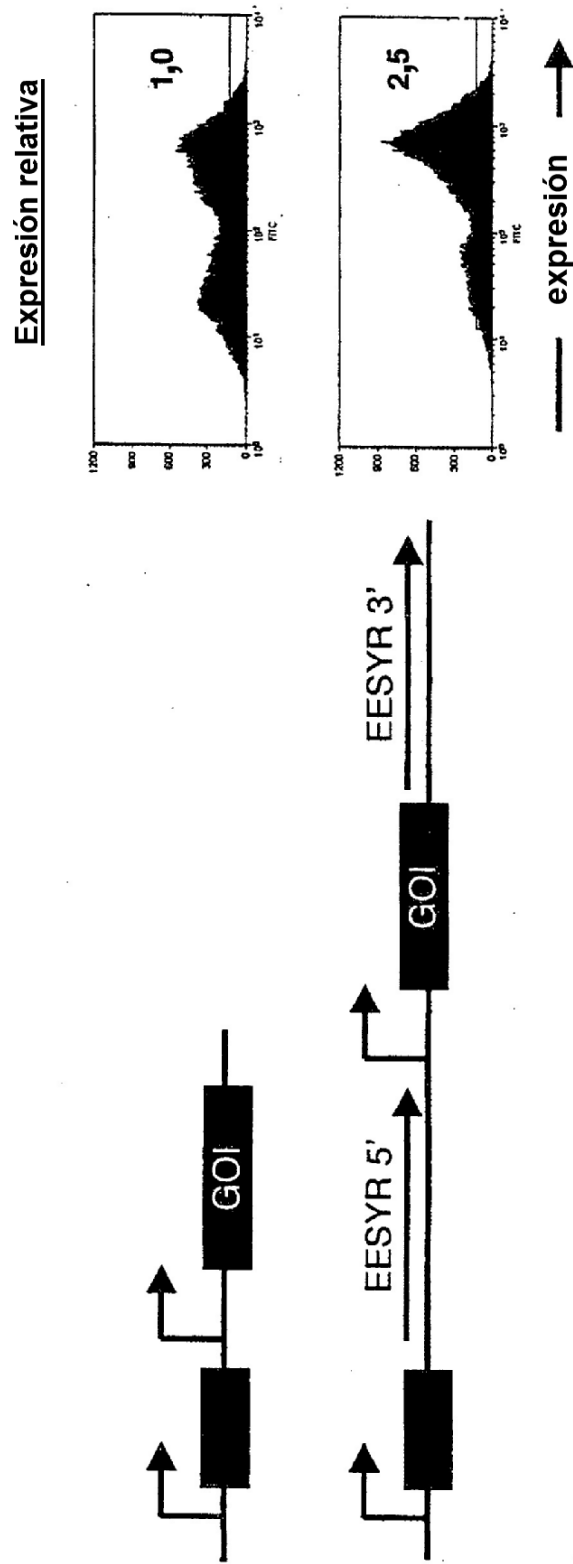


Fig. 5

