

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 384 391**

51 Int. Cl.:
C12N 15/67 (2006.01)
C12N 5/10 (2006.01)
C12N 15/79 (2006.01)
C12N 15/09 (2006.01)
C12N 15/11 (2006.01)
C12N 15/85 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Número de solicitud europea: **05810935 .6**
96 Fecha de presentación: **07.11.2005**
97 Número de publicación de la solicitud: **1809750**
97 Fecha de publicación de la solicitud: **25.07.2007**

54 Título: **Selección de células hospedantes que expresan proteína a altos niveles**

30 Prioridad:
08.11.2004 EP 04105593
08.11.2004 US 626301 P
05.07.2005 US 696610 P

45 Fecha de publicación de la mención BOPI:
04.07.2012

45 Fecha de la publicación del folleto de la patente:
04.07.2012

73 Titular/es:
CHROMAGENICS B.V.
ARCHIMEDESWEG 4
2333 CN LEIDEN, NL

72 Inventor/es:
OTTE, Arie Pieter;
VAN BLOKLAND, Henricus Johannes Maria;
KWAKS, Theodorus Hendrikus Jacobus y
SEWALT, Richard George Antonius Bernardus

74 Agente/Representante:
Lehmann Novo, Isabel

ES 2 384 391 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

Selección de células hospedantes que expresan proteína a altos niveles

La invención se refiere al campo de la biología molecular y la biotecnología. Más específicamente, la presente invención se refiere a medios y métodos para mejorar la selección de células hospedantes que expresan proteínas a altos niveles.

Las proteínas pueden producirse en diversas células hospedantes para un amplio intervalo de aplicaciones en biología y biotecnología, por ejemplo como productos biofarmacéuticos. Se prefieren para este fin células hospedantes eucariotas, y particularmente de mamífero, para la expresión de muchas proteínas, por ejemplo cuando dichas proteínas tienen ciertas modificaciones postraduccionales tales como glucosilación. Los métodos para dicha producción están bien establecidos, y suponen generalmente la expresión en una célula hospedante de un ácido nucleico (también denominado como "transgén") que codifica la proteína de interés. En general, el transgén, junto con un gen marcador seleccionable, se introduce en una célula precursora, se seleccionan células para la expresión del gen marcador seleccionable, y se identifican uno o más clones que expresan la proteína de interés a altos niveles, y se usan para la expresión de la proteína de interés.

Un problema asociado a la expresión de transgenes es que es impredecible, partiendo de la alta probabilidad de que el transgén se haga inactivo debido al silenciamiento génico (McBurney et al., 2002), y por lo tanto muchos clones de células hospedantes tiene que ensayarse en busca de la expresión elevada del transgén.

Se conocen métodos para seleccionar células hospedantes recombinantes que expresan niveles relativamente elevados de proteínas deseadas.

Un método describe el uso de proteínas marcadoras seleccionables con mutaciones en su secuencia codificante que disminuyen pero no destruyen la función del marcador (por ejemplo, documento WO 01/32901). La razón fundamental es que se requieren mayores niveles de la expresión del marcador mutante cuando se emplean condiciones de selección y por lo tanto se logra la selección para la expresión elevada del marcador, seleccionando concomitantemente con eso células hospedantes que también expresan el gen de interés en niveles elevados.

Otro método hace uso de un gen marcador de selección bajo el control de una secuencia promotora que se ha mutado, de manera que el promotor tiene un nivel de actividad sustancialmente inferior a aquel de su tipo salvaje correspondiente (patente U.S. 5.627.033).

Otro método describe el uso de una secuencia marcadora seleccionable dominante alterada, tal como neomicina fosfotransferasa con una secuencia Kozak de consenso alterada, para reducir el número de colonias a identificar e incrementar los niveles de expresión de un gen de interés que está coligado al marcador seleccionable dominante (patentes US 5.648.267 y 5.733.779). En realizaciones preferidas allí, el gen de interés se coloca en un intrón (artificial) en el marcador seleccionable dominante. El gen de interés y el marcador seleccionable dominante están en casetes transcripcionales diferentes, y cada uno contiene su propio promotor eucariota en este método (patentes US 5.648.267 y 5.733.779).

Otro método describe el uso de una unidad de transcripción monocistrónica que comprende un sitio de comienzo traduccional (Kozak) no óptimo unido a un marcador seleccionable para alterar la traducción del marcador seleccionable mediante el comienzo traduccional no óptimo para identificar secuencias génicas celulares que podrían superar tal alteración permitiendo la sobreproducción del marcador seleccionable (patente US 6.107.477).

Otro método usa el principio de un gen marcador seleccionable que contienen un intrón que no se produce naturalmente en el gen seleccionable, en el que el intrón es capaz de ser empalmado en una célula hospedante para proporcionar ARNm que codifica una proteína seleccionable, y en el que el intrón en el gen seleccionable reduce el nivel de proteína seleccionable producida a partir del gen seleccionable en la célula hospedante (Patente Europea 0724639 B1).

En todavía otro método, se usan constructos de ADN que comprenden un gen seleccionable situado en un intrón definido por un sitio dador de empalme 5' que comprende una secuencia dadora de empalme eficiente de manera que la eficiencia de empalmar un ARNm que tiene dicho sitio dador de empalme está entre alrededor de 80-99%, y un sitio aceptor de empalme 3', y un producto génico que codifica un producto de interés en dirección 3' del sitio aceptor de empalme 3', estando el gen seleccionable y el producto génico controlados por la misma región reguladora transcripcional (patente US 5.561.053).

En ciertos métodos, se hace uso de constructos de vectores de expresión policistrónicos. Un documento previo del uso de este principio describe un vector de expresión policistrónico que contiene secuencias que codifican tanto la proteína deseada como una proteína seleccionable, secuencias codificantes las cuales están gobernadas por el mismo promotor y separadas por codones de señales de inicio y de parada traduccionales (patente US 4.965.196). En realizaciones preferidas en el documento US 4.965.196, el marcador seleccionable es el gen DHFR amplificable.

En una realización particularmente preferida del sistema descrito en el documento US 4.965.196, la secuencia que codifica el marcador seleccionable está en dirección 3' de aquella que codifica el polipéptido deseado, de manera

que los procedimientos diseñados para seleccionar las células transformadas por el marcador seleccionable también seleccionarán la producción particularmente mejorada de la proteína deseada.

5 En mejoras adicionales basadas en el concepto de vectores de expresión multicistrónicos, se han descrito vectores bicistrónicos para la creación rápida y eficiente de estirpes celulares de mamíferos estables que expresan proteína recombinante. Estos vectores contienen un sitio de entrada del ribosoma interno (IRES) entre la secuencia codificante en dirección 5' para la proteína de interés y la secuencia codificante en dirección 3' del marcador de selección (Rees et al, 1996). Tales vectores están comercialmente disponibles, por ejemplo el vector pIRES1 de Clontech (CLONTECHniques, octubre 1996). Usando tales vectores para la introducción en células hospedantes, la selección de suficiente expresión de la proteína marcadora en dirección 3' selecciona entonces automáticamente niveles de transcripción elevados del ARNm multicistrónico, y por tanto se prevé una probabilidad fuertemente incrementada de expresión elevada de la proteína de interés usando tales vectores.

15 Preferiblemente en tales métodos, el IRES usado es un IRES que da un nivel relativamente bajo de traducción del gen marcador de selección, para mejorar además las ocasiones de seleccionar células hospedantes con un nivel de expresión elevado de la proteína de interés seleccionando la expresión de la proteína marcadora de selección (véase, por ejemplo, la publicación internacional WO 03/106684).

La presente invención tiene como meta proporcionar medios y métodos mejorados para la selección de células hospedantes que expresan niveles elevados de proteínas de interés.

Sumario de la invención

20 La presente invención usa un concepto nuevo y único para seleccionar células hospedantes que expresan niveles elevados de polipéptidos de interés, refiriéndose el concepto aquí como "reducción interdependiente recíproca". Es único con respecto a los enfoques de la técnica anterior por cuanto se usa un nivel extra de regulación para ajustar finamente la cantidad de productos de traducción procedentes de un transcrito multicistrónico, con lo que se reduce el nivel de traducción de un polipéptido marcador seleccionable, incrementando de ese modo el nivel de traducción del polipéptido de interés codificado en la misma unidad de transcripción multicistrónica, es decir, el nivel de expresión del polipéptido de interés depende directamente y más o menos recíprocamente del nivel de expresión del polipéptido marcador seleccionable (véase la Fig. 13 para una vista esquemática). Por el contrario, los enfoques que usan un IRES entre las secuencias codificantes para el polipéptido de interés y el polipéptido marcador seleccionable (por ejemplo Rees et al, 1996) implican la traducción independiente de ambos polipéptidos, de forma que en esos enfoques no hay un efecto directo de la eficiencia de traducción del polipéptido marcador seleccionable sobre aquella de la secuencia que codifica el polipéptido de interés.

35 En un aspecto, la invención proporciona una molécula de ADN que comprende una unidad de transcripción multicistrónica que codifica i) un polipéptido marcador seleccionable capaz de ser seleccionado en una célula hospedante eucariota, y ii) un polipéptido de interés, teniendo el polipéptido de interés una secuencia de iniciación de la traducción (y un codón de inicio) separada de la del polipéptido marcador seleccionable, caracterizada porque la secuencia codificante para el polipéptido de interés está en dirección 3' de la secuencia codificante para el marcador seleccionable en dicha unidad de transcripción multicistrónica, y la secuencia que codifica el marcador seleccionable no tiene ATG en la hebra codificante después del codón de inicio del polipéptido marcador seleccionable hasta el codón de inicio del polipéptido de interés, y caracterizada porque la secuencia de inicio de la traducción en la hebra codificante para el polipéptido marcador seleccionable se elige del grupo que consiste en:

- 40 a) una secuencia de inicio de la traducción que comprende una mutación en los nucleótidos en las posiciones -3 a -1 y/o +4 de la secuencia de consenso RCCATGG (en la que R es A o G, y en la que la A del codón de inicio ATG es nt +1);
- b) un codón de inicio GTG;
- c) un codón de inicio TTG;
- 45 d) un codón de inicio CTG;
- e) un codón de inicio ATT; y
- f) un codón de inicio ACG.

Preferiblemente, la eficiencia de la iniciación de la traducción del marcador seleccionable está disminuida.

50 Según la invención, la eficiencia disminuida de la iniciación de la traducción está provocada por tener una secuencia de inicio de la traducción no óptima del polipéptido marcador seleccionable.

En realizaciones preferidas de la misma, la secuencia de inicio de la traducción en la hebra codificante para el polipéptido marcador seleccionable comprende una secuencia ATG que define un codón de inicio, estando dicha secuencia ATG en un contexto no óptimo para el inicio de la traducción. Esto da como resultado un uso reducido de esta ATG como codón de inicio, cuando se compara con un codón de inicio ATG en el contexto óptimo.

En una realización más preferida, la secuencia de inicio de la traducción en la hebra codificante del polipéptido marcador seleccionable comprende un codón de inicio diferente del codón de inicio ATG, tal como una de las secuencias GTG, TTG, CTG, ATT o ACG, siendo las dos primeras de éstas las más preferidas. Dichos codones de inicio no ATG están flanqueados preferiblemente por secuencias que proporcionan un reconocimiento relativamente bueno de las secuencias no ATG como codones de inicio, de manera que al menos la traducción de inicio de algunos ribosomas a partir de estos codones de inicio, es decir, la secuencia de inicio de la traducción, comprende preferiblemente la secuencia ACC[codón de inicio no ATG]G o GCC[codón de inicio no ATG]G.

Según la invención, la secuencia que codifica el polipéptido marcador seleccionable no tiene ninguna secuencia ATG en la hebra codificante tras la secuencia que codifica el codón de inicio del polipéptido marcador seleccionable hasta la secuencia que codifica su codón de parada, y ninguna en ninguna de las secuencias en la misma hebra entre dicho codón de parada y el codón de inicio del polipéptido de interés (codón de inicio que se define preferiblemente por una ATG en la misma hebra codificante).

En ciertas realizaciones de la misma, cualquier secuencia ATG presente en el marco y que codifica metionina en la secuencia que codifica el polipéptido marcador seleccionable de tipo salvaje se ha mutado para que codifique valina, treonina, isoleucina o leucina.

En realizaciones preferidas, la proteína marcadora seleccionable proporciona resistencia frente a los efectos letales y/o inhibidores del crecimiento de un agente de selección, tal como un antibiótico.

Preferiblemente, la secuencia codificante del polipéptido de interés comprende una secuencia de inicio de la traducción óptima.

La invención proporciona adicionalmente casetes de expresión que comprenden una molécula de ADN según la invención, casetes de expresión los cuales comprenden además un promotor en dirección 5' de la unidad de expresión multicistónica y siendo funcionales en una célula hospedante eucariota para el inicio de la transcripción de la unidad de expresión multicistónica, y comprendiendo adicionalmente dichos casetes de expresión una secuencia de terminación de la transcripción en dirección 3' de la unidad de expresión multicistónica.

En realizaciones preferidas de los mismos, dichos casetes de expresión comprenden adicionalmente al menos un elemento de control de cromatina elegido del grupo que consiste en una región de unión a matriz o armazón (MAR/SAR), una secuencia aislante, un elemento de apertura de cromatina ubicuo (UCOE) y una secuencia antirrepresora. Las más preferidas en este aspecto son secuencias antirrepresoras, y en realizaciones preferidas dichas secuencias antirrepresoras se eligen del grupo que consiste en: a) una cualquiera de SEQ. ID. NO. 1 a SEQ. ID. NO. 66; b) fragmentos de una cualquiera de SEQ. ID. NO. 1 a SEQ. ID. NO. 66, en las que dichos fragmentos tienen actividad antirrepresora; c) secuencias que son al menos 70% idénticas en la secuencia nucleotídica a a) o a b), en las que dichas secuencias tienen actividad antirrepresora; y d) el complemento de una cualquiera de a) a c). En ciertas realizaciones preferidas, las secuencias antirrepresoras se eligen del grupo que consiste en STAR67 (SEQ. ID. NO. 66), STAR7 (SEQ. ID. NO. 7), STAR9 (SEQ. ID. NO. 9), STAR17 (SEQ. ID. NO. 17), STAR27 (SEQ. ID. NO. 27), STAR29 (SEQ. ID. NO. 29), STAR43 (SEQ. ID. NO. 43), STAR44 (SEQ. ID. NO. 44), STAR45 (SEQ. ID. NO. 45), STAR47 (SEQ. ID. NO. 47), STAR61 (SEQ. ID. NO. 61), y fragmentos o derivados funcionales de estas secuencias STAR. En ciertas realizaciones, el casete de expresión comprende STAR67, o un fragmento o derivado funcional de la misma, colocada en dirección 5' del promotor que dirige la expresión del gen multicistónico. En ciertas realizaciones, el gen multicistónico está flanqueado por ambos lados por al menos una secuencia antirrepresora. En ciertas realizaciones preferidas, se proporcionan casetes de expresión según la invención que comprenden en orden 5' a 3': secuencia antirrepresora A - secuencia antirrepresora B - [promotor - unidad de transcripción multicistónica según la invención (que codifica la proteína marcadora seleccionable funcional y, en dirección 3' de la misma, el polipéptido de interés) - secuencia de terminación de la transcripción] - secuencia antirrepresora C, en los que A, B y C pueden ser iguales o diferentes.

En ciertas realizaciones, el polipéptido de interés es una parte de una proteína multimérica, por ejemplo una cadena pesada o ligera de una inmunoglobulina.

La invención también proporciona moléculas de ADN que comprenden una secuencia que codifica un polipéptido marcador seleccionable funcional, capaz de ser seleccionado en una célula hospedante eucariota, caracterizado porque dicha molécula de ADN:

i) tiene una secuencia de inicio de la traducción escogida del grupo que consiste en:

a) una secuencia de inicio de la traducción que comprende una mutación en los nucleótidos en las posiciones -3 a -1 y/o +4 de la secuencia de consenso RCCATGG (en la que R es A o G, y en la que la A de la secuencia de inicio ATG es nt +1);

b) un codón de inicio GTG;

c) un codón de inicio TTG;

d) un codón de inicio CTG;

e) un codón de inicio ATT; y

f) un codón de inicio ACG, seguido de una secuencia codificante marcadora seleccionable de otro modo funcional, y

5 ii) porque la hebra codificante de la secuencia que define el polipéptido marcador seleccionable en dirección 3' del codón de inicio no óptimo y hasta el codón de parada está desprovista de secuencias ATG.

La invención proporciona también células hospedantes que comprenden moléculas de ADN según la invención.

10 La invención proporciona adicionalmente métodos para generar células hospedantes que expresan un polipéptido de interés, comprendiendo el método las etapas de: introducir en una pluralidad de células hospedantes precursoras un casete de expresión según la invención, cultivar las células en condiciones que seleccionan la expresión del polipéptido marcador seleccionable, y seleccionar al menos una célula hospedante productora del polipéptido de interés.

15 En un aspecto adicional, la invención proporciona métodos para la producción de un polipéptido de interés, comprendiendo los métodos cultivar una célula hospedante, comprendiendo dicha célula hospedante un casete de expresión según la invención, y expresar el polipéptido de interés a partir del casete de expresión. En realizaciones preferidas de la misma, el polipéptido de interés se aísla adicionalmente a partir de las células hospedantes y/o del medio de cultivo de células hospedantes.

Descripción detallada de la invención

20 En un aspecto, la invención proporciona una molécula de ADN que comprende una unidad de transcripción multicistrónica que codifica i) un polipéptido marcador seleccionable, capaz de ser seleccionado en una célula hospedante eucariota, y ii) un polipéptido de interés, teniendo el polipéptido de interés una secuencia de iniciación de la traducción separada de la del polipéptido marcador seleccionable, caracterizado porque la secuencia de codificante para el polipéptido de interés está en dirección 3' de la secuencia codificante para el marcador seleccionable en dicha unidad de transcripción multicistrónica, y la secuencia que codifica el marcador seleccionable no tiene ATG en la hebra codificante tras el codón de inicio del polipéptido marcador seleccionable hasta el codón de inicio del polipéptido de interés, y en la que la secuencia de inicio de la traducción en la hebra codificante para el polipéptido marcador seleccionable se elige del grupo que consiste en:

30 a) una secuencia de inicio de la traducción que comprende una mutación en los nucleótidos en las posiciones -3 a -1 y/o +4 de la secuencia de consenso RCCATGG (en la que R es A o G, y en la que la A de la secuencia de inicio ATG es nt +1);

b) un codón de inicio GTG;

c) un codón de inicio TTG;

d) un codón de inicio CTG;

e) un codón de inicio ATT; y

35 f) un codón de inicio ACG.

40 Dicha molécula de ADN puede usarse según la invención para obtener células hospedantes eucariotas que expresan niveles elevados del polipéptido de interés seleccionando la expresión del polipéptido marcador seleccionable. Subsiguientemente o de forma simultánea, pueden identificarse una o más células hospedantes que expresan el polipéptido de interés, y usarse adicionalmente para la expresión de niveles elevados del polipéptido de interés.

45 La expresión "gen monocistrónico" se define como un gen capaz de proporcionar una molécula de ARN que codifica un polipéptido. Una "unidad de transcripción multicistrónica", también denominada como un gen multicistrónico, se define como un gen capaz de proporcionar una molécula de ARN que codifica al menos dos polipéptidos. La expresión "gen bicistrónico" se define como un gen capaz de proporcionar una molécula de ARN que codifica dos polipéptidos. Por lo tanto, un gen bicistrónico está comprendido dentro de la definición de un gen multicistrónico. Un "polipéptido", como se usa en la presente memoria, comprende al menos cinco aminoácidos ligados por enlaces peptídicos, y puede ser, por ejemplo, una proteína o una parte, tal como una subunidad, de la misma. La mayoría de veces, los términos polipéptido y proteína se usan intercambiabilmente en la presente memoria. Un "gen" o una "unidad transcripcional", como se usa en la presente invención, puede comprender ADN cromosómico, ADNc, ADN artificial, combinaciones de los mismos, y similares. Las unidades de transcripción que comprenden varios cistrones se transcriben como un ARNm individual. En contraste con los enfoques usados en la técnica anterior, en las unidades de transcripción multicistrónicas de la invención, la traducción de la segunda región codificante (es decir, aquella para la proteína de interés) presente en ese ARN no depende del uso de los sitios de reinicio de la

traducción o sitios de entrada del ribosoma internos para el inicio de la traducción de la segunda región codificante, sino más bien depende de la eficiencia de la traducción de la primera región codificante (es decir, aquella para la proteína marcadora seleccionable) de una manera más o menos recíproca: cuanto mayor es el nivel de traducción del marcador seleccionable, menor es el nivel de traducción de la proteína de interés, y viceversa.

5 Una unidad de transcripción multicistrónica según la invención es preferiblemente una unidad de transcripción bicistrónica que codifica desde 5' hasta 3' un polipéptido marcador seleccionable y un polipéptido de interés. Por tanto, el polipéptido de interés está codificado en dirección 3' desde la secuencia codificante para el polipéptido marcador seleccionable.

10 También es posible incluir más secuencias codificantes en la misma unidad de transcripción, en dirección 3' del (primer) polipéptido de interés, por ejemplo que codifica un segundo polipéptido marcador seleccionable o un segundo polipéptido de interés. Tal secuencia codificante extra en dirección 3' de la secuencia codificante para el primer polipéptido de interés está precedida entonces preferiblemente por una secuencia que codifica un sitio de reinicio de la traducción o sitio de entrada del ribosoma interno (IRES), de manera que también el segundo polipéptido de interés o el segundo polipéptido marcador seleccionable es traducido, en este caso
15 independientemente de las dos secuencias codificantes anteriores. En tal realización, es posible por ejemplo que las secuencias codificantes para subunidades de una proteína multimérica, por ejemplo una cadena pesada y una ligera de un anticuerpo, sean codificadas como un primer y segundo polipéptido de interés (en cualquier orden), para codificar subunidades de una proteína multimérica en una única unidad de expresión multicistrónica. Sin embargo, se prefiere usar unidades de transcripción separadas para la expresión de diferentes polipéptidos de interés,
20 también cuando estos forman parte de una proteína multimérica (véase, por ejemplo, el ejemplo 13: la cadena pesada y ligera de un anticuerpo están codificadas cada una por una unidad de transcripción separada, siendo cada una de estas unidades de expresión una unidad de expresión bicistrónica según la invención).