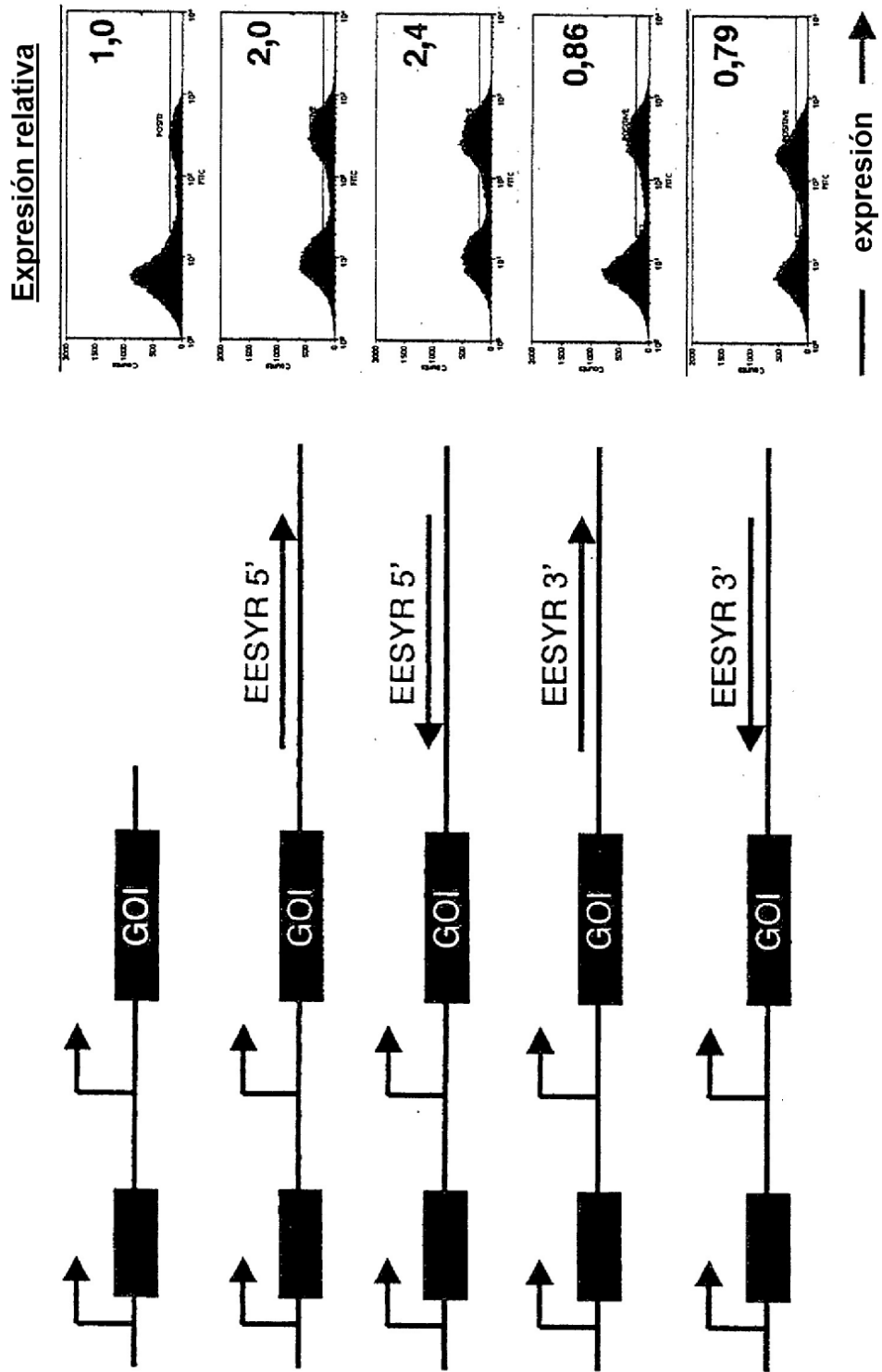
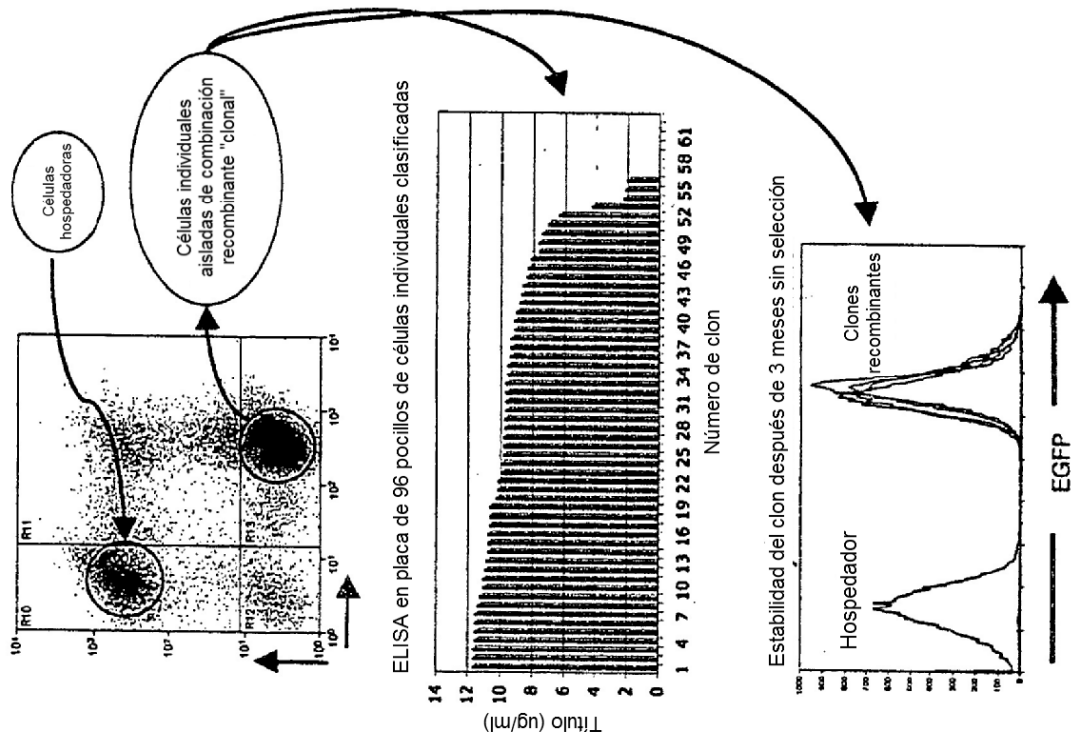