25 Las moléculas de ADN de la invención pueden estar presentes en forma de ADN bicatenario, que tiene con respecto al polipéptido marcador seleccionable y el polipéptido de interés una hebra codificante y una hebra no codificante, siendo la hebra codificante la hebra con la misma secuencia que el ARN traducido, excepto por la presencia de T en lugar de U. Por tanto, un codón de inicio AUG está codificado en la hebra codificante por una secuencia ATG, y la hebra que contiene esta secuencia ATG correspondiente al codón de inicio AUG en el ARN se designa como la hebra codificante del ADN. Resultará evidente que los codones de inicio o secuencias de iniciación de la traducción están de hecho presentes en una molécula de ARN, pero que estos pueden considerarse igualmente materializados
30 en una molécula de ADN que codifica dicha molécula de ARN; por tanto, siempre que la presente invención se refiera a un codón de inicio o secuencia de iniciación de la traducción, se pretende incluir la molécula de ADN correspondiente que tiene la misma secuencia que la molécula de ARN excepto por la presencia de una T en lugar de una U en la hebra codificante de dicha molécula de ADN, y viceversa, excepto cuando se especifique explícitamente otra cosa. En otras palabras, un codón de inicio es por ejemplo una secuencia AUG en ARN, pero la secuencia ATG correspondiente en la hebra codificante de ADN se designa igualmente como codón de inicio en la presente invención. Se usa lo mismo para la referencia de secuencias codificantes "en marco", lo que significa tripletes (3 bases) en la molécula de ARN que se traducen a un aminoácido, pero también han de interpretarse como las secuencias trinucleotídicas correspondientes en la hebra codificante de la molécula de ADN.

40 El polipéptido marcador seleccionable y el polipéptido de interés codificados por el gen multicistrónico tienen cada uno su propia secuencia de iniciación de la traducción, y por lo tanto tienen cada uno su propio codón de inicio (así como codón de parada), es decir, están codificados por marcos de lectura abiertos separados.

45 La expresión "marcador de selección" o "marcador seleccionable" se usa típicamente para referirse a un gen y/o proteína cuya presencia puede detectarse directa o indirectamente en una célula, por ejemplo un polipéptido que inactiva un agente de selección y protege a la célula hospedante de los efectos letales o inhibidores del crecimiento del agente (por ejemplo, un gen y/o proteína de resistencia a antibiótico). Otra posibilidad es que dicho marcador de selección induzca fluorescencia o un depósito de color (por ejemplo, proteína fluorescente verde (GFP) y derivados (por ejemplo, d2EGFP), luciferasa, lacZ, fosfatasa alcalina, etc.), que puede usarse para seleccionar células que expresan el polipéptido inductor del depósito de color, por ejemplo usando un clasificador celular activado por fluorescencia (FACS) para seleccionar células que expresan GFP. Preferiblemente, el polipéptido marcador
50 seleccionable según la invención proporciona resistencia frente a los efectos letales y/o inhibidores del crecimiento de un agente de selección. El polipéptido marcador seleccionable está codificado por el ADN de la invención. El polipéptido marcador seleccionable según la invención debe ser funcional en una célula hospedante eucariota, y por tanto poder seleccionarse en células hospedantes eucariotas. En principio, según la presente invención, puede usarse cualquier polipéptido marcador seleccionable que cumpla este criterio. Dichos polipéptidos marcadores seleccionables son bien conocidos en la técnica y se usan rutinariamente cuando han de obtenerse clones de células hospedantes eucariotas, y se proporcionan varios ejemplos en la presente memoria. En ciertas realizaciones, un marcador de selección usado para la invención es zeocina. En otras realizaciones, se usa blasticidina. El especialista en la técnica sabrá que existen y pueden usarse otros marcadores de selección, por ejemplo neomicina, puromicina, bleomicina, higromicina, etc. En otras realizaciones, se usa kanamicina. En aún otras realizaciones, se usa el gen DHFR como marcador seleccionable, que puede seleccionarse mediante metotrexato, especialmente aumentando la concentración de metotrexato pueden seleccionarse células para números aumentados de copias del gen DHFR. De forma similar, puede usarse el gen de glutamina sintetasa (GS), para el que es posible la selección
60

en células que tienen insuficiente GS (por ejemplo, células NS-0) cultivando en medio sin glutamina o, como alternativa, en células que tienen suficiente GS (por ejemplo, células CHO) añadiendo un inhibidor de GS, la metionina sulfoximina (MSX). En la tabla 1 de la patente de U.S. 5.561.053 se describen, por ejemplo, otros genes marcadores seleccionables que podrían usarse, y sus agentes de selección; véase también Kaufman, *Methods in Enzymology*, 185:537-566 (1990), para una revisión de estos.

Cuando han de seleccionarse dos unidades de transcripción multicistrónicas según la invención en una sola célula hospedante, cada una contiene preferiblemente la secuencia codificante de un marcador seleccionable diferente, para permitir la selección de ambas unidades de transcripción multicistrónicas. Por supuesto, ambas unidades de transcripción multicistrónicas pueden estar presentes en una sola molécula de ácido nucleico o, como alternativa, cada una puede estar presente en una molécula de ácido nucleico separada.

El término "selección" se define típicamente como el proceso de usar un marcador de selección/marcador seleccionable y un agente de selección para identificar células hospedantes con propiedades genéticas específicas (por ejemplo, que la célula hospedante contenga un transgén integrado en su genoma). Resulta evidente para un especialista en la técnica que son posibles numerosas combinaciones de marcadores de selección. Un antibiótico que es particularmente ventajoso es zeocina, porque la proteína de resistencia a zeocina (zeocina-R) actúa uniéndose al fármaco y volviéndolo inofensivo. Por lo tanto, es fácil valorar la cantidad de fármaco que extermina las células con bajos niveles de expresión de zeocina-R, mientras que permite sobrevivir a las de alta expresión. Todas las demás proteínas de resistencia a antibiótico de uso común son enzimas, y de este modo actúan catalíticamente (no 1:1 con el fármaco). Por tanto, el antibiótico zeocina es un marcador de selección preferido. Sin embargo, la invención funciona también con otros marcadores de selección.

Un polipéptido marcador seleccionable según la invención es la proteína que es codificada por el ácido nucleico de la invención, polipéptido el cual puede detectarse, por ejemplo debido a que proporciona resistencia a un agente de selección tal como un antibiótico. Por tanto, cuando se usa un antibiótico como agente de selección, el ADN codifica un polipéptido que confiere resistencia al agente de selección, polipéptido el cual es el polipéptido marcador seleccionable. Se conocen las secuencias de ADN que codifican dichos polipéptidos marcadores seleccionables, y en la presente memoria se proporcionan varios ejemplos de secuencias de tipo salvaje de ADN que codifica proteínas marcadoras seleccionables (Fig. 26-32). Resultará evidente que los mutantes o derivados de los marcadores seleccionables pueden usarse también adecuadamente según la invención, y están por lo tanto incluidos dentro del alcance de la expresión "polipéptido marcador seleccionable", a condición de que la proteína marcadora seleccionable todavía sea funcional. Por conveniencia, y como se acepta generalmente por los expertos, en muchas publicaciones así como en la presente memoria, a menudo el gen y la proteína que codifica resistencia a un agente de selección se designan como el "gen (de resistencia) del agente seleccionable" o la "proteína (de resistencia) del agente de selección", respectivamente, aunque los nombres oficiales pueden ser diferentes, por ejemplo el gen que codifica la proteína que confiere resistencia a neomicina (así como a G418 y kanamicina) se designa a menudo como gen (de resistencia) de neomicina (o neo^r), mientras que el nombre oficial es gen de aminoglicosido 3'-fosfotransferasa.

Para la presente invención, es beneficioso tener bajos niveles de expresión del polipéptido marcador seleccionable, de modo que sea posible una selección restrictiva. En la presente invención, esto se produce usando una secuencia codificante marcadora seleccionable con una eficiencia de traducción no óptima. Tras la selección, se seleccionarán solo las células que tengan no obstante niveles suficientes de polipéptido marcador seleccionable, lo que significa que dichas células deben tener suficiente transcripción de la unidad de transcripción multicistrónica y suficiente traducción del polipéptido marcador seleccionable, lo que proporciona una selección de células en las que la unidad de transcripción multicistrónica se ha integrado o está presente de otro modo en las células hospedantes en un lugar en el que los niveles de expresión de esta unidad de transcripción son elevados. Según la invención, la naturaleza (más o menos) recíprocamente interdependiente de la traducción del polipéptido de interés de aquella de la eficiencia de la traducción del marcador asegura niveles de traducción elevados del polipéptido de interés. Este nivel elevado de traducción se superpone sobre los niveles elevados de transcripción seleccionados por las condiciones de selección restrictivas (y que resultan del uso de un fuerte promotor en las células hospedantes para conducir la transcripción de la unidad de transcripción multicistrónica).

Las moléculas de ADN según la invención tienen la secuencia codificante para el polipéptido marcador seleccionable en dirección 5' de la secuencia codificante para el polipéptido de interés, para proporcionar un transcrito multicistrónico con traducción recíprocamente interdependiente del polipéptido de interés y el polipéptido marcador seleccionable. Por tanto, la unidad de transcripción multicistrónica comprende, en la dirección 5' a 3' (tanto en la hebra transcrita del ADN como en el ARN transcrito resultante) la secuencia codificante para el polipéptido marcador seleccionable y la secuencia codificante para el polipéptido de interés.

Según la invención, las regiones codificantes para el gen multicistrónico se traducen preferiblemente a partir del ORF dependiente de caperuza, y por lo tanto el polipéptido marcador seleccionable en principio también se produciría en abundancia. Para prevenir tal traducción elevada del cistron marcador seleccionable, según la invención la secuencia de ácido nucleico que codifica el polipéptido marcador seleccionable comprende una mutación (en el marco de lectura abierto (ORF) o en las secuencias de iniciación de la traducción que preceden al ORF, o en ambos) que reduce la eficiencia de la traducción del polipéptido marcador seleccionable en una célula

hospedante eucariota. La eficiencia de traducción debería ser menor que la de la secuencia de tipo salvaje correspondiente en la misma célula, es decir, la mutación debería dar como resultado menos polipéptido por célula por unidad de tiempo, y por tanto menos polipéptido marcador seleccionable. Este se puede detectar usando métodos habituales conocidos por la persona experta en la técnica. Por ejemplo, en el caso de selección con antibióticos, la mutación dará como resultado menor resistencia que la obtenida con la secuencia que no tiene tal mutación y por tanto eficiencia de traducción normal, diferencia la cual se puede detectar fácilmente determinando el número de colonias supervivientes tras un periodo de selección normal, que será menor cuando está presente una mutación que disminuye la eficiencia de la traducción. Como es bien conocido por la persona experta en la técnica, hay un número de parámetros que indican el nivel de expresión del polipéptido marcador, tales como la concentración máxima de agente de selección al que todavía son resistentes las células, el número de colonias supervivientes a una concentración dada, la velocidad de crecimiento (tiempo de doblaje) de las células en presencia de agente de selección, combinaciones de los anteriores, y similares.

Según la invención, la mutación que disminuye la eficiencia de la traducción es una mutación que disminuye la eficiencia de iniciación de la traducción. Esta se puede establecer, por ejemplo, según la presente invención, proporcionando la secuencia codificante del polipéptido marcador seleccionable con una secuencia de inicio de la traducción no óptima.

Según la invención, la eficiencia de iniciación de la traducción del gen marcador seleccionable en células eucariotas se puede disminuir adecuadamente según la invención mutando el codón de inicio y/o los nucleótidos en las posiciones -3 a -1 y +4 (en las que la A del codón de inicio ATG es nt +1), por ejemplo en la hebra codificante de la secuencia de ADN correspondiente, para proporcionar una secuencia de inicio de la traducción no óptima. Una secuencia de inicio de la traducción se designa a menudo en el campo como una "secuencia Kozak", y una secuencia Kozak óptima es RCCATGG, con el codón de inicio subrayado, siendo R una purina, es decir, A o G (véanse Kozak M, 1986, 1987, 1989, 1990, 1997, 2002). Por tanto, además del propio codón de inicio, es relevante el contexto del mismo, en particular los nucleótidos -3 a -1 y +4, y una secuencia de inicio de la traducción óptima comprende un codón de inicio óptimo (es decir, ATG) en un contexto óptimo (es decir, el ATG precedido directamente por RCC y seguido directamente por G). Una secuencia de inicio de la traducción no óptima se define aquí como cualquier secuencia que da al menos alguna traducción detectable en una célula eucariota (detectable debido a que el polipéptido marcador de selección es detectable), y que no tiene la secuencia de consenso RCCATGG (codón de inicio subrayado). La traducción por los ribosomas comienza preferiblemente en el primer codón de inicio que encuentra cuando escanea la caperuza de 5' del ARNm (mecanismo de escaneado), y es lo más eficaz cuando está presente una secuencia Kozak óptima (véanse Kozak M, 1986, 1987, 1989, 1990, 1997, 2002). Sin embargo, en un pequeño porcentaje de sucesos, se reconocen secuencias de iniciación de la traducción no óptimas y son usadas por el ribosoma para iniciar la traducción. La presente invención hace uso de este principio, y permite reducir e incluso ajustar finamente la cantidad de traducción y por tanto de la expresión del polipéptido marcador seleccionable, que puede usarse por lo tanto para aumentar la restricción del sistema de selección.

En una primera realización de la invención, el codón de inicio ATG del polipéptido marcador seleccionable (en la hebra codificante del ADN, que codifica el codón de inicio AUG correspondiente en el producto de transcripción de ARN) se deja intacto, pero las posiciones en -3 a -1 y +4 están mutadas de manera que ya no cumplen la secuencia Kozak óptima, por ejemplo proporcionando la secuencia TTTATGT como el sitio de inicio de la traducción (codón de inicio ATG subrayado). Está claro que se podrían usar otras mutaciones alrededor del codón de inicio en las posiciones -3 a -1 y/o +4 con resultados similares usando las enseñanzas de la presente invención, como se puede evaluar normal y fácilmente por la persona experta en la técnica. La idea de esta primera realización es que el codón de inicio ATG se coloca en un contexto "no óptimo" para el inicio de la traducción.

En una realización segunda y preferida, el propio codón de inicio ATG del polipéptido marcador seleccionable está mutado. Esto conducirá en general a niveles incluso menores de iniciación de la traducción que la primera realización. El codón de inicio ATG en esta segunda realización se muta a otro codón, lo que se ha informado que proporciona cierta iniciación de la traducción, por ejemplo a GTG, TTG, CTG, ATT o ACG (designados colectivamente en la presente memoria como "codones de inicio no óptimos"). En realizaciones preferidas, el codón de inicio ATG se muta a un codón de inicio GTG. Esto proporciona niveles de expresión aún menores (menor traducción) que con el codón de inicio ATG intacto, pero en un contexto no óptimo. Más preferiblemente, el codón de inicio ATG se muta a un codón de inicio TTG, que proporciona niveles de expresión aún menores del polipéptido marcador seleccionable que con el codón de inicio GTG (Kozak M, 1986, 1987, 1989, 1990, 1997, 2002; véanse también los ejemplos 9-13 aquí).

Para la segunda realización, es decir, cuando se usa un codón de inicio no ATG, es muy preferido proporcionar un contexto óptimo para dicho codón de inicio, es decir, los codones de inicio no óptimos preferiblemente están precedidos directamente por los nucleótidos RCC en posiciones -3 a -1 y están seguidos directamente por un nucleótido G (posición +4). Sin embargo, se ha informado que usando la secuencia TTGTGG (codón de inicio subrayado), se observa cierta iniciación al menos in vitro, así que, aunque muy preferido, puede no ser absolutamente necesario proporcionar un contexto óptimo para los codones de inicio no óptimos.

Las secuencias de ATG en la secuencia codificante para un polipéptido, pero que excluyen el codón de inicio ATG, se designan como "ATG internos", y si estas están en marco con el ORF y codifican por lo tanto metionina, la

metionina resultante en el polipéptido se designa como "metionina interna". Según la invención, la región codificante (después del codón de inicio, no incluyendo necesariamente el codón de inicio) que codifica el polipéptido marcador seleccionable está desprovista preferiblemente de cualquier secuencia ATG en la hebra codificante del ADN, hasta (pero sin incluir) el codón de inicio del polipéptido de interés (obviamente, el codón de inicio del polipéptido de interés puede ser, y de hecho es preferiblemente, un codón de inicio ATG). Esto se puede establecer mutando cualquiera de tal secuencia ATG en la secuencia codificante del polipéptido marcador seleccionable, tras su codón de inicio (como está claro a partir de las enseñanzas anteriores, el codón de inicio del propio polipéptido marcador seleccionable puede ser una secuencia ATG, aunque no es necesariamente así). Para este fin, preferiblemente se usa la degeneración del código genético para evitar mutar siempre que sea posible aminoácidos en el polipéptido marcador seleccionable. Por tanto, siempre que esté presente una ATG en la hebra codificante de la secuencia de ADN que codifica el polipéptido marcador seleccionable, ATG la cual no está en marco con el ORF del polipéptido marcador seleccionable, y por lo tanto no codifica una metionina interna en el polipéptido marcador seleccionable, la ATG se puede mutar de manera que el polipéptido resultante no tenga mutaciones en su secuencia de aminoácidos interna. Cuando la ATG es un codón en marco que codifica una metionina interna, el codón se puede mutar, y el polipéptido mutado resultante se puede comprobar normalmente para determinar la actividad del polipéptido marcador seleccionable. De esta manera, se puede escoger una mutación que conduzca a un polipéptido marcador seleccionable mutado que todavía es activo como tal (pueden existir diferencias cuantitativas, pero aquellas son menos relevantes, y de hecho incluso podría ser beneficioso tener menos variantes activas para los fines de la presente invención; el requisito mínimo es que el polipéptido marcador seleccionable todavía pueda ser seleccionado en células eucariotas). Los aminoácidos valina, treonina, isoleucina y leucina son estructuralmente similares a metionina, y por lo tanto los codones que codifican uno de estos aminoácidos son buenos candidatos de partida para ser ensayados en lugar de metionina en la secuencia codificante después del codón de inicio. Por supuesto, usando las enseñanzas de la presente invención, el experto en la técnica puede ensayar igualmente otros aminoácidos en lugar de metioninas internas, usando técnicas de biología molecular habituales para mutar el ADN codificante, y ensayos habituales para determinar la funcionalidad del polipéptido marcador seleccionable. Además de técnicas de biología molecular habituales para mutar el ADN, actualmente también es posible sintetizar a voluntad (si se requiere usando etapas de subclonación) secuencias de ADN que tienen una longitud suficiente para un ORF de un polipéptido marcador seleccionable, y tales secuencias de ADN sintéticas en la actualidad pueden pedirse comercialmente de diversas compañías. Por tanto, usando las enseñanzas de la presente invención, el experto en la técnica puede diseñar secuencias apropiadas según la invención que codifican un polipéptido marcador seleccionable (con una mutación que disminuye el inicio de la traducción, y preferiblemente que no tienen secuencias ATG internas), hacer que se sintetice esta secuencia, y someter a ensayo la molécula de ADN para determinar la funcionalidad del marcador seleccionable codificado introduciendo la molécula de ADN en células hospedantes eucariotas y someter a ensayo para determinar la expresión del polipéptido marcador seleccionable funcional. La disponibilidad comercial de tales secuencias también hace factible proporcionar sin demasiado problema secuencias codificantes marcadoras de selección que carecen de secuencias ATG internas, en las que la secuencia codificante de tipo salvaje del polipéptido marcador de selección comprende varias secuencias ATG internas de este tipo.

Proporcionando una secuencia codificante para un polipéptido marcador seleccionable que carece de cualquier secuencia ATG interna, se reducen las ocasiones del inicio inadvertido de la traducción por ribosomas que pasan la (primera) secuencia (no óptima) de inicio de la traducción del polipéptido marcador seleccionable en un trinucleótido ATG interno, de manera que los ribosomas continuarán barriendo la primera secuencia de inicio de la traducción óptima, es decir, la del polipéptido de interés.

Estará claro que no se requiere necesariamente de forma estricta eliminar todas las secuencias ATG internas de la secuencia codificante del marcador seleccionable, en tanto que la inmensa mayoría de ribosomas no usarán tales secuencias ATG internas como un sitio de inicio de la traducción, por ejemplo debido a que tales ATG internas habitualmente no están en un contexto óptimo para el inicio de la traducción. Sin embargo, también a la vista del hecho de que incluso una secuencia ATG en un contexto no óptimo da lugar a sucesos de inicio de la traducción significativos (en general mucho mayores que a partir de un codón de inicio no ATG en un contexto óptimo), también está claro que si están presentes varias secuencias ATG internas en la región codificante del marcador seleccionable, cada una de éstas puede sustraer una fracción de ribosomas que comienzan la traducción de esas ATG, evitando que alcance la unidad de traducción en dirección 3' que codifica el polipéptido de interés, y por tanto da como resultado menor expresión del polipéptido de interés. Esta es la razón por la que se prefiere tener una secuencia que codifica el polipéptido marcador seleccionable, secuencia la cual está libre de secuencias ATG internas. Posiblemente, si se desea, la relación de ribosomas que traducen el polipéptido de interés puede incluso potenciarse adicionalmente mutando cualesquiera sitios de inicio de la traducción potenciales no óptimos en la secuencia que codifica el polipéptido marcador seleccionable.

Claramente, cualquier secuencia presente en la hebra codificante entre el codón de parada del marcador seleccionable y el codón de inicio del polipéptido de interés se puede diseñar de manera que no comprenda secuencias ATG, u otros codones de inicio de la traducción potenciales no óptimos.

Claramente, se prefiere enormemente según la presente invención que la secuencia de inicio de la traducción del polipéptido de interés comprenda una secuencia de inicio de la traducción óptima, es decir, que tenga la secuencia

de consenso RCCATGG (codón de inicio subrayado). Esto dará como resultado una traducción muy eficiente del polipéptido de interés.

5 Como consecuencia de estas medidas, se proporciona el mecanismo de traducción recíprocamente interdependiente de la invención: los ribosomas barren desde la caperuza del ARNm que contiene la unidad de transcripción multicistrónica, y un pequeño porcentaje de ribosomas traducirá el primer marco de lectura abierto, es decir, la proteína marcadora seleccionable, debido a que se proporciona con una mutación que disminuye la eficiencia de la traducción, mientras que la inmensa mayoría de los ribosomas pasarán este marco de lectura abierto y comenzarán la traducción en el sitio de inicio de la traducción óptimo del polipéptido de interés, dando como resultado niveles bajos de proteína marcadora seleccionable y niveles elevados de polipéptido de interés.

10 Proporcionando la secuencia codificante del marcador con diferentes mutaciones que conducen a varios niveles de eficiencia reducida de la traducción, se incrementa la restricción de la selección y aumentan recíprocamente los niveles de expresión del polipéptido de interés, de manera que es posible el ajuste fino: por ejemplo usando un codón de inicio TTG para el polipéptido marcador de selección, sólo muy pocos ribosomas traducirán a partir de este codón de inicio, dando como resultado niveles muy bajos de proteína marcadora seleccionable, y por tanto una

15 elevada restricción de selección, y al mismo tiempo casi todos los ribosomas traducirán el polipéptido de interés.

Se demuestra aquí que esto se puede usar en un sistema de selección muy robusto, conduciendo a un porcentaje muy grande de clones que expresan el polipéptido de interés a niveles elevados, según se desee. Además, los niveles de expresión obtenidos para el polipéptido de interés parecen ser significativamente mayores que los obtenidos cuando se criba un número aún mayor de colonias usando los sistemas de selección conocidos hasta

20 ahora.

Además de una eficiencia reducida de la iniciación de la traducción, podría ser beneficioso proporcionar también una eficiencia reducida del alargamiento de la traducción del polipéptido marcador seleccionable, por ejemplo mutando la secuencia codificante del mismo de modo que comprenda varios codones no preferidos de la célula hospedante, para reducir adicionalmente los niveles de traducción del polipéptido marcador y permitir condiciones de selección aún más rigurosas, si se desea. En ciertas realizaciones, además de la mutación o mutaciones que reducen la eficiencia de la traducción según la invención, el polipéptido marcador seleccionable comprende adicionalmente una mutación que reduce la actividad del polipéptido marcador seleccionable en comparación con su contrapartida de tipo salvaje. Esto puede usarse para aumentar la restricción de la selección más aún. Como ejemplos no limitantes, puede mutarse la prolina en posición 9 en el polipéptido de resistencia a zeocina, por ejemplo, a Thr o Phe, y para el polipéptido de resistencia a neomicina, pueden mutarse adicionalmente el resto de aminoácido 182 ó 261 o ambos (véase, por ejemplo, el documento WO 01/32901).

25

30

En algunas realizaciones de la invención, se coloca una secuencia denominada espaciadora en dirección 3' de la secuencia que codifica el codón de inicio del polipéptido marcador seleccionable, secuencia espaciadora la cual es preferiblemente una secuencia en marco con el codón de inicio y que codifica unos pocos aminoácidos, y que no contiene una estructura secundaria (Kozak, 1990), y no contiene la secuencia ATG. Dicha secuencia espaciadora puede usarse para reducir adicionalmente la frecuencia de iniciación de la traducción si está presente una estructura secundaria en el ARN (Kozak, 1990) del polipéptido marcador seleccionable (por ejemplo, para zeocina, posiblemente de blastidina), y por tanto aumentar el rigor del sistema de selección según la invención.

35

La invención también proporciona una molécula de ADN que comprende la secuencia que codifica una proteína marcadora seleccionable funcional según la invención, molécula de ADN la cual:

40

i) tiene una secuencia de inicio de la traducción escogida del grupo que consiste en:

a) una secuencia de inicio de la traducción que comprende una mutación en los nucleótidos en las posiciones -3 a -1 y/o +4 de la secuencia de consenso RCCATGG (en la que R es A o G, y en la que la A de la secuencia de inicio ATG es nt +1);

45 b) un codón de inicio GTG;

c) un codón de inicio TTG;

d) un codón de inicio CTG;

e) un codón de inicio ATT; y

50 f) un codón de inicio ACG, seguido de una secuencia codificante marcadora seleccionable de otro modo funcional, y

ii) porque la hebra codificante de la secuencia que define el polipéptido marcador seleccionable en dirección 3' del codón de inicio no óptimo y hasta el codón de parada está desprovista de secuencias ATG. Tales moléculas de ADN engloban un producto intermedio útil según la invención. Estas moléculas se pueden preparar primero, se pueden introducir en células hospedantes eucariotas y se pueden ensayar para determinar la funcionalidad (para algunos marcadores esto es posible incluso en células hospedantes

55

procariotas), si se desea de manera (semi)cuantitativa, del polipéptido marcador seleccionable. Entonces pueden usarse adicionalmente para preparar una molécula de ADN según la invención, que comprende la unidad de transcripción multicistrónica.

5 Es un aspecto preferido de la invención proporcionar un casete de expresión que comprende la molécula de ADN según la invención, que tiene la unidad de transcripción multicistrónica. Dicho casete de expresión es útil para expresar secuencias de interés, por ejemplo en células hospedantes. Un "casete de expresión", como se usa en la presente memoria, es una secuencia de ácido nucleico que comprende al menos un promotor ligado funcionalmente a una secuencia de la que se desea la expresión. Preferiblemente, un casete de expresión contiene adicionalmente 10 secuencias de terminación de la transcripción y de poliadenilación. Pueden incluirse también otras secuencias reguladoras tales como potenciadores. Por tanto, la invención proporciona un casete de expresión que comprende en el siguiente orden: 5'- promotor - unidad de transcripción multicistrónica según la invención, que codifica un polipéptido marcador seleccionable y en dirección 3' de la misma un polipéptido de interés - secuencia de terminación de la transcripción - 3'. El promotor debe ser capaz de funcionar en una célula hospedante eucariota, es decir, debe ser capaz de dirigir la transcripción de la unidad de transcripción multicistrónica. El casete de expresión 15 puede contener opcionalmente otros elementos conocidos en la técnica, por ejemplo sitios de corte y empalme para comprender intrones, y similares. En algunas realizaciones, está presente un intrón detrás del promotor y antes de la secuencia que codifica el polipéptido de interés.

20 Para obtener la expresión de secuencias de ácidos nucleicos que codifican proteína, es bien conocido por los expertos en la técnica que las secuencias capaces de dirigir dicha expresión pueden estar ligadas funcionalmente con las secuencias de ácidos nucleicos que codifican la proteína, dando como resultado moléculas de ácidos nucleicos recombinantes que codifican una proteína en formato expresable. En la presente invención, el casete de expresión comprende una unidad de transcripción multicistrónica. En general, la secuencia promotora se coloca en dirección 5' de las secuencias que deberían expresarse. Están disponibles en la técnica vectores de expresión muy usados, por ejemplo, la serie de vectores pcDNA y pEF de Invitrogen, pMSCV y pTK-Hyg de BD Sciences, pCMV-Script de Stratagene, etc., que pueden usarse para obtener secuencias promotoras y/o terminadoras de la 25 transcripción adecuadas, secuencias de poliA, y similares.

30 Cuando la secuencia que codifica el polipéptido de interés se inserta apropiadamente con referencia a las secuencias que gobiernan la transcripción y traducción del polipéptido codificado, el casete de expresión resultante es útil para producir el polipéptido de interés, designado como expresión. Las secuencias que dirigen la expresión pueden incluir promotores, potenciadores y similares, y combinaciones de los mismos. Estos deberían ser capaces de funcionar en la célula hospedante, dirigiendo de ese modo la expresión de las secuencias de ácidos nucleicos que están ligadas funcionalmente a ellos. La persona experta en la técnica es consciente de que pueden usarse diversos promotores para obtener la expresión de un gen en células hospedantes. Los promotores pueden ser 35 constitutivos o regulados, y pueden obtenerse de diversas fuentes, incluyendo virus, fuentes procariotas o eucariotas, o pueden diseñarse artificialmente. La expresión de los ácidos nucleicos de interés puede ser a partir del promotor natural o un derivado del mismo, o a partir de un promotor enteramente heterólogo (Kaufman, 2000). Algunos promotores bien conocidos y muy usados para la expresión en células eucariotas comprenden promotores derivados de virus, tales como adenovirus, por ejemplo el promotor E1A, promotores derivados de citomegalovirus (CMV), tales como el promotor inmediato temprano (IE) de CMV (designado en la presente memoria como el promotor de CMV) (obtenible, por ejemplo, a partir de pcDNA, Invitrogen), promotores derivados del Virus del Simio 40 (SV40) (Das et al, 1985), y similares. Los promotores adecuados también se pueden derivar de células eucariotas, tales como promotores de metalotioneína (MT), el promotor del factor 1 α de alargamiento (EF-1 α) (Gill et al., 2001), el promotor de ubiquitina C o UB6 (Gill et al., 2001; Schorpp et al, 1996), el promotor de actina, un promotor de inmunoglobulina, promotores del choque término, y similares. Algunos promotores preferidos para obtener la expresión en células eucariotas, que son promotores adecuados en la presente invención, son el 45 promotor de CMV, un promotor de EF1-alfa de mamífero, un promotor de ubiquitina de mamífero tal como un promotor de ubiquitina C, o un promotor de SV40 (por ejemplo, obtenible de pIRES, n° de cat. 631605, BD Sciences). El ensayo de la función del promotor y la fortaleza de un promotor es una cuestión de rutina para una persona experta en la técnica, y en general puede comprender, por ejemplo, clonar un gen de ensayo, tal como lacZ, luciferasa, GFP, etc., detrás de la secuencia promotora, y ensayar la expresión del gen de ensayo. Por supuesto, los promotores se pueden alterar por supresión, adición, mutación de las secuencias allí, y se pueden ensayar para 50 determinar la funcionalidad, para encontrar secuencias promotoras nuevas, atenuadas, o mejoradas. Según la presente invención, se prefieren promotores fuertes que dan niveles elevados de transcripción en las células eucariotas de elección.

55 En ciertas realizaciones, una molécula de ADN según la invención es parte de un vector, por ejemplo un plásmido. Dichos vectores pueden manipularse fácilmente mediante métodos bien conocidos por la persona experta en la técnica, y pueden diseñarse, por ejemplo, para que sean capaces de replicarse en células procariotas y/o eucariotas. Además, muchos vectores pueden usarse directamente o en forma de un fragmento deseado aislado a partir del mismo para la transformación de células eucariotas, y se integrarán en todo o en parte en el genoma de 60 dichas células, dando como resultado células hospedantes estables que comprenden el ácido nucleico deseado en su genoma.

Los sistemas de expresión convencionales son moléculas de ADN en forma de un plásmido recombinante o un genoma vírico recombinante. El plásmido o el genoma vírico se introduce en células (hospedantes eucariotas) y preferiblemente se integra en sus genomas mediante métodos conocidos en la técnica. En realizaciones preferidas, la presente invención también usa estos tipos de moléculas de ADN para suministrar su sistema de expresión transgénico mejorado. Una realización preferida de la invención es el uso de ADN plasmídico para suministrar el sistema de expresión. Un plásmido contiene un número de componentes: los componentes convencionales, conocidos en la técnica, son un origen de replicación y un marcador seleccionable para la propagación del plásmido en células bacterianas; un marcador seleccionable que funciona en células eucariotas para identificar y aislar células hospedantes que poseen un sistema de expresión transgénico integrado; la proteína de interés, cuya transcripción a alto nivel es provocada por un promotor que es funcional en células eucariotas (por ejemplo el promotor/potenciador temprano inmediato principal de citomegalovirus humano, pCMV (Boshart et al., 1985); y terminadores transcripcionales víricos (por ejemplo el sitio de poliadenilación de SV40 (Kaufman y Sharp, 1982) para el transgén de interés y el marcador seleccionable.

El vector usado puede ser cualquier vector que sea adecuado para clonar ADN y que puede usarse para la transcripción de un ácido nucleico de interés. Cuando se usan células hospedantes, se prefiere que el vector sea un vector de integración. Alternativamente, el vector puede ser un vector que se replica episómicamente.

Es ampliamente apreciado que la estructura de cromatina y otros mecanismos de control epigenéticos pueden influir en la expresión de transgenes en células eucariotas (por ejemplo, Whitelaw et al., 2001). Las unidades de expresión multicistónicas según la invención forman parte de un sistema de selección con un régimen de selección más bien riguroso. Esto requiere generalmente niveles elevados de transcripción en las células hospedantes de elección. Para aumentar la oportunidad de encontrar clones de células hospedantes que sobrevivan al régimen de selección riguroso, y posiblemente para aumentar la estabilidad de la expresión en los clones obtenidos, será preferible generalmente aumentar la predecibilidad de la transcripción. Por lo tanto, en realizaciones preferidas, un casete de expresión según la invención comprende adicionalmente al menos un elemento de control de cromatina. Un "elemento de control de cromatina", como se usa en la presente memoria, es un término colectivo para secuencias de ADN que pueden tener de algún modo un efecto sobre la estructura de cromatina y con ello sobre el nivel de expresión y/o estabilidad de la expresión de los transgenes en su vecindad (funcionan "en cis", y por tanto se colocan preferiblemente a 5 kb, más preferiblemente a 2 kb, aún más preferiblemente a 1 kb del transgén) en células eucariotas. Dichos elementos se han usado algunas veces para aumentar el número de clones que tienen los niveles deseados de expresión transgénica. Los mecanismos mediante los cuales estos elementos funcionan pueden diferir para e incluso en diferentes clases de tales elementos, y no son completamente conocidos para todos los tipos de dichos elementos. Sin embargo, tales elementos se han descrito, y para los fines de la presente invención se eligen elementos de control de cromatina del grupo que consiste en regiones de unión a matriz o armazón (MARs/SARs) (por ejemplo, Phi-Van et al, 1990; documentos WO 02/074969, WO 2005/040377), aislantes (West et al, 2002), tales como el elemento aislante de beta-globina (5' HS4 del locus de beta-globina de pollo), scs, scs' y similares (por ejemplo, Chung et al, 1993, 1997; Kellum y Schedl, 1991; documentos WO 94/23046, WO 96/04390, WO 01/02553, WO 2004/027072), un elemento de apertura de cromatina ubicuo (UCOE) (documentos WO 00/05393, WO 02/24930, WO 02/099089, WO 02/099070), y secuencias antirrepresoras (también designadas como secuencias "STAR") (Kwaks et al, 2003; documento WO 03/004704). Los ejemplos no limitantes de secuencias MAR/SAR que se podrían usar en la actual invención son 5' MAR de lisozima de pollo (Phi-Van et al, 1990) o sus fragmentos, por ejemplo las regiones B, K, y F, como se describe en el documento WO 02/074969; secuencias de ADN que comprenden al menos un elemento de ADN doblado y al menos un sitio de unión para una proteína de unión a ADN, preferiblemente que contiene al menos 10% de dinucleótido TA, y/o al menos 12% de dinucleótido AT en un tramo de 100 pares de bases contiguas, tal como una secuencia seleccionada del grupo que comprende las secuencias SEQ ID Nos 1 a 27 en el documento WO 2005/040377, teniendo los fragmentos de una cualquiera de las SEQ ID Nos 1 a 27 en el documento WO 2005/040377 una longitud de al menos 100 nucleótidos y que tienen actividad MAR, secuencias que son al menos 70% idénticas en secuencia nucleotídica a una cualquiera de SEQ ID Nos 1 a 27 en el documento WO 2005/040377 o sus fragmentos y que tienen actividad MAR, en el que la actividad MAR se define como siendo capaz de unirse a matrices/armazones nucleares in vitro y/o de alterar la expresión de secuencias codificantes ligadas operablemente a un promotor; secuencias escogidas de una cualquiera de SEQ ID NO: 1 a 5 en el documento WO 02/074969, fragmentos de una cualquiera de una cualquiera de SEQ ID NO: 1 a 5 en el documento WO 02/074969 y que tienen actividad MAR, secuencias que son al menos 70% idénticas en secuencia nucleotídica a una cualquiera de SEQ ID NO: 1 a 5 en el documento WO 02/074969 o sus fragmentos y que tienen actividad MAR; secuencias escogidas de SEQ ID NO: 1 y SEQ ID NO: 2 en el documento WO 2004/027072, sus fragmentos funcionales y secuencias que son al menos 70% idénticas a ellas. Un ejemplo no limitante de secuencias aislantes que se podrían usar en la presente invención es una secuencia que comprende SEQ ID NO: 1 del documento WO 01/02553. Los ejemplos no limitantes de UCOEs que se podrían usar en la presente invención son secuencias representadas en las Figuras 2 y 7 del documento WO 02/24930, sus fragmentos funcionales y secuencias que son al menos 70% idénticas a ellas mientras que todavía retienen actividad, secuencias que comprenden SEQ ID NO: 28 del documento US 2005/181428, sus fragmentos funcionales y secuencias que son al menos 70% idénticas a ellas mientras todavía retienen actividad.

Preferiblemente, dicho elemento de control de cromatina es una secuencia antirrepresora, preferiblemente elegida del grupo que consiste en: a) una cualquiera de SEQ. ID. NO. 1 a SEQ. ID. NO. 66; b) fragmentos de una cualquiera

de SEQ. ID. NO. 1 a SEQ. ID. NO. 66, en el que dichos fragmentos tienen actividad antirrepresora (“fragmentos funcionales”); c) secuencias que son al menos 70% idénticas en la secuencia nucleotídica a a) o a b), en el que dichas secuencias tienen actividad antirrepresora (“derivados funcionales”); y d) el complemento de una cualquiera de a) a c). Preferiblemente, dicho elemento de control de cromatina se elige del grupo que consiste en STAR67 (SEQ. ID. NO. 66), STAR7 (SEQ. ID. NO. 7), STAR9 (SEQ. ID. NO. 9), STAR 17 (SEQ. ID. NO. 17), STAR27 (SEQ. ID. NO. 27), STAR29 (SEQ. ID. NO. 29), STAR43 (SEQ. ID. NO. 43), STAR44 (SEQ. ID. NO. 44), STAR45 (SEQ. ID. NO. 45), STAR47 (SEQ. ID. NO. 47), STAR61 (SEQ. ID. NO. 61), o un fragmento funcional o derivado de dichas secuencias STAR. En una realización particularmente preferida, dicha secuencia STAR es STAR67 (SEQ. ID. NO. 66), o un fragmento funcional o derivado de la misma. En ciertas realizaciones preferidas, STAR67 o un fragmento funcional o derivado de la misma está situado en dirección 5' de un promotor que dirige la expresión de la unidad de transcripción multicistónica. En otras realizaciones preferidas, los casetes de expresión según la invención están flanqueados en ambos lados por al menos una secuencia antirrepresora.

Las secuencias que tienen actividad antirrepresora como se usa aquí son secuencias que son capaces de contrarrestar al menos en parte el efecto represor de proteínas HP1 o HPC2 cuando estas proteínas se unen a ADN. Las secuencias que tienen actividad antirrepresora (algunas veces también denominadas como secuencias antirrepresoras o elementos antirrepresores aquí) adecuadas para la presente invención se han descrito en el documento WO 03/004704, y se acuñaron allí como secuencias “STAR” (siempre que una secuencia se cite como una secuencia STAR aquí, esta secuencia tiene actividad antirrepresora según la invención). Como ejemplo no limitante, las secuencias de 66 elementos antirrepresores, denominadas STAR1-65 (véase el documento WO 03/004704) y STAR 67, se presentan aquí como SEQ. ID. NOs. 1-65 y 66, respectivamente.

Según la invención, un fragmento funcional o derivado de un elemento antirrepresor dado se considera equivalente a dicho elemento antirrepresor cuando todavía tiene actividad antirrepresora. La presencia de tal actividad antirrepresora se puede comprobar fácilmente por la persona experta en la técnica, por ejemplo mediante el ensayo descrito más abajo. Los fragmentos funcionales o derivados se pueden obtener fácilmente por una persona experta en la técnica de biología molecular partiendo de una secuencia antirrepresora dada, y realizando supresiones, adiciones, sustituciones, inversiones y similares (véase, por ejemplo, el documento WO 03/004704). Un fragmento funcional o derivado también comprende ortólogos de otras especies, que se pueden encontrar usando las secuencias antirrepresoras conocidas mediante métodos conocidos por la persona experta en la técnica (véase, por ejemplo, el documento WO 03/004704). Por tanto, la presente invención engloba fragmentos de las secuencias antirrepresoras, en el que dichos fragmentos todavía tienen actividad antirrepresora. La invención también engloba secuencias que son al menos 70% idénticas en secuencia nucleotídica a dichas secuencias que tienen actividad antirrepresora, o a fragmentos funcionales de las mismas que tienen actividad antirrepresora, en tanto que estas secuencias que son al menos 70% idénticas tengan todavía la actividad antirrepresora según la invención. Preferiblemente, dichas secuencias son al menos 80% idénticas, más preferiblemente al menos 90% idénticas, y todavía más preferiblemente al menos 95% idénticas a la secuencia nativa de referencia o su fragmento funcional. Para fragmentos de una secuencia dada, el porcentaje de identidad se refiere a aquella porción de la secuencia nativa de referencia que se encuentra en el fragmento.

Las secuencias que tienen actividad antirrepresora según la invención se pueden obtener por diversos métodos, incluyendo, pero sin limitarse a, la clonación a partir del genoma humano o a partir del genoma de otro organismo, o, por ejemplo, amplificando secuencias antirrepresoras conocidas directamente a partir de tal genoma usando el conocimiento de las secuencias, por ejemplo mediante PCR, o se pueden sintetizar químicamente en parte o completamente.

Las secuencias que tienen actividad antirrepresora, y sus fragmentos funcionales o derivados, se definen estructuralmente aquí por su secuencia, y además se definen funcionalmente como secuencias que tienen actividad antirrepresora, que se puede determinar con el ensayo descrito más abajo.

Cualquier secuencia que tenga actividad antirrepresora según la presente invención debería ser al menos capaz de sobrevivir al siguiente ensayo funcional (véase el documento WO 03/004704, ejemplo 1).

Células U-2 OS humanas (ATCC HTB-96) se transfectan de forma estable con el plásmido pTet-Off (Clontech K1620-A) y con ácido nucleico que codifica una proteína de fusión LexA-represor que contiene el dominio de unión a ADN LexA y la región codificante de HP1 o HPC2 (proteínas del grupo Polycomb de *Drosophila* que reprimen la expresión génica cuando se unen a ADN; el ensayo funciona con cualquier proteína de fusión) bajo el control del sistema regulador transcripcional Tet-Off (Gossen y Bujard, 1992). Estas células se denominan más abajo como las células informadoras para el ensayo de actividad antirrepresora. Un plásmido informador, que proporciona resistencia a higromicina, contiene una secuencia poliligadora situada entre cuatro sitios del operador LexA y el promotor de SV40 que controla el gen de resistencia a zeocina. La secuencia a ensayar en busca de la actividad antirrepresora se puede clonar en dicho poliligador. La construcción de un plásmido informador adecuado, tal como pSelect, se describe en el ejemplo 1 y en la Fig. 1 del documento WO 00/004704. El plásmido informador se transfecta en las células informadoras, y las células se cultivan bajo selección con higromicina (25 µg/ml; selección en busca de la presencia del plásmido informador) y represión con tetraciclina (doxiciclina, 10 µg/ml; evita la expresión de la proteína de fusión LexA-represor). Después de 1 semana de crecimiento en estas condiciones, la concentración de doxiciclina se reduce hasta 0,1 ng/ml para inducir el gen de LexA, y después de 2 días se añade

zeocina hasta 250 µg/ml. Las células se cultivan durante 5 semanas, hasta que los cultivos de control (transfectados con plásmido informador vacío, es decir, que carece de una secuencia antirrepresora clonada en el poliligador) son exterminados por la zeocina (en este plásmido de control, el promotor de SV40 está representado por la proteína de fusión LexA-represor que está unida a los sitios de operación de LexA, dando como resultado una expresión insuficiente de zeocina en tales células para sobrevivir a la selección con zeocina). Una secuencia tiene actividad antirrepresora según la presente invención si, cuando dicha secuencia es clonada en el poliligador del plásmido informador, las células informadoras sobreviven la selección de 5 semanas bajo zeocina. Las células procedentes de tales colonias todavía se pueden propagar sobre nuevo medio que contiene zeocina después de la selección con zeocina durante 5 semanas, mientras que las células transfectadas con plásmidos informadores que carecen de secuencias antirrepresoras no se pueden propagar sobre nuevo medio que contiene zeocina. Cualquier secuencia no capaz de conferir tal crecimiento después de 5 semanas en zeocina en este ensayo no sirve como secuencia que tiene actividad antirrepresora, o fragmento funcional o derivado funcional de la misma según la presente invención. Como ejemplo, otros elementos de control de cromatina conocidos, tales como los ensayados por Van der Vlag et al (2000), incluyendo scs de *Drosophila* (Kellum y Schedl, 1991), 5'-HS4 del locus de β-globina de pollo (Chung et al, 1993, 1997) o Regiones de Adhesión de la Matriz (MARs) (Phi-Van et al.1990), no sobreviven a este ensayo.