Fig. 6



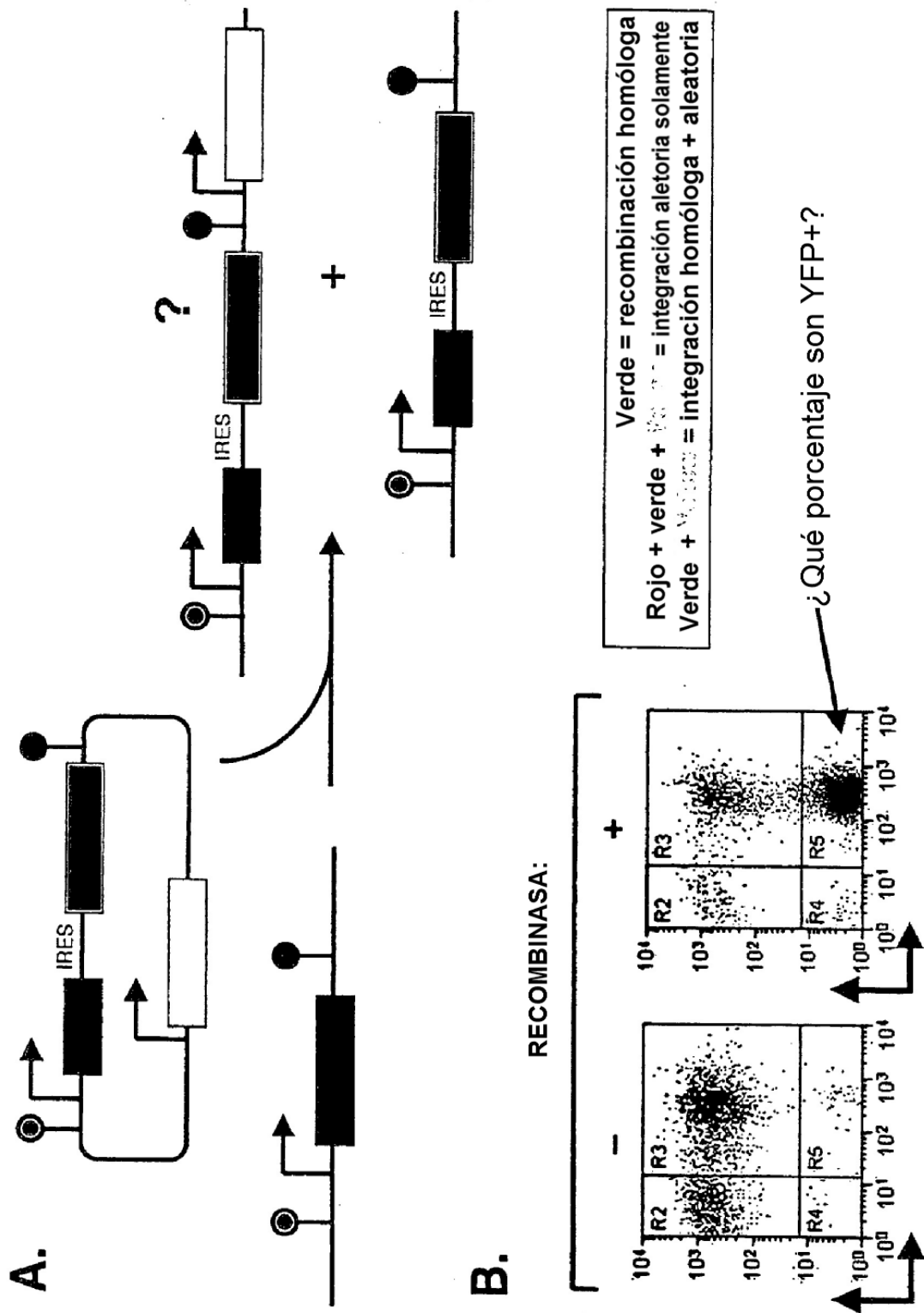


Fig. 8

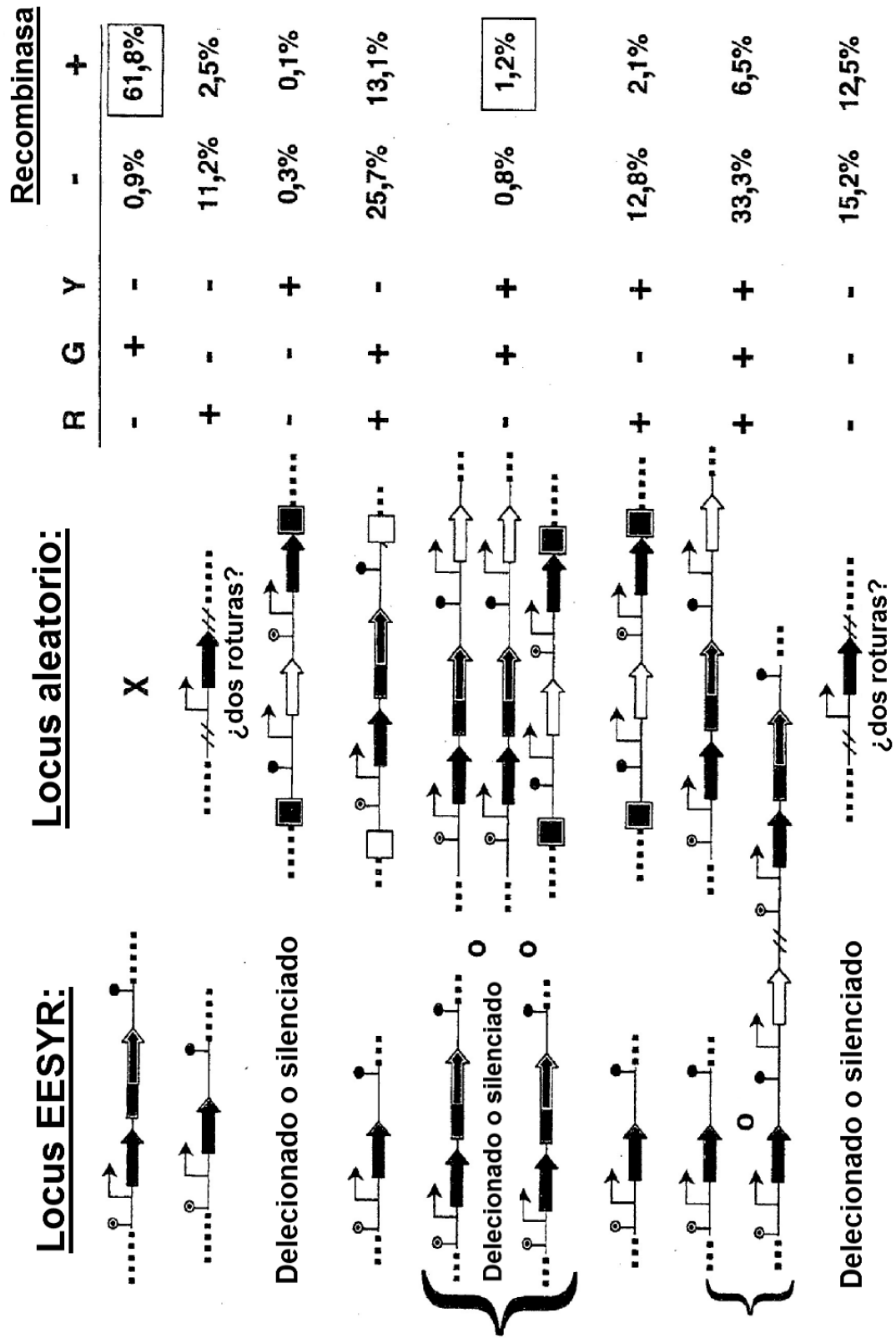