Además, se prefiere que la secuencia antirrepresora o su fragmento funcional o derivado confiera una mayor proporción de clones que sobreexpresan el informador cuando flanquea a un gen informador (por ejemplo luciferasa, GFP) que se integra en el genoma de células U-2 OS o CHO, en comparación a cuando dicho gen informador no está flanqueado por secuencias antirrepresoras, o flanqueado por secuencias que bloquean la expresión más débiles, tales como scs de *Drosophila*. Esto se puede verificar usando, por ejemplo, el vector pSDH, o vectores similares, como se describe en el ejemplo 1 y en la Fig. 2 del documento WO 03/004704. Los elementos antirrepresores pueden tener al menos una de tres consecuencias para la producción de proteína: (1) incrementan la predicibilidad de identificar líneas de células hospedantes que expresan una proteína a niveles industrialmente aceptables (proporciona la capacidad de hematocromatina adyacente para silenciar el transgén, de manera que la posición de integración tiene un efecto menos pronunciado sobre la expresión); (2) dan como resultado líneas de células hospedantes con mayores rendimientos proteicos; y/o (3) dan como resultado líneas de células hospedantes que presentan una producción proteica más estable durante un cultivo prolongado.

Cualquier secuencia STAR se puede usar en los casetes de expresión según la presente invención, pero las siguientes secuencias STAR son particularmente útiles: STAR67 (SEQ. ID. NO. 66), STAR7 (SEQ. ID. NO. 7), STAR9 (SEQ. ID. NO. 9), STAR 17 (SEQ. ID. NO. 17), STAR27 (SEQ. ID. NO. 27), STAR29 (SEQ. ID. NO. 29), STAR43 (SEQ. ID. NO. 43), STAR44 (SEQ. ID. NO. 44), STAR45 (SEQ. ID. NO. 45), STAR47 (SEQ. ID. NO. 47), STAR61 (SEQ. ID. NO. 61), o fragmentos funcionales o derivados de estas secuencias STAR.

En ciertas realizaciones, dicha secuencia antirrepresora, preferiblemente STAR67, se coloca en dirección 5' de dicho promotor, preferiblemente de manera que hay menos de 2kb entre el extremo 3' de la secuencia antirrepresora y el inicio de la secuencia promotora. En realizaciones preferidas, hay menos de 1kb, más preferiblemente menos de 500 nucleótidos (nt), todavía más preferiblemente menos de alrededor de 200, 100, 50, o 30 nt, entre el extremo 3' de la secuencia antirrepresora y el inicio de la secuencia promotora. En ciertas realizaciones preferidas, la secuencia antirrepresora se clona directamente en dirección 5' del promotor, dando como resultado sólo alrededor de 0-20 nt entre el extremo 3' de la secuencia antirrepresora y el inicio de la secuencia promotora.

Para la producción de proteínas multiméricas, pueden usarse dos o más casetes de expresión. Preferiblemente, ambos casetes de expresión son casetes de expresión multicistronicos según la invención, codificando cada uno una proteína marcadora seleccionable diferente, de modo que es posible la selección de ambos casetes de expresión. Esta realización ha demostrado dar buenos resultados, por ejemplo para la expresión de la cadena pesada y ligera de anticuerpos. Resultará evidente que ambos casetes de expresión pueden colocarse en una molécula de ácido nucleico, o los dos pueden estar presentes en una molécula de ácido nucleico separada, antes de introducirlos en células hospedantes. Una ventaja de colocarlos en una molécula de ácido nucleico es que los dos casetes de expresión están presentes a una relación predeterminada única (por ejemplo, 1:1) cuando se introducen en células hospedantes. Por otro lado, cuando están presentes en dos moléculas de ácido nucleico diferentes, esto permite la posibilidad de variar la relación molar de los dos casetes de expresión cuando se introducen en células hospedantes, lo que puede ser una ventaja si la relación molar preferida es diferente de 1:1 o cuando es desconocido previamente cuál es la relación molar preferida, de modo que pueda efectuarse fácilmente por el experto en la técnica la variación de la misma y encontrar empíricamente la óptima. Según la invención, preferiblemente al menos uno de los casetes de expresión, pero más preferiblemente cada uno de ellos, comprende un elemento de control de cromatina, más preferiblemente una secuencia antirrepresora.

En otra realización, las diferentes subunidades o partes de una proteína multimérica están presentes en un solo casete de expresión.

En lugar de o además de la presencia de una secuencia STAR colocada en dirección 5' de un promotor en un casete de expresión, se ha demostrado muy beneficioso proporcionar una secuencia STAR en ambos lados de un casete de expresión, de manera que el casete de expresión que comprende el transgén esté flanqueado por dos secuencias STAR, que en ciertas realizaciones son esencialmente idénticas entre sí.

Se muestra aquí que la combinación del primer elemento antirrepresor en dirección 5' de un promotor y que flanquea el casete de expresión por otras dos secuencias antirrepresoras proporciona resultados superiores.

Ya que al menos algunas secuencias antirrepresoras pueden ser direccionales (documento WO 00/004704), las secuencias antirrepresoras que flanquean el casete de expresión (antirrepresor A y B) se pueden colocar beneficiosamente en dirección opuesta una con respecto a la otra, de manera que el extremo 3' de cada una de estas secuencias antirrepresoras mira hacia el casete de expresión (y entre sí). Por tanto, en realizaciones preferidas, el lado 5' de un elemento antirrepresor encara el ADN/cromatina del cual se ha de disminuir la influencia sobre el transgén por dicho elemento antirrepresor. Para una secuencia antirrepresora en dirección 5' de un promotor en un casete de expresión, el extremo 3' mira hacia el promotor. Las secuencias de los elementos antirrepresores en el listado de secuencias (SEQ. ID. NOs. 1-66) se dan en la dirección 5' a 3', excepto que se indique de otro modo.

En ciertas realizaciones, se proporcionan unidades de transcripción o casetes de expresión según la invención, que comprenden además: a) una secuencia de pausa de la transcripción (TRAP) en dirección 5' del promotor que dirige la transcripción de la unidad de transcripción multicistrónica, estando dicha TRAP en una dirección 5' a 3'; o b) una secuencia TRAP en dirección 3' de dicho marco de lectura abierto del polipéptido de interés y preferiblemente en dirección 3' de la secuencia de terminación de la transcripción de dicha unidad de transcripción multicistrónica, estando dicha TRAP en una dirección 3' a 5'; o c) tanto a) como b); en el que una secuencia TRAP se define funcionalmente como una secuencia que, cuando se coloca en una unidad de transcripción, da como resultado un nivel reducido de transcripción en el ácido nucleico presente en el lado 3' de la TRAP cuando se compara con el nivel de transcripción observado en el ácido nucleico en el lado 5' de la TRAP. Los ejemplos no limitantes de secuencias TRAP son señales de terminación de la transcripción y/o de poliadenilación. En la SEQ. ID. NO. 142 se da un ejemplo no limitante de una secuencia TRAP. Los ejemplos de otras secuencias TRAP, los métodos para encontrarlas, y sus usos se han descrito en el documento WO 2004/055215.

Las moléculas de ADN que comprenden unidades de transcripción multicistrónicas y/o casetes de expresión según la presente invención pueden usarse para mejorar la expresión de ácidos nucleicos, preferiblemente en células hospedantes. Los términos "célula"/"célula hospedante" y "línea celular"/"línea celular hospedante" típicamente se definen respectivamente como una célula y poblaciones homogéneas de la misma que pueden mantenerse en cultivo celular mediante métodos conocidos en la técnica, y que tienen la capacidad de expresar proteínas heterólogas u homólogas.

Las células hospedantes procariotas pueden usarse para propagar y/o realizar ingeniería genética con las moléculas de ADN de la invención, especialmente cuando están presentes en plásmidos que pueden replicarse en células hospedantes procariotas tales como bacterias.

Una célula hospedante según la presente invención es preferiblemente una células eucariota, más preferiblemente una célula de mamífero, tal como una célula de roedor o una célula humana, o una fusión entre diferentes células. En ciertas realizaciones no limitantes, dicha célula hospedante es una célula de osteosarcoma U-2 OS, célula CHO (ovario de hámster chino), célula HEK 293, célula de mieloma HuNS-1, célula de retinoblastoma WERI-Rb-1, célula BHK, célula Vero, célula de mieloma de ratón no secretor Sp2/0-Ag 14, célula de mieloma de ratón no secretor NS0, célula de carcinoma de glándula suprarrenal NCI-H295R o una célula PER.C6®. En ciertas realizaciones de la invención, una célula hospedante es una célula que expresa al menos E1A, y preferiblemente también E1B, de un adenovirus. Como ejemplos no limitantes, tal célula puede derivar de, por ejemplo, células humanas, por ejemplo de un riñón (ejemplo: células HEK 293; véase Graham et al, 1977), pulmón (por ejemplo A549; véase por ejemplo el documento WO 98/39411) o retina (ejemplo: células HER comercializadas con la marca PER.C6®; véase la patente US 5.994.128), o de amniocitos (por ejemplo N52.E6, descritos en la patente US 6.558.948), y similarmente de otras células. Los métodos para obtener tales células se describen por ejemplo en las patentes US 5.994.128 y US 6.558.948. Las células PER.C6, para los fines de la presente invención, significan células procedentes de una pasada en dirección 5' o en dirección 3' de células tal como se depositaron con el nº ECACC 96022940, es decir, que tienen las características de esas células. Previamente se ha mostrado que tales células son capaces de expresar proteínas en niveles elevados (por ejemplo el documento WO 00/63403, y Jones et al, 2003). En otras realizaciones preferidas, las células hospedantes son células CHO, por ejemplo CHO-K1, CHO-S, CHO-DG44, CHO-DUKXB11, y similares. En ciertas realizaciones, dichas células CHO tienen un fenotipo dhfr⁻. Tales células hospedantes eucariotas pueden expresar polipéptidos deseados, y a menudo se usan para ese fin. Pueden obtenerse mediante la introducción de una molécula de ADN de la invención, preferiblemente en forma de un casete de expresión, en las células. Preferiblemente, el casete de expresión se integra en el genoma de las células hospedantes, que puede estar en diferentes posiciones en diversas células hospedantes, y la selección proporcionará un clon en el que el transgén está integrado en una posición adecuada, lo que conduce a un clon de célula hospedante con las propiedades deseadas en términos de niveles de expresión, estabilidad, características de crecimiento, y similares. Alternativamente, la unidad de transcripción multicistrónica puede seleccionarse como diana o seleccionarse al azar para su integración en una región cromosómica que es activa de manera transcripcional, por ejemplo detrás de un promotor presente en el genoma. La selección de células que contienen el ADN de la invención puede efectuarse seleccionando el polipéptido marcador seleccionable, usando métodos habituales conocidos por la persona experta en la técnica. Cuando dicha unidad de transcripción multicistrónica se

integra detrás de un promotor en el genoma, puede generarse in situ un casete de expresión según la invención, es decir, en el genoma de las células hospedantes.

Preferiblemente, las células hospedantes provienen de un clon estable que puede seleccionarse y propagarse según procedimientos convencionales conocidos por el experto en la técnica. Un cultivo de un clon de este tipo es capaz de producir el polipéptido de interés, si las células comprenden la unidad de transcripción multicistrónica de la invención. Las células según la invención son capaces de crecer preferiblemente en cultivo en suspensión en medio libre de suero.

En realizaciones preferidas, la molécula de ADN que comprende la unidad de transcripción multicistrónica de la invención, preferiblemente en forma de un casete de expresión, se integra en el genoma de la célula hospedante eucariota según la invención. Esto proporcionará una herencia estable de la unidad de transcripción multicistrónica.

La selección de la presencia del polipéptido marcador seleccionable, y por tanto de la expresión, puede efectuarse durante la obtención inicial de las células, y podría reducirse o detenerse del todo después de que se han obtenido clones estables. Sin embargo, también es posible aplicar el agente de selección durante etapas posteriores continuamente, o sólo ocasionalmente, posiblemente a menores niveles que durante la selección inicial de las células hospedantes.

Un polipéptido de interés según la invención puede ser cualquier proteína, y puede ser una proteína monomérica o una (parte de una) proteína multimérica. Una proteína multimérica comprende al menos dos cadenas polipeptídicas. Los ejemplos no limitantes de una proteína de interés según la invención son enzimas, hormonas, cadenas de inmunoglobulina, proteínas terapéuticas como proteínas anticancerosas, proteínas de coagulación sanguínea tales como Factor VIII, proteínas multifuncionales tales como eritropoyetina, proteínas de diagnóstico, o proteínas o fragmentos de las mismas útiles con fines de vacunación, todas conocidas por el experto en la técnica.

En ciertas realizaciones, un casete de expresión de la invención codifica una cadena pesada o ligera de inmunoglobulina o una parte de unión a antígeno, derivado y/o análogo de la misma. En una realización preferida, se proporciona una unidad de expresión de proteína según la invención, en la que dicha proteína de interés es una cadena pesada de inmunoglobulina. En aún otra realización preferida, se proporciona una unidad de expresión de proteína según la invención, en la que dicha proteína de interés es una cadena ligera de inmunoglobulina. Cuando estas dos unidades de expresión de proteína están presentes en la misma célula (hospedante), se ensambla una proteína multimérica, y más específicamente una inmunoglobulina. Por tanto, en ciertas realizaciones, la proteína de interés es una inmunoglobulina, tal como un anticuerpo, que es una proteína multimérica. Preferiblemente, dicho anticuerpo es un anticuerpo humano o humanizado. En ciertas realizaciones del mismo, es un anticuerpo de IgG, IgA o IgM. Una inmunoglobulina puede estar codificada por las cadenas pesada y ligera en casetes de expresión diferentes, o en un único casete de expresión. Preferiblemente, la cadena pesada y ligera están presentes cada una en un casete de expresión separado, teniendo cada uno su propio promotor (que puede ser el mismo o diferente para los dos casetes de expresión), comprendiendo cada uno una unidad de transcripción multicistrónica según la invención, siendo la cadena pesada y ligera el polipéptido de interés, y codificando preferiblemente cada uno una proteína marcador seleccionable diferente, de modo que puede efectuarse la selección del casete de expresión de ambas cadenas pesada y ligera cuando los casetes de expresión se introducen y/o están presentes en una célula hospedante eucariota.

El polipéptido de interés puede ser de cualquier fuente, y en ciertas realizaciones es una proteína de mamífero, una proteína artificial (por ejemplo, una proteína de fusión o proteína mutada), y preferiblemente es una proteína humana.

Obviamente, las configuraciones de los casetes de expresión de la presente invención pueden usarse también cuando el objetivo último no es la producción de un polipéptido de interés, sino el ARN mismo, por ejemplo para la producción de cantidades aumentadas de ARN a partir de un casete de expresión, que puede usarse con fines de regular otros genes (por ejemplo, ARNi, ARN anticodificante), terapia génica, producción de proteína in vivo, etc.

En un aspecto, la invención proporciona un método para generar una célula hospedante que expresa un polipéptido de interés, comprendiendo el método las etapas de: a) introducir en una pluralidad de células precursoras un casete de expresión según la invención, y b) cultivar las células generadas en condiciones que seleccionan la expresión del polipéptido marcador seleccionable, y c) seleccionar al menos una célula hospedante que produce el polipéptido de interés. Este nuevo método proporciona un resultado muy bueno en términos de la relación de clones obtenidos frente a clones con expresión elevada del polipéptido deseado. Usando las condiciones más rigurosas, es decir, la eficiencia de la traducción más débil para el polipéptido marcador seleccionable (usando la secuencia de inicio de la traducción más débil), se obtienen muchas menos colonias usando la misma concentración de agente de selección que con sistemas de selección conocidos, y un porcentaje relativamente elevado de los clones obtenidos produce el polipéptido de interés en niveles elevados. Además, los niveles obtenidos de expresión parecen mayores que los obtenidos cuando se usa un número incluso mayor de clones usando los sistemas de selección conocidos.

Es una ventaja adicional que el sistema de selección sea veloz, debido a que no requiere la amplificación del número de copias del transgén. Por tanto, las células con números de copia bajos de las unidades de transcripción

multicistricas ya proporcionan niveles de expresi3n elevados. N3meros elevados de copias del transg3n pueden tender a la inestabilidad gen3tica y al silenciamiento inducido por la repetic3n (por ejemplo Kim et al, 1998; McBurney et al, 2002). Por lo tanto, una ventaja adicional de las realizaciones de la invenci3n con n3meros relativamente bajos de copias del transg3n es que se anticipa que menores n3meros de copias tienen menos tendencia a la recombinaci3n y al silenciamiento inducido por repetic3n, y por lo tanto se anticipan menos problemas a este respecto cuando se usan c3lulas hospedantes con un n3mero limitado de copias del transg3n en comparaci3n con las c3lulas hospedantes obtenidas usando un sistema de amplificaci3n en el que puede haber en el genoma de la c3lula cientos o incluso miles de copias de las secuencias codificantes del marcador seleccionable y de la prote3na de inter3s. La presente invenci3n proporciona ejemplos de niveles de expresi3n elevados, que usan el sistema de selecci3n de unidad de transcripci3n multicistricas, mientras que el n3mero de copias del transg3n es relativamente bajo, es decir, menor que 30 copias por c3lula, o incluso menor que 20 copias por c3lula. Por tanto, la presente invenci3n permite la generaci3n de c3lulas hospedantes seg3n la invenci3n que comprenden menos de 30 copias de la unidad de transcripci3n multicistricas en el genoma de las c3lulas hospedantes, preferiblemente menos de 25, m3s preferiblemente menos de 20 copias, mientras que al mismo tiempo proporcionan suficientes niveles de expresi3n del polip3ptido de inter3s para fines comerciales, por ejemplo m3s de 15, preferiblemente m3s de 20 pg/c3lula/d3a de un anticuerpo.

Aunque pueden obtenerse clones que tienen n3meros de copias relativamente bajos de las unidades de transcripci3n multicistricas y niveles elevados de expresi3n, el sistema de selecci3n de la invenci3n puede combinarse no obstante con m3todos de amplificaci3n para mejorar a3n m3s los niveles de expresi3n. Esto puede lograrse, por ejemplo, mediante la amplificaci3n de un gen dhfr cointegrado usando metotrexato, por ejemplo colocando dhfr en la misma mol3cula de 3cido nucleico que la unidad de transcripci3n multicistricas de la invenci3n, o mediante cotransfecci3n cuando dhfr est3 en una mol3cula de ADN separada.

En un aspecto, la invenci3n proporciona un m3todo para producir un polip3ptido de inter3s, comprendiendo el m3todo cultivar una c3lula hospedante, comprendiendo dicha c3lula hospedante una mol3cula de ADN que comprende una unidad de expresi3n multicistricas o un casete de expresi3n seg3n la invenci3n, y expresar el polip3ptido de inter3s a partir de la secuencia codificante para el polip3ptido de inter3s.

La c3lula hospedante para este aspecto es una c3lula hospedante eucariota, preferiblemente una c3lula de mam3fero, tal como una c3lula CHO, adicionalmente como se describe anteriormente.

La introducci3n de 3cido nucleico que se va a expresar en una c3lula puede hacerse mediante uno de varios m3todos, que son conocidos como tales por el experto en la t3cnica, dependiendo tambi3n del formato del 3cido nucleico a introducir. Dichos m3todos incluyen, pero no se limitan a, transfecci3n, infecci3n, inyecci3n, transformaci3n, y similares. Las c3lulas hospedantes adecuadas que expresan el polip3ptido de inter3s pueden obtenerse mediante selecci3n como se describe anteriormente.

En ciertas realizaciones, el agente de selecci3n est3 presente en el medio de cultivo al menos parte del tiempo durante el cultivo, ya sea en concentraciones suficientes para seleccionar c3lulas que expresan el polip3ptido marcador seleccionable o en concentraciones menores. En realizaciones preferidas, el agente de selecci3n ya no est3 presente en el medio de cultivo durante la fase de producci3n cuando se expresa el polip3ptido.

El cultivo de una c3lula se realiza para posibilitar que metabolice y/o crezca y/o se divida y/o produzca prote3nas recombinantes de inter3s. Esto puede lograrse mediante m3todos bien conocidos por los expertos en la t3cnica, e incluye, pero sin limitarse a, proporcionar nutrientes para la c3lula. Los m3todos comprenden crecimiento adherente a superficies, crecimiento en suspensi3n, o combinaciones de los mismos. El cultivo puede realizarse por ejemplo en platos, matraces giratorios o en biorreactores, usando sistemas discontinuos, semicontinuos, continuos tales como sistemas de perfusi3n y similares. Para conseguir la producci3n a gran escala (continua) de prote3nas recombinantes mediante el cultivo celular, se prefiere en la t3cnica tener c3lulas capaces de crecer en suspensi3n, y se prefiere tener c3lulas aptas para ser cultivadas en ausencia del suero derivado de animal o ser humano o de componentes s3ricos derivados de animal o ser humano.

Las condiciones para hacer crecer o multiplicar c3lulas (v3ase, por ejemplo, Tissue Culture, Academic Press, Kruse and Paterson, editores (1973)) y las condiciones para la expresi3n del producto recombinante son conocidas para el experto en la t3cnica. En general, los principios, protocolos y t3cnicas pr3cticas para maximizar la productividad de cultivos de c3lulas de mam3fero pueden encontrarse en Mammalian Cell Biotechnology: a Practical Approach (M. Butler, ed., IRL Press, 1991).

En una realizaci3n preferida, la prote3na expresada se recoge (aisla) a partir de c3lulas o del medio de cultivo, o de ambos. Entonces puede purificarse adicionalmente usando m3todos conocidos, por ejemplo filtraci3n, cromatograf3a en columna, etc., por m3todos generalmente conocidos por el experto en la t3cnica.

El m3todo de selecci3n seg3n la presente invenci3n funciona en ausencia de elementos de control de cromatina, pero se obtienen resultados mejorados cuando las unidades de expresi3n multicistricas se proporcionan con dichos elementos. El m3todo de selecci3n seg3n la presente invenci3n funciona particularmente bien cuando se usa un casete de expresi3n seg3n la invenci3n, que comprende al menos una secuencia antirrepresora. Dependiendo

del agente y condiciones de selección, la selección puede hacerse en ciertos casos tan rigurosa que sólo unas pocas o incluso ninguna célula hospedante sobrevive a la selección, a menos que estén presentes secuencias antirrepresoras. Por tanto, la combinación del nuevo método de selección y las secuencias antirrepresoras proporciona un método muy atractivo para obtener sólo números limitados de colonias con una oportunidad enormemente mejorada de expresión elevada del polipéptido de interés en las mismas, mientras que al mismo tiempo los clones obtenidos que comprenden los casetes de expresión con secuencias antirrepresoras proporcionan la expresión estable del polipéptido de interés, es decir, tienden menos al silenciamiento u otros mecanismos de reducción de la expresión que los casetes de expresión convencionales.

En ciertas realizaciones, casi no se obtienen clones cuando está presente una secuencia no antirrepresora en el casete de expresión según la invención, que proporciona una selección muy rigurosa. El nuevo sistema de selección descrito aquí también proporciona por lo tanto la posibilidad de ensayar partes de elementos antirrepresores en busca de funcionalidad, analizando los efectos de tales secuencias cuando están presentes en casetes de expresión de la invención en condiciones de selección. Este cribado fácil, que proporciona en muchos casos una diferencia de blanco y negro casi completa o incluso completa, puede contribuir por lo tanto a identificar partes funcionales o derivados procedentes de secuencias antirrepresoras. Cuando se ensayan secuencias antirrepresoras conocidas, este ensayo se puede usar para caracterizarlas adicionalmente. Cuando se ensayan fragmentos de secuencias antirrepresoras conocidas, el ensayo proporcionará fragmentos funcionales de tales secuencias antirrepresoras conocidas.

La práctica de esta invención empleará, a menos que se indique de otro modo, técnicas convencionales de inmunología, biología molecular, microbiología, biología celular y ADN recombinante, que están dentro de las habilidades de la técnica. Véanse, por ejemplo, Sambrook, Fritsch y Maniatis, *Molecular Cloning: A Laboratory Manual*, 2ª edición, 1989; *Current Protocols in Molecular Biology*, Ausubel FM, et al., eds., 1987; la serie *Methods in Enzymology* (Academic Press, Inc.); *PCR2: A Practical Approach*, MacPherson MJ, Hams BD, Taylor GR, eds., 1995; *Antibodies: A Laboratory Manual*, Harlow and Lane, ed., 1988.

La invención se explica adicionalmente en los siguientes ejemplos. Los ejemplos no limitan la invención en modo alguno. Sirven simplemente para aclarar la invención.

Ejemplos

Los Ejemplos 1-7 describen el ensayo de STAR67 y otras secuencias STAR en diversas configuraciones y en diversas circunstancias. Los Ejemplos 8 y posteriores describen el nuevo sistema de selección de la invención.

Ejemplo 1: Construcción de vectores STAR 67

Se aisló una nueva secuencia antirrepresora (STAR) usando un cribado genético como se describe en el documento WO 03/004704, y esta nueva secuencia se denominó STAR67. Se ensayaron los efectos de STAR67 sobre la expresión de transgenes en líneas de células de mamífero. Se describe aquí la construcción de los diversos constructos.

35 Materiales y métodos

Se crearon tres plásmidos (Fig.1):

- A) CMV-d2EGFP-ires-Zeo (Control de CMV),
- B) STAR67-CMV-d2EGFP-ires-Zeo (CMV-STAR67),
- C) STAR7-STAR67-CMV-d2EGFP-ires-Zeo-STAR7 (CMV-STAR67 7/7)

A continuación se describe la construcción del constructo A. El plásmido pd2EGFP (Clontech 6010-1) se modificó mediante inserción de un ligador en el sitio *Bs*WI para producir pd2EGFP-link. El ligador, obtenido mediante la hibridación de oligonucleótidos GTACGGATATCAGATCTTTAATTAAG (SEQ. ID. NO. 67) y GTACCTTAATTAAGATCTGATAT (SEQ. ID. NO. 68), introdujo sitios para las endonucleasas de restricción *Pac*I, *Bgl*II, y *Eco*RV. Esto creó el sitio de clonación múltiple MCSII para la inserción de elementos STAR. Después, los cebadores GATCAGATCTGGCGGCCATTTAAATCGTCTCGCGGTTTCGGTGATGACG G (SEQ. ID. NO. 69) y (AGGCGGATCCGAATGTATTAGAAAAATAACAAATAGGGG (SEQ. ID. NO. 70) se usaron para amplificar una región de 0,37 kb procedente de pd2EGFP, que se insertó en el sitio *Bgl*II de pIRES (Clontech 6028-1) para producir pIRES-stuf. Esto introdujo sitios para las endonucleasas de restricción *Asc*I y *Swa*I en MCSI, y actúa como un "fragmento relleno" para evitar la interferencia potencial entre elementos STAR y promotores adyacentes. pIRES-stuf se dirigió con *Bgl*II y *Fsp*I para liberar un fragmento de ADN compuesto del fragmento relleno, el promotor de CMV, el elemento IRES (flanqueado por sitios de clonación múltiples MCS A y MCS B), y la señal de poliadenilación de SV40. Este fragmento se ligó con el vector de cadena principal de pd2EGFP-link, producido mediante digestión con *Bam*HI y *Stu*I, para producir pIRES-link.

El marco de lectura abierto del gen de resistencia a zeocina se insertó en los sitios *Bam*HI/*Not*I en dirección 3' del pIRES según lo siguiente: el ORF de resistencia a zeocina se amplificó mediante PCR con los cebadores GATCGGATCCTTCGAAATGGCCAAGTTGACCAGTGC (SEQ. ID. NO. 71) y

AGGCGCGCCGCAATTCTCAGTCCTGCTCCTC (SEQ. ID. NO. 72) del plásmido pCMV/zeo (Invitrogen, número de catálogo V50120), se dirigió con *Bam*HI y *Not*I, y se ligó con pIRES-link digerido con *Bam*HI/*Not*I para producir pIRES-link-zeo. El ORF informador de d2EGFP se introdujo en pIRES-link-zeo mediante amplificación de pd2EGFP (Clontech 6010-1) con los cebadores GATCGAATTCTCGCGAATGGTGAGCAAGCAGATCCTGAAG (SEQ. ID. NO. 73) y AGGCGAATTCACCGGTGTTAAACTTACACCCACTCGTGCAGGCTGCCCA GG (SEQ. ID. NO. 74), y la inserción del casete de d2EGFP digerido con *Eco*RI en el sitio *Eco*RI en el plásmido pIRES-link-zeo. Esto creó el constructo A, CMV-d2EGFP-IRES-Zeo (Control de CMV).

STAR 67 se clonó en dirección 5' del promotor de CMV, en el sitio *Asc*I (quedando alrededor de 15 nt entre STAR67 y el promotor). Esto creó el constructo B, STAR67-CMV-d2EGFP-ires-Zeo (CMV-STAR67).

STAR67 también se ensayó en el contexto de un casete que contiene también STAR7, clonado direccionalmente en los sitios 5' *Sal*I y *Xba*I y los sitios 3' *Bgl*II y *Pac*I para flanquear el casete completo con STAR7. Este es el constructo C (CMV-STAR67 7/7).

Ejemplo 2: STAR67 potencia el nivel de expresión de los promotores de CMV, EFl α y UB6 en células CHO transfectadas establemente.

Se ensayó si la presencia de STAR67 adyacente a los promotores de CMV, EFl α y UB6 influye en el nivel de expresión de estos promotores en células CHO. Para este fin se usan los constructos A y B (Fig. 1) descritos en el Ejemplo 1, modificados para los promotores respectivos:

- 1 CMV-d2EGFP-ires-Zeo (Control de CMV)
- 2 STAR67-CMV-d2EGFP-ires-Zeo (CMV-STAR67)
- 3 EFl α -d2EGFP-ires-Zeo (Control de EFl α)
- 4 STAR67-EFl α -d2EGFP-ires-Zeo (EFl α -STAR67)
- 5 UB6-d2EGFP-ires-Zeo (Control de UB6)
- 6 STAR67-UB6-d2EGFP-ires-Zeo (UB6-STAR67).

Materiales y métodos

Los promotores de UB6 y EFl α se intercambiaron por el promotor de CMV en los plásmidos descritos en la Fig. 1. El promotor de UB6 se clonó según lo siguiente. Se amplificó mediante PCR un rellenedor de 0,37 kb de ADN procedente del plásmido pd2EGFP como se describe en el ejemplo 1, usando cebadores identificados mediante SEQ. ID. NOs. 69 y 70. El rellenedor de ADN resultante se clonó en el sitio *Bgl*II de pUB6/V5-His [Invitrogen V250-20], creando pUB6-stuf. A partir de pUB6-stuf, se clonó un fragmento *Asc*I-*Sac*I en CMV-d2EGFP-IRES-Zeo, a partir del que se eliminó el promotor de CMV.

El promotor de EFl α se amplificó mediante PCR con pEFl α /V5-His [Invitrogen V920-20] como molde, usando los cebadores GATCGGCGCGCCATTTAAATCCGAAAAGTGCCACCTGACG (SEQ. ID. NO. 79) y AGGCGGGACCCCTCACGACACCTGAAATGGAAG (SEQ. ID. NO. 80). El fragmento de PCR se clonó en los sitios *Asc*I y *Ppu*MI de CMV-d2EGFP-IRES-Zeo, a partir del que se eliminó el promotor de CMV.

La línea celular de ovario de hámster chino CHO-K1 (ATCC CCL-61) se cultivó en medio HAMS-F12 + 10% de suero fetal de ternera que contiene 2 mM de glutamina, 100 U/ml de penicilina, y 100 microgramos/ml de estreptomina a 37°C/5% de CO₂. Las células se transfectaron con los plásmidos usando Lipofectamine 2000 (Invitrogen) como se describe por el fabricante. De forma breve, las células se sembraron en vasos de cultivo y se hicieron crecer toda la noche hasta 70-90% de confluencia. El reactivo Lipofectamine se combinó con ADN plasmídico a una relación de 6 microlitros por microgramo (por ejemplo, para una cápsula de Petri de 10 cm, 20 microgramos de ADN y 120 microlitros de Lipofectamine), y se añadieron a las células. Después de la incubación durante toda la noche, la mezcla de transfección se substituyó por medio reciente, y las células transfectadas se incubaron adicionalmente. Después del cultivo durante toda la noche, las células se tripsinizaron y se sembraron en vasijas de cultivo recientes con medio reciente. Después de otra incubación durante toda la noche, se añadió zeocina hasta una concentración de 50 μ g/ml, y las células se cultivaron adicionalmente. Después de otros tres días, el medio se substituyó por medio reciente que contiene zeocina (100 μ g/ml) y se cultivó adicionalmente. Cuando las colonias individuales se hicieron visibles (aproximadamente diez días después de la transfección), el medio se retiró y se substituyó por medio reciente sin zeocina. Los clones individuales se aislaron y se transfirieron a placas de 24 pocillos en medio sin zeocina. Un día después del aislamiento de las colonias, se añadió zeocina al medio. La expresión del gen informador d2EGFP se evaluó aproximadamente tres semanas después de la transfección. Los niveles de expresión de d2EGFP en las colonias se midieron después de períodos de dos semanas. Después de las dos semanas iniciales después de la transfección cuando se llevaron a cabo las primeras medidas de d2EGFP, las colonias se cultivaron en medio sin zeocina u otros antibióticos. Esto continuó durante el resto del experimento.

Resultados

La Fig. 2 muestra que la transfección del constructo que contiene STAR67 clonada en dirección 5' del promotor de CMV dio como resultado un número de colonias de CHO que expresan niveles significativamente mayores de

proteína d2EGFP, en comparación con el control “vacío” sin STAR67, Control de CMV. La media de la señal de d2EGFP en las 10 colonias transfectadas con el plásmido de Control de CMV fue 34, cuando se mide 25 días después de la transfección. 60 días después de la transfección, la media de la señal de d2EGFP en estas 10 colonias se redujo a 13, indicando que la expresión no es estable a lo largo del tiempo. En comparación, la media de la señal de d2EGFP en las 10 colonias transfectadas con el plásmido STAR67-CMV fue 42 cuando se mide después de 25 días, y 32 medida 60 días después de la transfección. Por tanto, 60 días después de la transfección, un constructo de CMV que engloba STAR67 comunicó un factor 2,5 mayor de nivel de expresión conducido por el promotor de CMV de la proteína informadora en clones establemente transfectados. De forma importante, 25 días después de la transfección, tras la primera medida, la presión de selección se retiró cultivando las colonias en medio sin zeocina. Por tanto, las colonias que contienen el constructo STAR67 son más estables a lo largo del tiempo en ausencia de presión de selección que las colonias que no contienen un constructo STAR67.

La Fig. 3 muestra que la transfección del constructo que contiene STAR67 clonado en dirección 5' del promotor de EFl α dio como resultado un número de colonias de CHO que expresan niveles significativamente mayores de proteína d2EGFP, en comparación con el control “vacío” sin STAR67, Control de EFl α . La media de la señal de d2EGFP en las 10 colonias transfectadas con el plásmido de Control de EFl α fue 31 cuando se mide 25 días después de la transfección. 60 días después de la transfección, la media de la señal de d2EGFP en estas 10 colonias fue 26. En comparación, la media de la señal de d2EGFP en las 10 colonias transfectadas con el plásmido EFl α -STAR67 fue 60 cuando se midió después de 25 días, y 76 cuando se midió 60 días después de la transfección. Por tanto, tanto después de 25 como de 60 días después de la transfección, un constructo de EFl α que engloba STAR67 comunicó un factor 2,9 mayor de nivel de expresión conducida por el promotor de EFl α de la proteína informadora en clones establemente transfectados.

La Fig. 4 muestra que la transfección del constructo que contiene STAR67 clonado en dirección 5' del promotor de UB6 dio como resultado un número de colonias de CHO que expresan niveles significativamente mayores de proteína d2EGFP, en comparación con el control “vacío” sin STAR67, Control de UB6. La media de la señal de d2EGFP en las 10 colonias transfectadas con el plásmido de control de UB6 fue 51 cuando se mide 25 días después de la transfección. 60 días después de la transfección, la media de la señal de d2EGFP en estas 10 colonias fue 29, indicando que la expresión no fue estable a lo largo del tiempo. En comparación, la media de la señal de d2EGFP en las 10 colonias transfectadas con el plásmido UB6-STAR67 fue 218 cuando se mide después de 25 días, y 224 cuando se mide 60 días después de la transfección. Por tanto, 25 días después de la transfección, un constructo de UB6 que engloba STAR67 comunicó un factor 4,3 mayor de nivel de expresión conducida por el promotor de UB6 de la proteína informadora en clones establemente transfectados. Después de 60 días, este factor fue 7,7, debido a la inestabilidad de la expresión en las colonias del control y a la estabilidad en las colonias de UB6-STAR67. De forma importante, 25 días después de la transfección, tras la primera medida, se retiró la presión de selección cultivando las colonias en medio sin zeocina. Por tanto, las colonias que contienen el constructo de STAR67 son más estables a lo largo del tiempo en ausencia de presión de selección de las colonias que no contienen un constructo de STAR67.

En conclusión, STAR67 incrementa la expresión de tres promotores diferentes, no relacionados.

Ejemplo 3: STAR67 potencia el nivel de expresión a partir de promotores de CMV, EFl α y UB6 en células PER.C6 establemente transfectadas.

Se ensayó si la presencia de STAR67 adyacente a los promotores de CMV, EFl α y UB6 influye en el nivel de expresión de estos promotores en otro tipo celular distinto de células CHO, a saber, células PER.C6 humanas. Se usaron los mismos constructos como en el ejemplo 1.

Materiales y métodos

Transfección, cultivo y análisis de células PER.C6

Se cultivaron células PER.C6® en medio DMEM + piridoxina + 9% de suero fetal bovino (no inactivado con calor), 8,9 mM de MgCl₂, 100 U/ml de penicilina, y 100 microgramos/ml de estreptomycin a 37°C/10% de CO₂. Las células se transfectaron con los plásmidos usando Lipofectamine 2000 (Invitrogen) como se describe por el fabricante. De forma breve, las células se sembraron en placas de 6 pocillos y se hicieron crecer toda la noche hasta 70-90% de confluencia. El reactivo de Lipofectamine se combinó con ADN plasmídico a una relación de 15 microlitros por 3 microgramos (por ejemplo para una cápsula de Petri de 10 cm, 20 microgramos de ADN y 120 microlitros de Lipofectamine) y se añadió después de 30 minutos de incubación a 25°C a las células. Después de la incubación durante 6 horas, la mezcla de transfección se sustituyó por medio reciente, y las células transfectadas se incubaron adicionalmente. Después del cultivo durante toda la noche, las células se tripsinizaron y se sembraron (diluciones 1:15, 1:30, 1:60, 1:120) en cápsulas de Petri recientes (90 mm) con medio reciente con zeocina añadida hasta una concentración de 100 μ g/ml, y las células se cultivaron adicionalmente. Cuando las colonias se hicieron visibles, los clones individuales se aislaron raspando y se transfirieron a placas de 24 pocillos en medio con zeocina. Cuando se hicieron crecer hasta ~70% de confluencia, las células se transfirieron a placas de 6 pocillos. Las colonias estables se expandieron durante dos semanas en placas de 6 pocillos antes de que se determinase la señal de d2EGFP en un citómetro de flujo XL-MCL Beckman Coulter. La media de la señal de d2EGFP se tomó como medida para el

nivel de expresión de d2EGFP. Las colonias se midieron durante una segunda vez después de dos semanas. Después, las colonias se cultivaron adicionalmente en ausencia de zeocina.

Resultados

5 La Fig. 5 muestra que la transfección del constructo que contiene STAR67 clonada en dirección 5' del promotor de CMV dio como resultado un número de colonias de PER.C6 que expresan niveles significativamente mayores de proteína d2EGFP, en comparación con el control "vacío" sin STAR67, Control de CMV. La media de la señal de d2EGFP en las 10 colonias transfectadas con el plásmido de Control de CMV fue 37 cuando se midió 30 días después de la transfección. 60 días después de la transfección, la media de la señal de d2EGFP en estas 10 colonias se redujo a 14, indicando que la expresión no fue estable a lo largo del tiempo. En comparación, la media de la señal de d2EGFP en las 10 colonias transfectadas con el plásmido STAR67-CMV fue 101 cuando se midió después de 30 días, y 45 cuando se midió 60 días después de la transfección. Por tanto, 60 días después de la transfección, un constructo de CMV que engloba STAR67 comunicó un factor de 3,2 mayor de nivel de expresión conducida por el promotor de CMV de la proteína informadora en clones establemente transfectados.

15 La Fig. 6 muestra que la transfección del constructo que contiene STAR67 clonada en dirección 5' del promotor de EF1 α dio como resultado un número de colonias de PER.C6 que expresan niveles significativamente mayores de proteína d2EGFP, en comparación con el control "vacío" sin STAR67, Control de EF1 α . La media de la señal de d2EGFP en las 10 colonias transfectadas con el plásmido de Control de EF1 α fue 5 cuando se midió 30 días después de la transfección. 60 días después de la transfección, la media de la señal de d2EGFP en estas 10 colonias fue 6. En comparación, la media de la señal de d2EGFP en las 10 colonias transfectadas con el plásmido EF1 α -STAR67 fue 25 cuando se midió después de 30 días, y 20 cuando se midió 60 días después de la transfección. Por tanto, tanto después de 30 días como de 60 días después de la transfección, un constructo de EF1 α que engloba STAR67 comunicó un factor de 4 mayor de nivel de expresión conducida por el promotor de EF1 α de la proteína informadora en clones establemente transfectados.

25 La Fig. 7 muestra que la transfección del constructo que contiene STAR67 clonada en dirección 5' del promotor de UB6 dio como resultado un número de colonias de PER.C6 que expresan niveles significativamente mayores de proteína d2EGFP, en comparación con el control "vacío" sin STAR67, Control de UB6. La media de la señal de d2EGFP en las 10 colonias transfectadas con el plásmido de Control de UB6 fue 4 cuando se midió 30 días después de la transfección. 60 días después de la transfección, la media de la señal de d2EGFP en estas 10 colonias fue 2. En comparación, la media de la señal de d2EGFP en las 10 colonias transfectadas con el plásmido UB6-STAR67 fue 27 cuando se midió después de 30 días, y 18 cuando se midió 60 días después de la transfección. Por tanto, tanto después de 30 días como de 60 días después de la transfección, un constructo de UB6 que engloba STAR67 comunicó un factor 7 a 9 veces mayor de nivel de expresión conducida por el promotor de UB6 de la proteína informadora en clones establemente transfectados.

35 Por tanto, la colocación de STAR67 en dirección 5' del promotor dio como resultado niveles de expresión proteica significativamente mayores en comparación con constructos sin STAR67, también en células PER.C6. Por tanto, STAR 67 funciona en tipos celulares diferentes, no relacionados.

Ejemplo 4: Nueva configuración de STAR67 combinada con otros elementos de STAR para potenciar el promotor de SV40 en células CHO.

40 Se ensayó si la presencia de STAR67 adyacente al promotor de SV40 influye en el nivel de expresión de este promotor, ya sea sola o en combinación con otro elemento STAR, en este ejemplo STAR7. Los constructos que se usaron para este fin (véase la Fig. 8) son:

- 1 SV40-d2EGFP-ires-Zeo (Control de SV40)
- 2 STAR67-SV40-d2EGFP-ires-Zeo (SV40-STAR67)
- 3 STAR7-SV40-d2EGFP-ires-Zeo-STAR7 (SV40-STAR7/7)
- 45 4 STAR7-STAR67-SV40-d2EGFP-ires-Zeo-STAR7 (SV40-STAR67 7/7)

Materiales y métodos

50 El promotor de SV40 se amplificó mediante PCR con pIRES como molde usando los cebadores TTGGTTGGGGCGCGCCGCAGCACCATGGCCTGAAATAACCTCTGAAAGAG G (SEQ. ID. NO. 81) y TTGGTTGGGAGCTCAAGCTTTTTGCAAAGCCTAGGCCTCCAAAAAGCC TCCTC (SEQ. ID. NO. 82).

El fragmento de PCR se clonó en los sitios Ascl y SacI de CMV-d2EGFP-IRES-Zeo, a partir del cual se retiró el promotor de CMV.

Las células CHO se transfectaron, las colonias se aislaron y se propagaron, y se analizaron como en el Ejemplo 2.

55 Resultados

La Fig. 8 muestra que la transfección del constructo que contiene STAR67 clonada en dirección 5' del promotor de SV40 (SV40-STAR67), o STAR7 clonada para flanquear todo el constructo (SV40-STAR 7/7), no dio como resultado colonias de CHO que expresan niveles significativamente mayores de la proteína d2EGFP, en comparación con el control "vacío" sin los STARs (Control de SV40). La media de la señal de d2EGFP en las 18 colonias transfectadas con el plásmido de Control de SV40 fue 86, cuando se midió 40 días después de la transfección. En comparación, la media de la señal de d2EGFP en las 18 colonias transfectadas con el plásmido SV40-STAR67 fue 82 cuando se midió después de 40 días, y la media de la señal de d2EGFP en las 18 colonias transfectadas con el plásmido SV40-STAR 7/7 fue 91 cuando se midió después de 40 días. Por tanto, no se observó ningún efecto significativo de estos elementos de STAR sobre el promotor de SV40 en células CHO.