Fig. 9

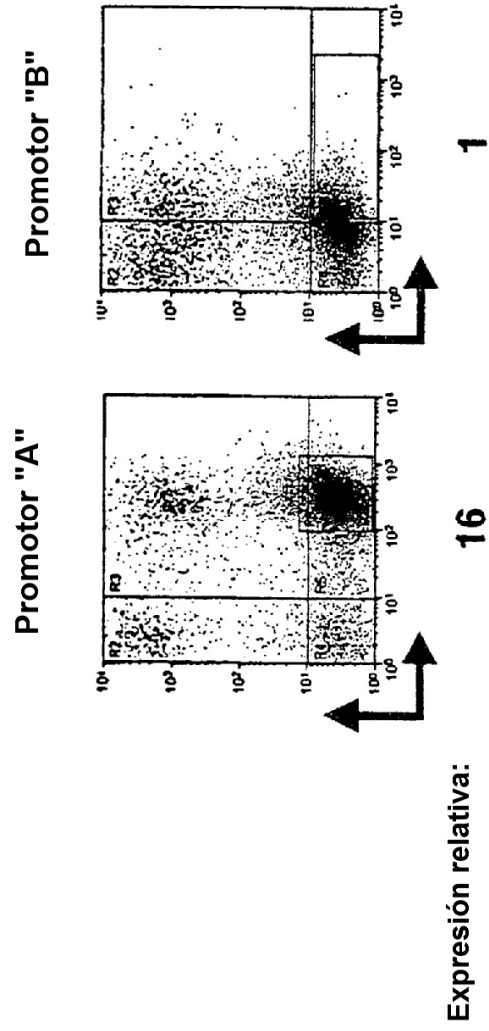


Fig. 10

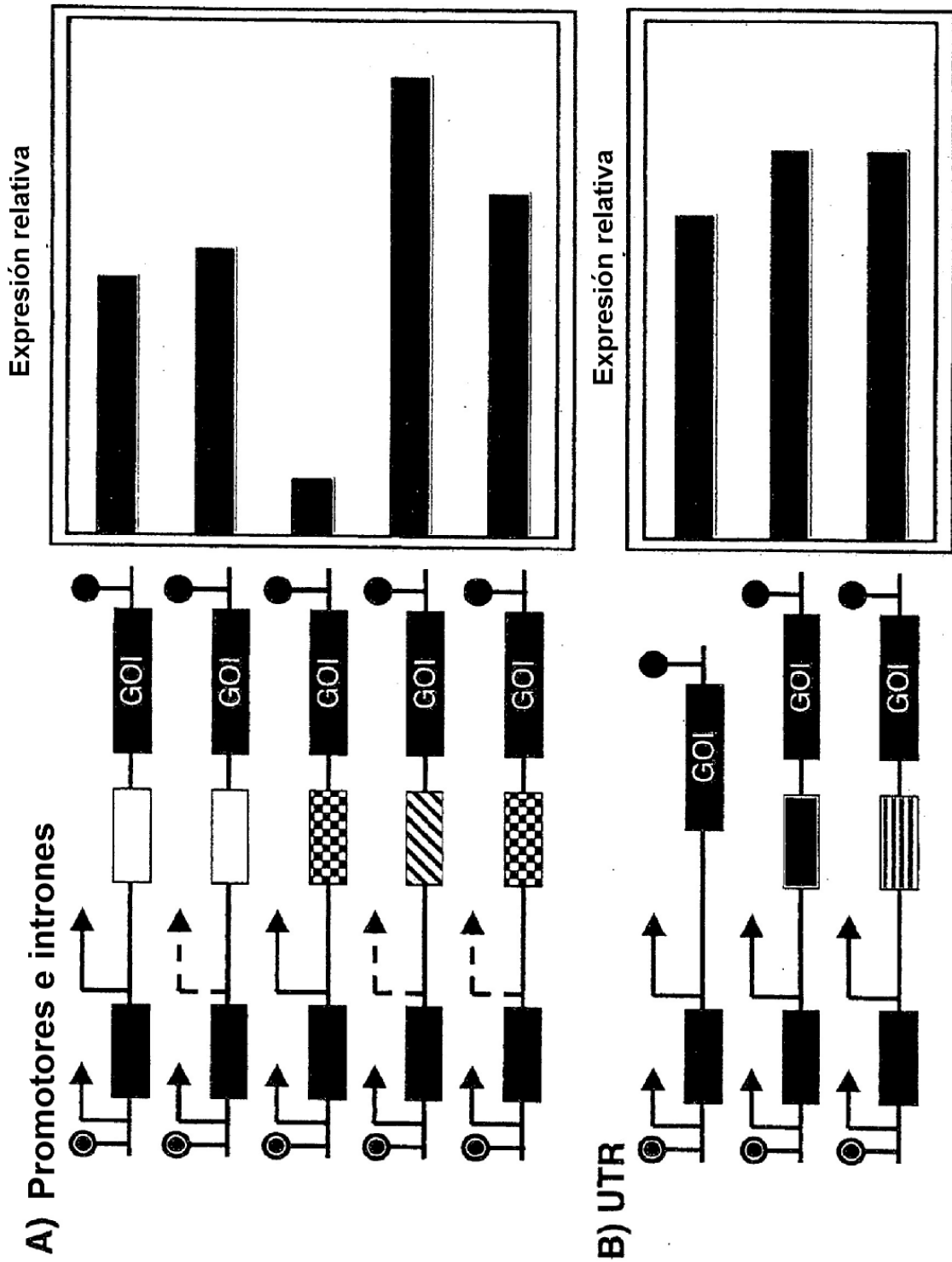
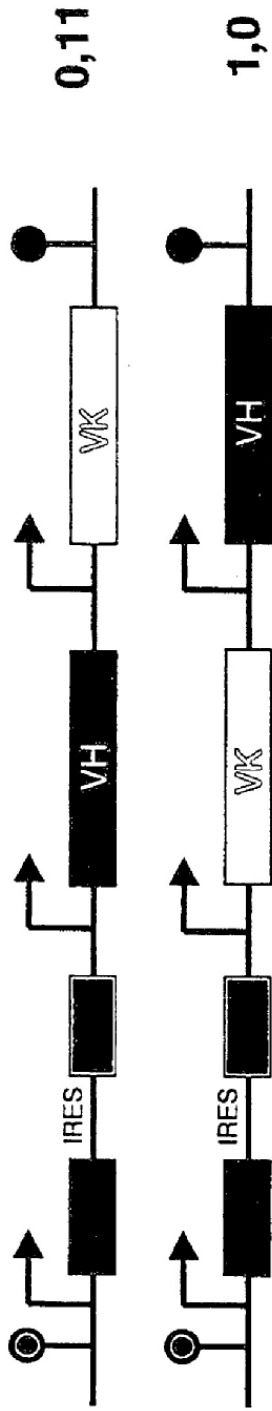