Parece que los niveles de expresión a partir del promotor de SV40 en estas células ya son bastante elevados, e incluso mucho mayores que los observados con el promotor de CMV, que se considera que es un promotor muy potente. Esta expresión de fondo elevada en ausencia de un elemento STAR usando el promotor de SV40 en células CHO puede explicar por qué no se observó ningún efecto significativo de STAR67 sola, o STAR7 que flanquea el transgén.

Sin embargo, la media de la señal de d2EGFP en las 18 colonias transfectadas con el plásmido SV40-STAR67 7/7 fue 209 cuando se midió después de 40 días, que es un factor 2,4 veces mayor que la media de las 18 colonias de control (86). Por tanto, cuando se usa el elemento STAR67 en combinación con otro elemento STAR, esto da como resultado un número de colonias de CHO establemente transfectadas que muestran niveles de expresión de d2EGFP significativamente mayores.

Por lo tanto, en esta nueva configuración (5'- secuencia STAR A - secuencia STAR C - promotor - ácido nucleico que codifica una proteína de interés - secuencia STAR B - 3', en la que, en el presente ejemplo, las secuencias STAR A y B son STAR7, y la secuencia STAR C es STAR67), parece que los elementos STAR funcionan incluso mejor que en las configuraciones descritas hasta ahora.

Se han realizado experimentos en los que los elementos STAR7 flanqueantes se sustituyeron flanqueando elementos STAR6 o flanqueando elementos STAR4, en combinación con STAR67 en dirección 5' del promotor de SV40 (SV40-STAR67 6/6 y SV40-STAR67 4/4, respectivamente, usando la misma nomenclatura como antes), y se observó expresión mejorada también con estas combinaciones. Esto demuestra que los elementos STAR7 flanqueantes pueden de hecho intercambiarse por otras secuencias STAR, y todavía se observa la mejora de la nueva configuración del casete de expresión con las secuencias STAR.

Ejemplo 5: Una combinación de STAR67 y STAR7 potencia los niveles de expresión de anticuerpo conducida por UB6 en células CHO establemente transfectadas.

En el ejemplo 4 se demostró que la combinación de STAR67 y STAR7 potenció los niveles de expresión de la proteína d2EGFP en células CHO. Aquí se ensayó si la combinación de STAR67 y STAR7 se podría usar para la producción de un anticuerpo. Se escogió un anticuerpo frente a la molécula EpCAM (Huls et al, 1999) como proteína de ensayo, y se usó el promotor de UB6.

Materiales y métodos

Plásmidos

El ADNc de cadena pesada (HC-EpCAM) se clonó en un constructo que engloba el promotor de UB6. El HC-EpCAM se acopló al gen de resistencia a zeocina mediante una secuencia IRES. El ADNc de cadena ligera (LC-EpCAM) también se clonó en un constructo que engloba el promotor de UB6. El LC-EpCAM se acopló al gen de resistencia a puromicina mediante una secuencia IRES. Juntos, estos dos plásmidos representan el Control de UB6 (HC+LC) (Fig. 9).

Para ensayar los efectos de STAR67 y STAR7, se clonó STAR67 en ambos constructos HC+LC, en dirección 5' de los promotores de UB6. STAR7 se clonó para flanquear todos los casetes, tanto en el extremo 5' como 3' (Fig. 9). Estos dos plásmidos representan STAR7-STAR67-UB6 (HC+LC) STAR7.

Transfección y cultivo de células CHO

La línea celular de ovario de hámster chino CHO-K1 (ATCC CCL-61) se transfectó y cultivo como en el ejemplo 2, usando zeocina (100 µg/ml) y puromicina (2,5 µg/ml) para la selección. Un día después de la transfección, se añadió zeocina al medio de cultivo. Cuando las primeras colonias se hicieron visible (aproximadamente 7 días después de la adición de zeocina), se retiró el medio de cultivo y se sustituyó por medio de cultivo que contiene puromicina. Después de aproximadamente 7 días, las colonias se aislaron y se transfirieron a placas de 24 pocillos, en medio de cultivo que contiene zeocina solamente.

Resultados

La Fig. 9 muestra que la transfección de constructos de anticuerpo que contienen STAR67 clonada en dirección 5' del promotor de UB6 y dos elementos STAR7 clonados para flanquear los casetes completos dio como resultado un número de colonias de CHO que expresan niveles significativamente mayores de anticuerpo EpCAM (medido mediante ELISA usando un anticuerpo anti-IgG humana) en comparación con el control "vacío" sin STAR67 y STAR7, Control de UB6 (HC+LC). La media de la producción de EpCAM en las 18 colonias transfectadas con el plásmido de Control de UB6 (HC+LC) fue 2,7 pg/célula/día, cuando se mide 25 días después de la transfección. Los agentes de selección zeocina y puromicina se retiraron después de 25 días. 45 días después de la transfección, la media de la producción de EpCAM en estas 18 colonias fue 2,7 pg/célula/día. En comparación, la media de la señal de producción de EpCAM en las 18 colonias transfectadas con el plásmido STAR7-STAR67-UB6 (HC+LC)-STAR7 fue 6,7 cuando se midió después de 25 días, y 7,7 pg/célula/día cuando se midió 45 días después de la transfección. Por tanto, tanto después de 30 como de 45 días después de la transfección, un constructo de UB6 que engloba STAR67/STAR7 comunicó un factor de 2,5 a 2,9 veces mayor de nivel de expresión de EpCAM conducida por el promotor de UB6 en clones de CHO establemente transfectados.

Por tanto, la colocación de STAR67 en dirección 5' del promotor y dos elementos STAR7 para flanquear los casetes dio como resultado niveles de expresión del anticuerpo EpCAM significativamente mayores en CHO, en comparación con constructos sin STAR67/STAR7.

Ejemplo 6: STAR67 no es un bloqueador del potenciador, mientras que STAR6 y STAR7 sí lo son.

Todos los elementos STAR conocidos hasta ahora que se ensayaron para determinar esa propiedad, incluyendo STAR6 y STAR7, son bloqueadores del potenciador (documento WO 03/004704, Kwaks et al, 2003). La actividad bloqueante del potenciador se ensaya colocando un elemento STAR entre un potenciador fuerte y un promotor. Aquí, se ensayó si también STAR67 es un bloqueador del potenciador.

Materiales y métodos

El gen d2EGFP se amplificó mediante PCR usando los cebadores TTGGTTGGTCATGAATGGTGAGCAAGGGCGAGGAGCTGTTC (SEQ.ID.NO. 75) y ATTCTCTAGACTACACATTGATCCTAGCAGAAGCAC (SEQ.ID.NO. 76), y se clonó en el plásmido pGL3-promoter (Promega) usando los sitios de restricción NcoI y XbaI para sustituir el gen de luciferasa para crear el plásmido pGL3-promoter-GFP. Un ligador, creado hibridando los oligo CGATATCTTGGAGATCTACTAGTGGCGCGCCTTGGGCTAGCT (SEQ. ID. NO. 77) y GATCAGCTAGCCCAAGGCGGCCACTAGTAGATCTCCAAGATATCGAGCT (SEQ. ID. NO. 78), se clonó en los sitios SacI y BglIII para crear múltiples sitios de clonación. El sitio BglIII original se destruyó con la ligación del ligador, creando un nuevo sitio BglIII único en el ADN ligador. El potenciador de SV40 se cortó del plásmido pGL3-basic (Promega) usando BsaBI y BamHI, y se clonó en el pGL3-ligador-promotor-GFP usando los sitios EcoRV y BglIII creando el plásmido pGL3-potenciador-promotor-GFP. El elemento STAR40 se colocó en dirección 5' del potenciador de SV40 usando los sitios KpnI y SacI para evitar la acción del potenciador en las secuencias en dirección 5'. Finalmente, los elementos STAR 6, 7 y 67 se colocaron entre el potenciador de SV40 y el promotor mínimo de SV40 usando los sitios de restricción SpeI y AscI.

Transfección y cultivo de células CHO

La línea celular de ovario de hámster chino CHO-K1 (ATCC CCL-61) se transfectó como en el ejemplo 2. Un día después de la transfección, los niveles de d2EGFP se midieron en el citómetro de flujo Epics XL (Beckman Coulter).

Resultados

La Fig. 10 muestra que STAR67 no es un bloqueador del potenciador, mientras que STAR6 y STAR7 actúan como bloqueadores del potenciador en el mismo ensayo. STAR6, STAR7 y STAR67 se clonaron entre el potenciador de SV40 y un promotor de SV40 mínimo en dirección 5' del gen de d2EGFP. Cuando no se clonó ningún elemento STAR entre el potenciador y el promotor, se produjo una fuerza activación transcripcional (arbitrariamente establecida a 100%). Cuando STAR6 o STAR7 se colocaron entre el potenciador y el promotor, la transcripción cayó hasta los niveles de valores iniciales, indicando que STAR6 y STAR7 son potentes bloqueadores del potenciador. Por el contrario, cuando STAR67 se clonó entre el potenciador y el promotor, los niveles de transcripción relativos todavía fueron 80% del control, indicando que STAR67 no es un buen bloqueador del potenciador, estando esto en contraste con STAR6 y STAR7, así como otros elementos STAR, como se describe previamente (documento WO 03/004704, Kwaks et al, 2003).

Ejemplo 7: STAR67 potencia los niveles de expresión de anticuerpo dirigida por UB6 y CMV en células CHO establemente transfectadas

En el ejemplo 5 se demostró que la combinación de STAR67 y STAR7 potenció los niveles de expresión del anticuerpo EpCAM en células CHO, en el contexto de dos plásmidos distintos, que contenían las cadenas pesada y ligera. En este ejemplo se ensayó si se podría usar STAR67 para la producción de anticuerpo EpCAM cuando las cadenas tanto pesada como ligera se colocan en un plásmido. Para cada marcador seleccionable se usó selección simultánea.

Materiales y métodos

Plásmidos

El ADNc de cadena pesada (HC-EpCAM) se coloca bajo el control del promotor de UB6 y se acopla al gen de resistencia a zeocina mediante una secuencia IRES. El ADNc de cadena ligera (LC-EpCAM) se coloca bajo el control del promotor de CMV y se acopla al gen de resistencia a puromicina mediante una secuencia IRES. Básicamente, estos son los constructos usados en el ejemplo 5. Estos dos casetes de expresión se colocaron en un plásmido, de una manera tal que la transcripción de las dos unidades de expresión tuvo direcciones opuestas. En el plásmido de control, los promotores de UB6 y CMV se separaron mediante un relleno de 500 pb (control de EpCAM) (Fig. 10). En otro plásmido, STAR67 se colocó entre el promotor de UB6 y CMV (EpCAM STAR67) (Fig. 10).

Transfección y cultivo de células CHO

La línea celular de ovario de hámster chino CHO-K1 (ATCC CCL-61) se transfirió y cultivó como en el ejemplo 2, usando zeocina (100 µg/ml) y puromicina (2,5 µg/ml) para la selección. En el ejemplo 5 se usó selección consecutiva para ambos marcadores de selección. Por el contrario, aquí estaban presentes simultáneamente ambos agentes de selección en el medio de cultivo. Un día después de la transfección, se añadió zeocina y puromicina al medio de cultivo. El medio de selección estaba presente hasta que se aislaron colonias (aproximadamente 14 días después de la transfección). Después de que las colonias se aislaron y se transfirieron a placas de 24 pocillos, las células se cultivaron en presencia de zeocina y puromicina.

Resultados

La Fig. 11 muestra que la transfección del constructo de anticuerpo que contiene STAR67 clonada entre los promotores de UB6 y CMV dio como resultado un número de colonias de CHO que expresan anticuerpo EpCAM (medido mediante ELISA usando un anticuerpo anti-IgG humana). La media de la producción de EpCAM en las 19 colonias transfectadas con el plásmido EpCAM STAR67 fue 9,8 pg/célula/día, cuando se midió 25 días después de la transfección.

Por el contrario, sorprendentemente no sobrevivieron colonias de la transfección con el plásmido de Control de EpCAM. Cuando se llevó a cabo la selección con zeocina o puromicina sola, las colonias de Control de EpCAM sobrevivieron. Sin embargo, cuando la presión de selección se incrementó poniendo presión de selección tanto en la cadena pesada como en la cadena ligera, estas condiciones sólo permitieron que sobreviviesen colonias que tienen una STAR67 presente en el plásmido transfectado.

Los resultados muestran también que la incorporación de STAR67 tiene un efecto beneficioso sobre dos promotores, los promotores de UB6 y CMV, que se colocan en dirección 5' y en dirección 3' de un elemento STAR67. Esto indica que STAR67 puede operar de manera bidireccional.

La diferencia entre el plásmido de Control de EpCAM y el plásmido de EpCAM STAR67 es blanco-negro, en el sentido de que solamente la transfección de EpCAM STAR67 da como resultado el establecimiento de colonias, cuando la presión de selección es elevada. Esto abre una oportunidad para usar esta configuración plasmídica para identificar la región en STAR67 que es responsable de mediar este efecto. Se colocan porciones más pequeñas, solapantes, de STAR67 entre los promotores de UB6 y CMV, que dirigen la molécula de EpCAM. Cuando una porción de STAR67 es funcional, las colonias sobrevivirán cuando se usan simultáneamente tanto zeocina como puromicina como agente de selección. Cuando una porción de STAR67 no es funcional, ninguna colonia sobrevivirá en condiciones de selección idénticas.

Por tanto, la colocación de STAR67 en dirección 5' del promotor dio como resultado niveles de expresión del anticuerpo EpCAM significativamente mayores en CHO, en comparación con constructos sin STAR.

Ejemplo 8: Construcción y ensayo de un producto génico de resistencia a zeocina sin metionina interna

La idea básica detrás del desarrollo de un nuevo sistema de selección es colocar el gen que codifica el gen de resistencia en dirección 5' de un gen de interés, y que un promotor dirija la expresión de este ARNm bicistrónico. La traducción del ARNm bicistrónico es tal que sólo en un pequeño porcentaje de sucesos de traducción el gen de resistencia se traducirá en proteína, y la mayoría del tiempo el gen de interés en dirección 3' se traducirá en proteína. Por tanto, la eficacia de la traducción del gen de resistencia en dirección 5' se debe impedir de forma importante en comparación con la eficacia de la traducción del gen de interés en dirección 3'. Para lograr esto, se pueden realizar tres etapas según la invención:

- 1) en el gen de resistencia en el ARNm, el ribosoma de búsqueda no debería encontrar preferiblemente otro AUG, puesto que cualquier AUG en dirección 3' puede servir como codón de inicio de la traducción, dando como resultado una menor eficacia de la traducción del segundo gen de interés en dirección 3'. Por tanto, preferiblemente cualquier AUG en el ARNm del gen de resistencia tendrá que ser sustituido. En el caso en

el que este AUG sea un codón funcional que codifique una metionina, este aminoácido tendrá que ser sustituido por un aminoácido diferente, por ejemplo por una leucina (Fig. 12A y B);

2) el codón de inicio del gen de resistencia debe tener un mal contexto (ser parte de una secuencia de inicio de la traducción no óptima); es decir, los ribosomas deben comenzar la traducción en este codón de inicio sólo en un número limitado de sucesos, y por tanto en la mayoría de los sucesos deben continuar buscando un codón de inicio mejor, más óptimo (Fig. 12C-E). Se pueden distinguir tres restricciones diferentes: a) el codón de inicio ATG normal, pero colocado en un mal contexto (TTTATGT (denominado ATGmut) (Fig. 12C), b) preferiblemente cuando se coloca en un contexto óptimo, GTG puede servir como codón de inicio (ACCGTGG) (Fig. 12D), y c) preferiblemente cuando se coloca en un contexto óptimo, TTG puede servir como un codón de inicio (ACCTTGG) (Fig. 12E). La condición de traducción más rigurosa es el codón TTG, seguido del codón GTG (Fig. 12). Se espera que el ARNm de Zeo con un TTG como codón de inicio produzca la proteína menos resistente a zeocina, y por tanto comunicará la resistencia a zeocina funcional más baja a las células (Figs. 12, 13).

3) preferiblemente, el codón de inicio normal (ATG) del gen de interés en dirección 3' debería tener un contexto de traducción óptimo (por ejemplo ACCATGG) (Fig. 13A-D). Esto garantiza que, después de que se han llevado a cabo las etapas 1 y 2, en la mayoría de los sucesos el codón de inicio del gen de interés funcionará como codón de inicio del ARNm bicistrónico.

En el ejemplo 8, se lleva a cabo la etapa 1, esto es, en el gen de resistencia a zeocina, una metionina interna existente se sustituye por otro aminoácido (Fig. 12B-E). Es importante que después de tal cambio la proteína Zeo todavía confiera resistencia a Zeocina a las células transfectadas. Puesto que no se sabe de antemano qué aminoácido cumplirá este criterio, se han intentado tres aminoácidos diferentes: leucina, treonina y valina. Los diferentes constructos con aminoácidos distintos se han ensayado entonces para determinar su capacidad para conferir todavía resistencia a zeocina a las células transfectadas.

Materiales y métodos

Construcción de los plásmidos

El marco de lectura abierto de Zeo original tiene la siguiente secuencia alrededor del codón de inicio: AAACCATGGCC (codón de inicio en negrita; SEQ. ID. NO. 83). Este es un codón de inicio con un contexto traduccional óptimo (Fig. 12A). En primer lugar, el contexto óptimo del codón de inicio del marco de lectura abierto de Zeo se cambió mediante amplificación a partir del plásmido pCMV-zeo [Invitrogen V50120], con el par de cebadores ZEOforwardMUT (SEQ. ID. NO. 84): GATCTCGCGATACAGGATTTATGTTGGCCAAGTTGACCAAGTGCCGTTCCG y ZEO-WTreverse (WT=tipo salvaje; SEQ. ID. NO. 85): AGGCGAATTCAGTCCTGCTCCTCGGC, usando como molde pCMV-ZEO (Invitrogen; V50120). El producto amplificado se cortó con NruI-EcoRI, y se ligó en pcDNA3, dando como resultado pZEOATGmut.

El marco de lectura abierto de Zeo original contiene ATG en el marco, que codifica metionina en la posición 94 de los aminoácidos (de 124). Este ATG interno, que codifica la metionina en la posición 94, se cambió de tal manera que la metionina se cambió en leucina, treonina o valina, respectivamente:

1) Para sustituir el codón interno para metionina en el marco de lectura abierto de Zeo por el codón para leucina (Fig. 12B), se amplificó parte del marco de lectura abierto de Zeo usando el par de cebadores ZEOforwardMUT (SEQ. ID. NO. 84) y ZEO-LEUreverse (SEQ. ID. NO. 86): AGGCCCCGCCCCACGGCTGCTCGCCGATCTCGGTCAAGGCCGGC. El producto de PCR se cortó con BamHI-BglI y se ligó en pZEOATGmut. Esto dio como resultado pZEO(leu).

Para sustituir el codón interno para metionina en el marco de lectura abierto de Zeo por el codón para treonina (no mostrado, pero como en Fig. 12B), se amplificó parte del marco de lectura abierto de Zeo usando el par de cebadores ZEOforwardMUT (SEQ. ID. NO. 84) y ZEO-THRreverse (SEQ. ID. NO. 87): AGGCCCCGCCCCACGGCTGCTCGCCGATCTCGGTGGTGCCGGC. El producto de PCR se cortó con BamHI-BglI y se ligó en pZEOATGmut. Esto dio como resultado pZEO(thr).

Para sustituir el codón interno para metionina en el marco de lectura abierto de Zeo por el codón para valina (no mostrado, pero como en la Fig. 12B) (GTG), se amplificó parte del marco de lectura abierto de Zeo usando el par de cebadores ZEOforwardMUT (SEQ. ID. NO. 84) y ZEO-VALreverse (SEQ. ID. NO. 88): AGGCCCCGCCCCACGGCTGCTCGCCGATCTCGGTCCACGCCGG. El producto de PCR se cortó con BamHI-BglI y se ligó en pZEOATGmut. Esto dio como resultado pZEO(val).

Transfección y cultivo de células

La línea celular de ovario de hámster chino CHO-K1 (ATCC CCL-61) se cultivó en medio HAMS-F12 + 10% de suero fetal de ternera que contiene 2 mM de glutamina, 100 U/ml de penicilina, y 100 microgramos/ml de estreptomina a 37°C/5% CO₂. Las células se transfectaron con los plásmidos usando Lipofectamine 2000 (Invitrogen) como se describe por el fabricante. De forma breve, las células se sembraron en vasijas de cultivo y se hicieron crecer toda la

noche hasta 70-90% de confluencia. El reactivo de Lipofectamine se combinó con ADN plasmídico a una relación de 6 microlitros por microgramo (por ejemplo para una cápsula de Petri de 10 cm, 20 microgramos de ADN y 120 microlitros de Lipofectamine), y se añadió a las células. Después de la incubación durante toda la noche, la mezcla de transfección se sustituyó por medio reciente, y las células transfectadas se incubaron adicionalmente. Tras el cultivo durante toda la noche, las células se tripsinizaron y se sembraron en vasijas de cultivo recientes con medio reciente que contiene zeocina (100 µg/ml). Cuando las colonias individuales se hicieron visibles (aproximadamente diez días después de la transfección), las colonias se contaron.

Resultados

Se transfectaron cuatro plásmidos en células CHO-K1, 1) pZEO(WT), 2) pZEO(leu), 3) pZEO(thr), y 4) pZEO(val). Las células se seleccionaron en 100 µg/ml de zeocina. La transfección de pZEO(leu) dio como resultado un número igual de colonias resistentes a zeocina en comparación con el control pZEO (WT). pZEO(thr) y pZEO(val) dieron menos colonias, pero las diferencias no fueron del orden de una magnitud. Por tanto, se concluyó que los cambios de la metionina interna en leucina, treonina o valina dieron todos como resultado una proteína de resistencia a zeocina que todavía es capaz de conferir resistencia a zeocina a las células transfectadas. De forma más bien arbitraria, se escogió pZEO(leu) como punto de partida para crear diferentes codones de inicio en el marco de lectura abierto de Zeo. Por tanto, en los ejemplos a continuación, las metioninas de inicio así como internas se sustituyen siempre por leucina, para zeocina, pero también para otros genes marcadores seleccionables, como será claro a partir de ejemplos posteriores.

Ejemplo 9: Creación y ensayo de constructos bicistrónicos de zeocina-d2EGFP con diferentes eficacias de traducción.

Para crear un ARNm bicistrónico que engloba un ARNm de resistencia a zeocina mutado con menos eficacia traduccional, y el gen de d2EGFP como gen de interés en dirección 3', se optimizó en primer lugar el codón de inicio del gen de d2EGFP (etapa 3 en el ejemplo 8). Después de eso, se crearon las diferentes versiones del gen de resistencia a zeocina. Las diferencias entre estas versiones son que tienen diferentes codones de inicio, con distinta eficacia traduccional (etapa 2 en el Ejemplo 9, Figura 12C-E). Estas diferentes versiones del gen de resistencia a zeocina se clonaron en dirección 5' del gen de d2EGFP modificado (Fig. 13).

Materiales y métodos

Creación de plásmidos

El ORF informador de d2EGFP se introdujo en pcDNA3. La secuencia alrededor del codón de inicio de este ADNc de d2EGFP es **GAATTCATGGG** (codón de inicio en negrita; SEQ. ID. NO. 89), que no es óptima. Como una primera etapa, se amplificó d2EGFP a partir de pd2EGFP (Clontech 6010-1) con los cebadores d2EGFPforwardBamHI (SEQ. ID. NO. 90): GATCGGATCCTATGAGGAATTCGCCACCA TGGTGAGCAAGGGCGAGGAG y d2EGFPreverseNotI (SEQ. ID. NO. 91): AAGGAAAAAAGCGGCCGCCTACACATTGATCCTAGCAGAAG. Este producto contiene ahora un codón de inicio con un contexto traduccional óptimo (**ACCATGG**). Esto creó pd2EGFP, y subsiguientemente el marco de lectura abierto de Zeo se ligó en pd2EGFP, dando como resultado pZEO-d2EGFP. Se señala aquí que la optimización de la secuencia de inicio traduccional del gen de interés (aquí: EGFP como un gen modelo) no es esencial pero sí preferida a fin de inclinar todavía más la frecuencia de iniciación de la traducción hacia el gen de interés.

Ahora se obtuvieron tres clases de constructos:

1) ATG como un codón de inicio en el gen de resistencia a Zeo, pero en un mal contexto (**TTTATGT**) (no mostrado, pero como en la Fig. 13B) y seguido de secuencia espaciadora, en lugar del ATG óptimo (Fig. 13 A).

La secuencia espaciadora se coloca en dirección 3' de la secuencia de ATG. En el ARN de zeocina (y posiblemente en la blasticidina), está presente una estructura secundaria, que provoca que se retrase temporalmente el ribosoma. Debido a esto, en algunos casos el ribosoma puede usar un codón de inicio pobre, a pesar de ser un codón de inicio malo o estar en un contexto no óptimo para la iniciación de la traducción. Esto provoca que aumente la oportunidad de la traducción, y, en el caso de la actual invención, hace por lo tanto más baja la rigurosidad para la selección. Para disminuir este efecto, y por tanto disminuir adicionalmente la eficacia de la iniciación de la traducción, se introduce una secuencia espaciadora que no contiene una estructura secundaria (Kozak, 1990). Por tanto, se introduce el término "espacio", y se usa en los nombres del plásmido y de los cebadores para indicar la presencia de tal secuencia espaciadora. El espaciador elimina la "secuencia que retrasa el ribosoma" de la vecindad del codón de iniciación, provocando con ello que el ribosoma comience a traducir menos frecuentemente, y por tanto aumentando la rigurosidad de la selección según la invención. El espaciador introduce algunos aminoácidos extra en la secuencia codificante. Esto se ha hecho en algunos casos tanto para zeocina como para blasticidina, como será manifiesto a partir de los ejemplos. La nomenclatura de los plásmidos y cebadores en general en lo siguiente sigue estas líneas: el nombre del polipéptido marcador seleccionable se cita mediante abreviatura (por ejemplo Zeo, Blas, etc); se menciona el codón de inicio (por ejemplo ATG, GTG, TTG); cuando este

codón de inicio se coloca en un contexto no óptimo para la iniciación de la traducción, se usa la adición "mut" (esto sólo se hace habitualmente para codones de inicio ATG, puesto que la combinación de un contexto no óptimo con un codón de inicio no ATG habitualmente no da como resultado una iniciación suficiente de la traducción para permitir la selección), y cuando se usa una secuencia espaciadora detrás del codón de inicio se usa la adición "space" (esto se hace habitualmente para codones de inicio "ATGmut" para los marcadores seleccionables Zeo o Blas).

El marco de lectura abierto de Zeo se amplificó con el par de cebadores ZEOforwardBamHI-**ATGmut**/space (SEQ. ID. NO. 93):

GATCGGATCCTTGG**TTTATG**TCGATCCAAAGACTGCCAAATCTAGATCCGA

GATTTTCAGGAGCTAAGGAAGCTAAAGCCAAGTTGACCAGTGAAGTTC

(en el que la secuencia tras la secuencia subrayada comprende la secuencia espaciadora), y ZEOwTreverse (SEQ. ID. NO. 85), el producto de PCR se cortó con EcoRI-BamHI, y se ligó en pd2EGF, cortado con EcoRI-BamHI, creando pZEO-**ATGmut/space**-d2EGFP.

2) GTG como un codón de inicio en el gen de resistencia a Zeo, en lugar de ATG (Fig. 13C).

El marco de lectura abierto de Zeo se amplificó con el par de cebadores ZEOforwardBamHI-**GTG** (SEQ. ID. NO. 94): GATCGGATCC**ACCGTGG**CCAAGTTGACCAGTGCCGTTC y ZEOwTreverse (SEQ. ID. NO. 85), el producto de PCR se cortó con EcoRI-BamHI, y se ligó en pd2EGFP, cortado con EcoRI-BamHI, creando pZEO-**GTG-d2EGFP**.

3) TTG como un codón de inicio en el gen de resistencia a Zeo, en lugar de ATG (Fig. 13D). El marco de lectura abierto de Zeo se amplificó con el par de cebadores ZEOforwardBamHI-**TTG**: GATCGGATCC**ACCTTGG**CCAAGTTGACCAGTGCCGTTC (SEQ. ID. NO. 95) y ZEOwTreverse (SEQ. ID. NO. 85), el producto de PCR se cortó con EcoRI-BamHI, y se ligó en pd2EGFP, cortado con EcoRI-BamHI, creando pZEO-**TTG**-d2EGFP.

Transfección, cultivo y análisis de células CHO

La línea celular de ovario de hámster chino CHO-K1 (ATCC CCL-61) se cultivó en medio HAMS-F12 + 10% de suero fetal de ternera que contiene 2 mM de glutamina, 100 U/ml de penicilina y 100 microgramos/ml de estreptomina a 37°C/5% CO₂. Las células se transfectaron con los plásmidos usando Lipofectamine 2000 (Invitrogen) como se describe por el fabricante. De forma breve, las células se sembraron en vasijas de cultivo y se hicieron crecer toda la noche hasta 70-90% de confluencia. El reactivo de Lipofectamine se combinó con ADN plasmídico a una relación de 15 microlitros por 3 microgramos (por ejemplo para una cápsula de Petri de 10 cm, 20 microgramos de ADN y 120 microlitros de Lipofectamine), y se añadió después de 30 minutos de incubación a 25°C a las células. Después de la incubación durante toda la noche, la mezcla de transfección se sustituyó por medio reciente, y las células transfectadas se incubaron adicionalmente. Después del cultivo durante toda la noche, las células se tripsinizaron y se sembraron en vasijas de cultivo recientes con medio reciente. Después de otra incubación durante toda la noche, se añadió zeocina a una concentración de 50 µg/ml, y las células se cultivaron adicionalmente. Después de otros tres días, el medio se sustituyó por medio reciente que contiene zeocina (100 µg/ml) y se cultivó adicionalmente. Cuando las colonias individuales se hicieron visibles (aproximadamente diez días después de la transfección), el medio se eliminó y se sustituyó por medio reciente sin zeocina. Los clones individuales se aislaron y se transfirieron a placas de 24 pocillos en medio sin zeocina. Un día después de aislar las colonias, se añadió zeocina al medio. La expresión del gen informador d2EGFP se evaluó aproximadamente 3 semanas después de la transfección. Los niveles de expresión de d2EGFP en las colonias se midió después de períodos de dos semanas.

Resultados

Células CHO-K1 se transfectaron con constructos que contienen los genes ATGmut/space Zeo (Fig. 13B), GTG Zeo (Fig. 13C) y TTG Zeo (Fig. 13D) como genes de selección, clonándose todos ellos en dirección 5' del gen informador d2EGFP. Estos tres constructos carecían de elementos STAR (Control) o tenían elementos STAR 7 y 67 en dirección 5' del promotor de CMV, y STAR 7 en dirección 3' del gen de d2EGFP (Fig. 14). La Fig. 14 muestra que ambos constructos de control (sin elementos STAR) con ATGmut/space Zeo (A) y GTG Zeo (B) dieron colonias que expresaron la proteína d2EGFP. El nivel medio de expresión de d2EGFP de 24 colonias ATGmut/space Zeo fue 46, y de colonias GTG Zeo fue 75. Este mayor nivel medio de expresión en colonias GTG Zeo puede reflejar la mayor rigurosidad de GTG, en comparación con ATGmut/space (ejemplo 8). La adición de elementos STAR 7 y 67 a dos constructos dio como resultado colonias que tuvieron mayores niveles medios de expresión de d2EGFP. La transfección del constructo ATGmut/space Zeo STAR 7/67/7 dio como resultado colonias con un nivel medio de expresión de d2EGFP de 118, que es un factor 2,6 veces mayor que la media en las células de control (46). La adición de los elementos STAR al constructo GTG Zeo dio como resultado un nivel medio de expresión de d2EGFP de 99, que es un factor 1,3 veces mayor que la media en las células de control (75).

De forma importante, no se establecieron colonias cuando se transfectó el constructo TTG Zeo. Sin embargo, el constructo con TTG Zeo, flanqueado con STARs 7 y 67, dio como resultado el establecimiento de 6 colonias, con un nivel medio de expresión de d2EGFP de 576 (Fig. 14C). De este modo, la rigurosidad de la traducción más elevada, provocada por el codón de inicio TTG (Fig. 12), conduce a los niveles de expresión de d2EGFP más elevados, como se predice en Fig. 13. Los resultados también indican que la rigurosidad del TTG Zeo solo (sin elementos STAR) es al menos, en algunos experimentos, demasiado elevada para que las colonias sobrevivan. Sin embargo, en experimentos independientes posteriores (véase más abajo), se encontraron algunas colonias con este constructo sin elementos STAR, indicando que la rigurosidad del sistema de selección con el codón de inicio TTG en el marcador de selección de zeocina no excluye necesariamente el hallazgo de colonias cuando no están presentes elementos STAR, y que el número de colonias obtenidas puede variar entre experimentos.

Se concluye que el uso de elementos STAR en combinación con el sistema de selección riguroso según la invención permite identificar fácilmente productores elevados del gen de interés.

Ejemplo 10: Establecimiento de un mayor número de colonias TTG Zeo STAR en comparación con un constructo IRES-Zeo.

Los resultados en el ejemplo 9 indican que el TTG Zeo tiene una eficacia de la traducción extremadamente rigurosa, que puede ser demasiado elevada para conferir resistencia a zeocina a las células. La transfección se aumentó de escala para ensayar si habría algunas colonias que tienen tales niveles elevados de expresión de manera que sobrevivieran. El aumento de escala del experimento podría resolver también la cuestión de si la media elevada de TTG Zeo STAR 7/67/7 sería mayor cuando se analizaran más colonias.

Materiales y métodos

Se transfectaron células CHO-K1 con los constructos que tienen el gen TTG Zeo como marcador de selección, con y sin elementos STAR 7 y 67 (Fig. 15). Las transfecciones, la selección, el cultivo, etc., fueron como en el ejemplo 9, excepto que se usaron 6 veces más células, ADN y Lipofectamine 2000. Las transfecciones y la selección se realizaron en cápsulas de Petri.

Resultados

La Fig. 15A muestra que la transfección con el constructo TTG Zeo STAR 7/67/7 dio como resultado la generación de muchas colonias con una señal media de d2EGFP de 560. Esto es tan elevado como en el ejemplo 9, excepto que ahora se analizaron 58 colonias. Cuando se compara con un constructo con el gen de resistencia a zeocina colocado detrás de una secuencia IRES (Fig. 15B), el nivel medio de expresión de d2EGFP fue 61, y cuando los elementos STAR 7 y 67 se añadieron a tal constructo, el nivel medio de expresión de d2EGFP fue 125, un factor de 2 por encima del control (Fig. 15B). La media de las colonias de TTG Zeo STAR 7/67/7 fue por lo tanto un factor 9,2 veces mayor que las colonias de IRES-Zeo sin STAR, y un factor de 4,5 veces mayor que las colonias de STAR7/67/7 IRES Zeo.

Una observación es que la forma de la curva de todas las colonias que se expresan difiere entre TTG Zeo STAR7/67/7 e IRES-Zeo STAR 7/67/7. En el primer caso (TTG Zeo), la curva se estabiliza, mientras que en el segundo caso (IRES-Zeo) la curva tiene una forma más "exponencial". La meseta en la curva de TTG Zeo podría indicar que las células han alcanzado un nivel máximo de expresión de d2EGFP, por encima del cual los niveles de expresión de d2EGFP se hacen tóxicos y las células mueren. Sin embargo, parece más tarde que los valores elevados estuvieron próximos al valor máximo que se puede detectar con los ajustes del detector del analizador de FACS. En experimentos posteriores, los ajustes del analizador de FACS se cambiaron para permitir la detección de valores mayores, y en algunos casos se midieron de hecho, en experimentos independientes posteriores, valores mayores que los obtenidos aquí (véase más abajo).

Debido al aumento de escala de las transfecciones, se pudieron escoger tres colonias con el constructo TTG Zeo sin STAR. Los niveles de expresión de d2EGFP de estas colonias fueron 475, 158 y 43. La última colonia murió pronto después de la primera medición. Este resultado indica que el constructo TTG Zeo puede conferir resistencia a zeocina, dando como resultado colonias que también pueden dar en algunos casos niveles elevados de expresión. Por tanto, el nuevo método de selección según la invención se puede aplicar con casetes de expresión que no contienen elementos de control de cromatina, aunque se prefiere claramente usar casetes de expresión que comprenden al menos uno de tales elementos, preferiblemente un elemento STAR.

Los resultados indican que los elementos STAR permiten un sistema de selección más riguroso según la invención, tal como se ejemplifica en este ejemplo, dando como resultado la recogida de colonias que tienen un nivel medio de expresión proteica muy elevado.

Ejemplo 11: Creación y ensayo de constructos bicistrónicos Blastidina-d2EGFP con eficacias de traducción diferenciales.

Hay cuatro ATG internos en el gen de resistencia a blastidina, ninguno de los cuales codifica una metionina (Fig. 25A). Estos ATG se han de eliminar sin embargo (Fig. 25B), puesto que servirán como codón de inicio cuando se

haya modificado el codón de inicio ATG (o su contexto), y esto dará como resultado péptidos que no se asemejan a la proteína de resistencia a blasticidina. De forma más importante, estos ATGs evitarán la traducción eficaz del gen de interés, como se representa mediante d2EGFP en este ejemplo con fines ilustrativos. Para eliminar los ATGs internos, se amplificó en primer lugar el marco de lectura abierto de la proteína de resistencia a blasticidina con cuatro pares de cebadores, generando 4 fragmentos de proteína de resistencia a blasticidina. Los pares de cebadores fueron:

A) BSDBamHlforward (SEQ. ID. NO. 96):

GATCGGATCCACCATGGCCAAGCCTTTGTCTCAAG

BSD 150reverse (SEQ. ID. NO. 97):

GTAAATGATATACGTTGACACCAG

B) BSD 150forward (SEQ. ID. NO. 98): CTGGTGTCAACGTATATCATTTTAC

BSD250reverse (SEQ. ID. NO. 99): GCCCTGTTCTCGTTTCCGATCGCG

C) BSD250forward (SEQ. ID. NO. 100):

CGCGATCGGAAACGAGAACAGGGC

BSD350reverse (SEQ. ID. NO. 101): GCCGTCGGCTGTCCGTCCTACTGTCC

D) BSD350forward (SEQ. ID. NO. 102): GGACAGTGACGGACAGCCGACGGC

BSD399reverse (SEQ. ID. NO. 103):

GATCGAATTCTTAGCCCTCCCACACGTAACCAGAGGGC

Los fragmentos A a D se aislaron de un gel de agarosa y se mezclaron juntos. Seguidamente, sólo se usaron los cebadores BSDBamHlforward y BSD399reverse para crear el ADNc de la proteína de resistencia a blasticidina de longitud completa, pero sustituyendo todos los ATG internos. La blasticidina reconstituida se cortó entonces con EcoRI-BamHI, y se clonó en pZEO-**GTG**-d2EGFP, cortado con EcoRI-BamHI (que libera Zeo), dando como resultado pBSDmut-d2EGFP. Todo el marco de lectura abierto de la proteína de resistencia a blasticidina se secuenció para verificar que se habían sustituido todos los ATGs.

Con este gen mutado que codifica proteína de resistencia a blasticidina (Blas), se obtuvieron tres clases de constructos (Fig. 25C-E):

1) ATG como codón de inicio, pero en un mal contexto y seguido de secuencia espaciadora. El marco de lectura abierto de la proteína de resistencia a blasticidina mutado en pBSD-d2EGFP se amplificó usando los cebadores BSDforwardBamHIAvrII-ATGmut/space (SEQ. ID. NO. 104):

**GATCGGATCCTAGGTTGGTTTATGTCGATCCAAAGACTGCCAAATCTAGAT
CCGAGATTTTCAGGAGCTAAGGAAGCTAAAGCCAAGCCTTTGTCTCAAGA
AG,**

y BSD399reverseEcoRIAvrII (SEQ. ID. NO. 105):

GATCGAATTCCTAGGTTAGCCCTCCCACACGTAACCAGAGGGC,

el producto de la PCR se cortó con BamHI-EcoRI, y se ligó en pZEO-**GTG**-d2EGFP, cortado con EcoRI-BamHI. Esto da como resultado pBSD-**ATGmut/space**-d2EGFP.

2) GTG como codón de inicio en lugar de ATG. El marco de lectura abierto de la proteína de resistencia a blasticidina mutado en pBSDd2EGFP se amplificó usando los cebadores BSDforwardBamHIAvrII-**GTG** (SEQ. ID. NO. 106): GATCGGATCCTAGG**ACCGTGG**CCAAGCCTTTGTCTCAAGAAG y BSD399reverseEcoRIAvrII (SEQ. ID. NO. 105), el producto de la PCR se cortó con BamHI-EcoRI, y se ligó en pZEO-**GTG**-d2EGFP, cortado con EcoRI-BamHI. Esto da como resultado pBSD-**GTG**-d2EGFP.

3) TTG como codón de inicio en lugar de ATG. El marco de lectura abierto de blasticidina mutado en pBSD-d2EGFP se amplificó usando los cebadores BSDforwardBamHIAvrII-**TTG** (SEQ. ID. NO. 107): GATCGGATCCTAGG**ACCTTGG**CCAAGCCTTTGTCTCAAGAAG y BSD399reverseEcoRIAvrII (SEQ. ID. NO. 105), el producto de la PCR se cortó con BamHI-EcoRI, y se ligó en pZEO-**GTG**-d2EGFP, cortado con EcoRI-BamHI. Esto da como resultado pBSD-**TTG**-d2EGFP.

Resultados

Se transfectaron células CHO-K1 con constructos que contienen los genes GTG Blas (Fig. 16A) y TTG Blas (Fig. 16B) como gen de selección, clonándose todos en dirección 5' del gen informador d2EGFP. La selección tuvo lugar en presencia de 20 µg/ml de blasticidina. Los dos constructos carecían de elementos STAR (Control), o tenían elementos STAR 7 y 67 en dirección 5' del promotor de CMV, y STAR7 en dirección 3' del gen d2EGFP (Fig. 16). La Fig. 16 muestra que ambos constructos de control (sin elementos STAR) con GTG Blas (A) y TTG Blas (B) dieron colonias que expresaron la proteína d2EGFP. La señal media de d2EGFP de 24 colonias GTG Blas fue 14,0 (Fig. 16A), y la de las colonias TTG Blas fue 81 (Fig. 16B). Este mayor nivel medio de expresión en colonias TTG Blas puede reflejar la mayor rigurosidad de TTG, en comparación con GTG (véase también el ejemplo 9). Sin embargo, sólo sobrevivieron 8 colonias bajo las condiciones TTG más restrictivas.

La adición de elementos STAR 7 y 67 a los constructos dio como resultado colonias que tuvieron mayores niveles medios de expresión de d2EGFP. La transfección del constructo GTG Blas STAR 7/67/7 dio como resultado colonias con un nivel medio de expresión de d2EGFP de 97,2 (Fig. 16A), que es un factor 6,9 veces mayor que la media en las células de control (14,0). La adición de elementos STAR al constructo TTG Blas dio como resultado una señal media de d2EGFP de 234,2 (Fig. 16B), que es un factor 2,9 mayor que la media en las células de control (81). Sin embargo, obsérvese nuevamente que sólo sobrevivieron 8 colonias a las condiciones duras de selección de TTG Blas, mientras que sobrevivieron 48 colonias con TTG Blas STAR 7/67/7. Cuando se comparan solamente los cinco valores más elevados, la media de los cinco TTG Blas más elevados fue 109,1, y la media de los cinco TTG Blas STAR 7/67/7 más elevados fue 561,2, que es un factor 5,1 veces mayor.

Los resultados indican que los elementos STAR permiten un sistema de selección más riguroso, dando como resultado la recogida de colonias que tienen un nivel medio de expresión proteica muy elevado. También muestran que esta selección no está restringida a la proteína de resistencia a zeocina sola, sino que también se pueden usar otros polipéptidos marcadores de selección, en este caso la proteína de resistencia a blasticidina.

Ejemplo 12: Estabilidad de la expresión de d2EGFP en el nuevo sistema de selección.

Las colonias descritas en el ejemplo 10 se cultivaron adicionalmente en varias condiciones para evaluar la estabilidad de la expresión de d2EGFP a lo largo de un período de tiempo prolongado.

Resultados

Las colonias que contienen TTG Zeo STAR 7/67/7 en la Fig. 15A se cultivaron durante 70 días adicionales en presencia de 100 µg/ml de zeocina. Como se muestra en la Fig. 17, la señal media de d2EGFP aumentó desde 560,2 después de 35 días hasta 677,2 después de 105 días. Excepto por algunas colonias raras, todas las colonias tuvieron un mayor nivel de expresión de d2EGFP.

Cuando el nivel de zeocina se redujo hasta 20 µg/ml de zeocina, todavía hubo un incremento en el nivel medio de expresión de d2EGFP, desde 560,2 después de 35 días hasta 604,5 después de 105 días (Fig. 18).

Cuando no había presión de selección en absoluto debido a la eliminación de la zeocina del medio de cultivo, aproximadamente 50% de las colonias se convirtieron en un mosaico, esto es, dentro de una colonia se manifestaron células que no expresan d2EGFP. Esto dio como resultado una reducción de los niveles de expresión de d2EGFP a menos de 50% de los niveles originales. Si la señal se hizo menor que 67% (disminución de al menos un tercio) desde la señal original, la colonia se consideró que era inestable con respecto a la expresión de d2EGFP. De las 57 colonias originales, 27 colonias permanecieron estables según este criterio; la señal media de d2EGFP de estas colonias después de 35 días (mientras todavía estaban bajo presión de selección) fue 425,6, mientras que la señal media de d2EGFP sin presión de selección después de 65 días fue 290,0. Cuando se midió después de 105 días, la señal media en las 27 colonias fue 300,9. Por tanto, después de una disminución inicial, los niveles de expresión en las 27 colonias permanecieron estables según este criterio (Fig. 19).