Fig. 11

A) Anticuerpo: orientación de cadena pesada y ligera Expresión relativa



B) Optimización de codones

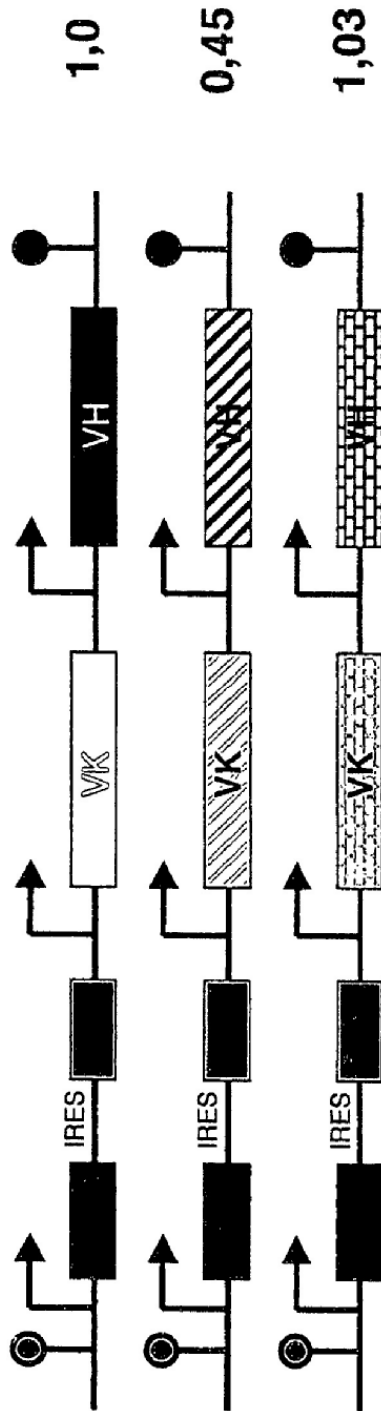


Fig. 12