Seis de las colonias se sometieron a una ronda de subclonación. Las células se sembraron en placas de 96 pocillos de tal manera que cada pocillo contuvo aproximadamente 0,3 células. No había zeocina en el medio, de manera que, desde el comienzo, los subclones crecieron sin presión de selección. De cada colonia original, se aislaron al azar seis subclones y se hicieron crecer en placas de 6 pocillos hasta el análisis. En la Fig. 23 se comparan los valores originales de los clones originales, como ya se mostraron en la Fig. 15 A, con uno de los subclones. En uno de los seis clones (clon 25), no había ningún subclón con señal de d2EGFP en el intervalo del clon original. Sin embargo, en cinco de seis casos, al menos uno de los subclones tuvo niveles de expresión de d2EGFP iguales al clon progenitor. Estos niveles de expresión se determinaron después de 50 días sin presión de selección. Se concluye que una ronda de subclonación es suficiente para obtener un número elevado de colonias que permanecen estables para la expresión elevada en ausencia de presión de selección. Esto se ha confirmado en un experimento similar (no mostrado).

Se comparó el número de copias que se integraron en las colonias TTG Zeo STAR 7/67/7. Se aisló ADN cuando las colonias estuvieron 105 días bajo presión de selección de zeocina (véase la Fig. 17). Como se muestra en la Fig. 24, se pudieron distinguir dos subpoblaciones. En la Fig. 24, el corte se realizó a 20 copias, y se calculó el valor R^2 y se

muestra. También se muestra el valor R^2 a partir de datos con copias mayores de 20. En el intervalo de la señal de d2EGFP de 100 a 800, hubo un grado elevado de dependencia del número de copias, como se denota mediante un R^2 relativamente elevado de 0,5685 (Fig. 24). Sin embargo, la población de colonias que fluctúa alrededor de una señal de d2EGFP de 800, se observó una variación elevada en el número de copias (Fig. 24), como se denota con R^2 bajo de 0,0328. Juntos, los datos muestran que, en el nuevo sistema de selección, en colonias que contienen constructos TTG Zeo STAR 7/67/7, hay una expresión de d2EGFP dependiente del número de copias hasta ~20 copias. También, aunque la dependencia del número de copias se pierde cuando hay >20 copias, todavía una proporción sustancial de las colonias con una señal de d2EGFP elevada (>800) tienen no más de 30 copias (Fig. 24). Esta combinación entre expresión elevada de d2EGFP y un número relativamente bajo de copias (entre 10 y 30) puede ser importante para identificar colonias que siguen siendo relativamente estables sin presión de selección. Es una ventaja tener clones con números relativamente bajos de copias (menos de alrededor de 30, más preferiblemente menos de alrededor de 20) que dan niveles elevados de expresión, debido a que se cree que tales clones son menos susceptibles a la inestabilidad genética. El presente sistema de selección permite generar tales clones, incluyendo a partir de células CHO.

15 **Ejemplo 13: Creación y ensayo de constructos bicistrónicos zeocina-blasticidina-EpCAM con eficacias de traducción diferenciales.**

Para ensayar el sistema de selección en la producción de un anticuerpo, se tomó como ejemplo el anticuerpo anti-EpCAM (véase también el ejemplo 5).

Resultados

20 Se creó un plásmido en el que tanto la cadena pesada (HC) como la cadena ligera (LC) se colocaron cada una en una unidad de transcripción separada (Fig. 20-22). La expresión de ambas cadenas fue conducida por el promotor de CMV. En dirección 5' de la cadena pesada de EpCAM, se colocó el gen de resistencia a zeocina, ya sea con el ATGmut/space (Fig. 20), GTG (Fig. 21) o TTG (Fig. 22) como codón de inicio (véase el ejemplo 9). En dirección 5' de la cadena ligera de EpCAM, se colocó el gen de resistencia a blasticidina, ya sea con el ATGmut/space (Fig. 20), GTG (Fig. 21) o TTG (Fig. 22) como codón de inicio (véase el ejemplo 11). Se obtuvieron dos tipos de constructos, un constructo sin elementos STAR (Control) y un constructo con una combinación de elementos STAR 7 y 67. Los elementos STAR se colocaron según lo siguiente: en dirección 5' de cada promotor de CMV (es decir, uno para la unidad de transcripción que comprende HC, y uno para la unidad de transcripción que comprende LC) se colocó STAR 67, y el constructo resultante se flanqueó con un elemento STAR 7 5' y 3' (Figs. 20-22). Todos los constructos se transfectaron a células CHO-K1, y se seleccionaron en 100 $\mu\text{g/ml}$ de zeocina y 20 $\mu\text{g/ml}$ de blasticidina (al mismo tiempo). Después de la selección, se aislaron colonias independientes y se propagaron bajo presión de selección continua (usando 100 $\mu\text{g/ml}$ de zeocina y 20 $\mu\text{g/ml}$ de blasticidina). La Fig. 20 muestra que la combinación STAR 7/67/7 tuvo un efecto beneficioso sobre la producción de EpCAM. El ATGmut/space Zeo y ATGmut/space Blas no tuvieron efectos sobre el número de colonias que se formaron con plásmidos que contienen o no elementos STAR. Sin embargo, los niveles medios de expresión de EpCAM de 24 colonias de control frente a colonias STAR 7/67/7 osciló desde 0,61 pg/célula/día en el control hasta 3,44 pg/célula/día en el constructo STAR7/67/7 (Fig. 20). Esto es un factor de 5,6 de incremento. Puesto que hubo muchas colonias en el control ATGmut/space con 0 pg/célula/día, también se comparó la producción media de EpCAM en las cinco colonias más elevadas. En el ATGmut/space de control, esta fue de 3,0 pg/célula/día, frente a 7,8 pg/célula/día con el constructo ATGmut/space STAR 7/67/7, un incremento de un factor de 2,6.

La Fig. 21 también muestra que la combinación STAR 7/67/7 tuvo un efecto beneficioso sobre la producción de EpCAM, usando el codón de inicio GTG para los marcadores. Con el constructo GTG Zeo y GTG Blas STAR 7/67/7, se formaron aproximadamente 2 veces más colonias. También, los niveles medios de expresión de EpCAM de 24 colonias de control frente a colonias STAR 7/67/7 osciló desde 2,44 pg/célula/día en el control hasta 6,51 pg/célula/día en el constructo STAR7/67/7 (Fig. 21). Este es un factor de 2,7 de incremento. También se comparó la producción media de EpCAM en las cinco colonias más elevadas. En el GTG de control, esta fue 5,7 pg/célula/día, frente a 13,0 pg/célula/día con el constructo GTG STAR 7/67/7, un incremento de un factor de 2,3. También obsérvese que la producción media de EpCAM mediada por el codón de inicio GTG para los marcadores de selección fue significativamente mayor que con el codón de inicio ATGmut/space.

La Fig. 22 muestra que con el constructo de control TTG Zeo y TTG Blas no se formaron colonias, de manera similar como en el ejemplo 9. Con el constructo STAR 7/67/7 TTG se formaron colonias. Los niveles medios de expresión de EpCAM de las colonias STAR 7/67/7 TTG fue 10,4 pg/célula/día (Fig. 22). Esto es nuevamente mayor que con el ATGmut/space y GTG como codón de inicio (véanse las Figs. 20, 21 para comparación). La producción media de EpCAM en las cinco colonias TTG STAR 7/67/7 más elevadas fue 22,5 pg/célula/día.

Los resultados muestran que el sistema de selección también se puede aplicar a dos polipéptidos producidos simultáneamente, en este caso dos polipéptidos de una proteína multimérica, según sea el caso, un anticuerpo. La producción de EpCAM sigue estrechamente los resultados obtenidos con d2EGFP. El TTG como codón de inicio es más restrictivo que el codón de inicio GTG, que a su vez es más restrictivo que el ATGmut/space (Figs. 12 y 13). Una mayor rigurosidad da como resultado un número decreciente de colonias, sin colonias en el caso del control de

TTG, que no tiene elementos STAR, y una mayor restricción del marcador de selección está acoplado a una mayor expresión de la proteína de interés.

Ejemplo 14: Creación y ensayo de constructos bicistrónicos GTG zeocina-d2EGFP adicionales con eficacias de traducción diferenciales.

5 En el Ejemplo 8 se describieron diferentes versiones del gen de resistencia a zeocina con codones de inicio mutados. Además de los codones GTG descritos (Ejemplo 8, Fig. 33A), son posibles codones de inicio modificados adicionales con eficacia traduccional distinta. Estas versiones diferentes del gen de resistencia a zeocina se crearon (Fig. 33) y se clonaron en dirección 5' del gen de d2EGFP modificado, como en el Ejemplo 9.

Materiales y métodos

10 Creación de plásmidos

Se obtuvieron cuatro constructos GTG adicionales:

1) GTG como un codón de inicio en el gen de resistencia a Zeo (Fig. 33A), pero seguido de una secuencia espaciadora (Fig. 33B). El marco de lectura abierto de mutspace-Zeo se amplificó con el par de cebadores GTGspaceBamHIF (SEQ. ID. NO. 122):

15 GAATTCGGATCCACCGTGGCGATCCAAAGACTGCCAAATCTAG y

(en el que la secuencia que sigue a la secuencia subrayada comprende la secuencia espaciadora), y ZEOwTreverse (SEQ. ID. NO. 85), el producto de PCR se cortó con EcoRI-BamHI, y se ligó en pd2EGFP, cortado con EcoRI-BamHI, creando pZEO- GTGspace-d2EGFP.

20 2) GTG como un codón de inicio en el gen de resistencia a Zeo, pero en un mal contexto (TTTGTG) (Fig. 33C). El marco de lectura abierto de Zeo se amplificó con el par de cebadores ZEOwTGTGBamHIF (SEQ. ID. NO. 123):

GAATTCGGATCCTTTGTGGCCAAGTTGACCAGTGCCGTTCCG y

ZEOwTreverse (SEQ. ID. NO. 85), el producto de PCR se cortó con EcoRI-BamHI, y se ligó en pd2EGFP, cortado con EcoRI-BamHI, creando pZEO(leu)-TTTGTG-d2EGFP.

25 3) GTG como un codón de inicio en el gen de resistencia a Zeo, en lugar de ATG (Fig. 33A), pero con una mutación adicional en el marco de lectura abierto de Zeo en Pro9, que se substituyó por treonina (Thr) (Fig. 33D). La mutación Thr9 se introdujo amplificando el marco de lectura abierto de Zeo con el par de cebadores ZEOwForwardGTG-Thr9 (SEQ. ID. NO. 124):

AATTGGATCCACCGTGGCCAAGTTGACCAGTGCCGTTACCGTGCTC

30 y ZEOwTreverse (SEQ. ID. NO. 85), el producto de PCR se cortó con EcoRI-BamHI, y se ligó en pd2EGFP, cortado con EcoRI-BamHI, creando pZEO-GTG-Thr9-d2EGFP.

35 4) GTG como un codón de inicio en el gen de resistencia a Zeo, en lugar de ATG (Fig. 33A), pero con una mutación adicional en el marco de lectura abierto de Zeo en Pro9, que se substituyó por fenilalanina (Phe) (Fig. 33E). La mutación Phe9 se introdujo amplificando el marco de lectura abierto de Zeo con el par de cebadores ZEOwForward GTG-Phe9 (SEQ. ID. NO. 125):

AATTGGATCCACCGTGGCCAAGTTGACCAGTGCCGTTTTCTGCTC

y ZEOwTreverse (SEQ. ID. NO. 85), el producto de la PCR se cortó con EcoRI-BamHI, y se ligó en pd2EGFP, cortado con EcoRI-BamHI, creando pZEO-GTG-Phe9-d2EGFP.

Transfección, cultivo y análisis de células CHO

40 La transfección, el cultivo y el análisis de células CHO-K1 se realizó como en el Ejemplo 8.

Resultados.

45 Células CHO-K1 se transfectaron con constructos que contienen los genes de GTG Zeo (Fig. 33A), GTG space Zeo (Fig. 33B), TTT GTG Zeo (también denominado: GTGmut Zeo) (Fig. 33C), GTG Thr9 Zeo(leu) (Fig. 33D) y GTG Phe9 Zeo(leu) (Fig. 33E) como gen de selección, clonándose todos ellos en dirección 5' del gen informador de d2EGFP. Estos cinco constructos carecían de elementos STAR (Control) o tenían elementos STAR 7 y 67 en dirección 5' del promotor de CMV, y STAR 7 en dirección 3' del gen de d2EGFP (Fig. 33). La Fig. 34 muestra que, de los constructos de control sin elementos STAR, solamente el constructo GTG Zeo sin elementos STAR dio colonias que expresaron la proteína d2EGFP. Por el contrario, todos los constructos que contienen elementos STAR dieron colonias que expresaron proteína d2EGFP. La señal media de fluorescencia de d2EGFP de 11 colonias de

Control de GTG Zeo fue 20,3, de 13 colonias GTG Zeo con los STAR 7/67/7 104,9, de 24 colonias GTG space Zeo 7/67/7 201,5, de 6 colonias TTT GTG Zeo 7/67/7 310,5, de 22 colonias GTG Thr9 Zeo 7/67/7 423, y de 16 colonias GTG Phe9 Zeo 550,2 (Fig. 34).

5 Las mayores restricciones de las nuevas mutaciones GTG se correlacionan con mayores señales medias de fluorescencia (Fig. 34). Sin embargo, el TTT GTG Zeo 7/67/7 dio solamente dos colonias con nivel elevado de expresión, y unas pocas colonias con un nivel bajo de expresión. Esto puede indicar que esta mutación está al borde de la restricción que estas células pueden soportar con una concentración fija de zeocina añadida al medio de cultivo.

10 Las mutaciones Thr9 y Phe9 no influyen en la eficacia de la traducción de los mutantes de Zeo. En su lugar, reducen la funcionalidad de la proteína de resistencia a zeocina, evitando una interacción óptima entre las dos mitades de la proteína de resistencia a zeocina (Dumas et al, 1994). Esto implica que se ha de producir más proteína para lograr resistencia frente a la zeocina en el medio de cultivo. Como consecuencia, todo el casete se ha de transcribir a un mayor nivel, dando como resultado eventualmente un mayor nivel de expresión de d2EGFP.

15 Se concluye que el uso de las eficacias de traducción descritas del ARNm de resistencia a zeocina da como resultado mayores niveles de expresión de la proteína d2EGFP, esto en combinación con elementos STAR.

20 Este ejemplo demuestra además la posibilidad de proporcionar un ajuste fino de la restricción del sistema de selección de la invención para lograr niveles óptimos de expresión de una proteína de interés. Claramente, la persona experta en la técnica será capaz de combinar estas y otras posibilidades dentro de los conceptos descritos aquí (por ejemplo mutar la zeocina en la posición 9 a otros aminoácidos, o mutarla en otras posiciones; usar un GTG u otro codón de inicio en un contexto de iniciación de la traducción no óptimo para zeocina u otros marcadores de selección; o mutar otros marcadores de selección para reducir su funcionalidad, por ejemplo usar una secuencia que codifica un gen de resistencia a neomicina que tiene una mutación en el resto de aminoácidos 182 o 261, o ambos, véase por ejemplo el documento WO 01/32901), y similar, para proporcionar tal ajuste fino, y, simplemente ensayando, determinar una combinación adecuada de rasgos para el marcador de selección, conduciendo a una expresión mejorada del polipéptido de interés.

Ejemplo 15: Creación y ensayo de constructos bicistrónicos TTG Zeocina-d2EGFP adicionales con eficacias de traducción diferenciales.

30 En el Ejemplo 8 se describieron diferentes versiones del gen de resistencia a zeocina con codones de inicio mutados. Además de los codones TTG descritos (Fig. 35A), son posibles codones de inicio modificados adicionales, con distinta eficacia traduccional. Estas diferentes versiones del gen de resistencia a zeocina se crearon y se clonaron en dirección 5' del gen de d2EGFP modificado (Fig. 35).

Materiales y métodos

Creación de plásmidos

Se obtuvieron tres constructos TTG adicionales:

35 1) TTG como un codón de inicio en el gen de resistencia a Zeo (Fig. 35A), pero seguido de una secuencia espaciadora (Fig. 35B). El marco de lectura abierto de Zeo (con la secuencia espaciadora) se amplificó con el par de cebadores TTGspaceBamHI (SEQ. ID. NO. 126):

GAATTCGGATCCACCT**TTG**GCGATCCAAAGACTGCCAAATCTAG y

40 ZEOWTreverse (SEQ. ID. NO. 85), el producto de la PCR se cortó con EcoRI-BamHI, y se ligó en pd2EGFP, cortado con EcoRI-BamHI, creando pZEO-**TTGspace**-d2EGFP.

2) TTG como codón de inicio en el gen de resistencia a Zeo en lugar de ATG (Fig. 35A), pero con una mutación adicional en el marco de lectura abierto de Zeo en Pro9, que se sustituyó por treonina (Thr) (Fig. 35C). La mutación Thr9 se introdujo amplificando el marco de lectura abierto de Zeo con el par de cebadores ZEOForwardTTG-Thr9 (SEQ. ID. NO. 127):

45 AATTGGATCCACCT**TTG**GCCAAGTTGACCAGTGCCGTT**ACCG**TGCTC

y ZEOWTreverse (SEQ. ID. NO. 85), el producto de la PCR se cortó con EcoRI-BamHI, y se ligó en pd2EGFP, cortado con EcoRI-BamHI, creando pZEO-**TTG**-Thr9-d2EGFP.

50 3) TTG como un codón de inicio en el gen de resistencia a Zeo, en lugar de ATG (Fig. 35A), pero con una mutación adicional en el marco de lectura abierto de Zeo en Pro9, que se sustituyó por fenilalanina (Phe) (Fig. 35D). La mutación Phe9 se introdujo amplificando el marco de lectura abierto de Zeo con el par de cebadores ZEOForwardTTG-Phe9 (SEQ. ID. NO. 128):

AATTGGATCCACCT**TTG**GCCAAGTTGACCAGTGCCGTT**TTCT**TGCTC

y ZEOwTreverse (SEQ. ID. NO. 85), el producto de la PCR se cortó con EcoRI-BamHI, y se ligó en pd2EGFP, cortado con EcoRI-BamHI, creando pZEO-**TTG**-Phe9-d2EGFP.

Resultados

5 Se transfectaron células CHO-K1 con constructos que contienen los genes TTG Zeo (Fig. 35A), TTGspace Zeo (Fig. 35B), TTG Thr9 Zeo (Fig. 35C) y TTG Phe9 Zeo (Fig. 35D) como gen de selección, clonándose todos ellos en dirección 5' del gen informador de d2EGFP. Estos cuatro constructos carecían de elementos STAR (Control), o tenían elementos STAR 7 y 67 en dirección 5' del promotor de CMV, y STAR 7 en dirección 3' del gen de d2EGFP (Fig. 35). La Fig. 36 muestra que, de los constructos de control sin elementos STAR, solamente el constructo TTG Zeo sin elementos STAR dio colonias que expresaron la proteína d2EGFP. Por el contrario, todos los constructos
10 que contienen elementos STAR dieron colonias que expresaron la proteína d2EGFP. La señal media de fluorescencia de d2EGFP de 3 colonias de Control de TTG Zeo fue 26,8, de 24 colonias TTG Zeo con los STARS 7/67/7 426,8, de 24 colonias TTGspace Zeo 7/67/7 595,7, de 2 colonias TTG Thr9 Zeo 7/67/7 712,1, y de 3 colonias TTG Phe9 Zeo 677,1 (Fig. 36).

15 Las mayores restricciones de las nuevas mutaciones TTG se correlacionan con mayores señales medias de fluorescencia (Fig. 36). Sin embargo, los constructos TTG Thr9 Zeo 7/67/7 y TTG Phe9 Zeo 7/67/7 solamente dieron dos colonias de expresión elevada cada una, y unas pocas colonias de baja expresión. Esto puede indicar que estas mutaciones están en el borde de la restricción que las células pueden soportar con una concentración fija de zeocina añadida al medio de cultivo.

20 Se concluye que el uso de las eficacias de traducción descritas del ARNm de resistencia a zeocina da como resultado mayores niveles de expresión de la proteína d2EGFP, esto en combinación con elementos STAR.

Ejemplo 16: Creación y ensayo de constructos bicistrónicos de puromicina-d2EGFP con eficacias de traducción diferenciales.

25 Hay tres ATGs internos en el gen de resistencia a puromicina, cada uno de los cuales codifica una metionina (Fig. 28, Fig. 37A). Estos ATGs se han de eliminar (Fig. 37B, C), puesto que servirán como codón de inicio cuando se ha modificado el codón de inicio ATG (o su contexto), y esto dará como resultado péptidos que no se asemejan a la proteína de resistencia a puromicina. De forma más importante, estos ATGs evitarán la traducción eficiente del gen de interés, como se representa mediante d2EGFP en este ejemplo con fines ilustrativos. Las metioninas se cambiaron por leucina, como en la proteína de resistencia a zeocina (ejemplo 8). Sin embargo, en lugar de usar el codón TTG para leucina (por ejemplo en zeocina en el ejemplo 8), ahora se escogió el codón CTG para leucina
30 (en seres humanos, para leucina, el codón CTG se usó más a menudo que el codón TTG). Para eliminar los ATGs internos, se amplificó en primer lugar el marco de lectura abierto de la proteína de resistencia a puromicina con 4 pares de cebadores, generando 4 fragmentos de proteína de resistencia a puromicina. Los pares de cebadores fueron:

PURO BamHI F (SEQ. ID. NO. 129):

35 GATCGGATCCATGGTTACCGAGTACAAGCCCACGGT,

PURO300 R LEU (SEQ. ID. NO. 130):

CAGCCGGGAACCGCTCAACTCGGCCAGGCGCGGGC; y

PURO300FLEU (SEQ. ID. NO. 131):

40 CGAGTTGAGCGGTTCCCGGCTGGCCGCGCAGCAACAGCTGGAAGGCCTC,

PURO600RLEU (SEQ. ID. NO. 132):

AAGCTTGAATTCAGGCACCGGGCTTGCGGGTCAGGCACCAGGTC.

45 Esto genera dos productos de PCR, que corresponden a la parte 5' y 3' del gen de resistencia a puromicina. Los dos productos se añadieron juntos y se amplificaron con PURO BamHI F (SEQ. ID. NO. 129) - PURO600RLEU (SEQ. ID. NO. 132). El producto de PCR resultante se cortó con BamHI-EcoRI y se ligó, creando pCMV- ATGPURO (leu). La secuenciación de este clon verificó que se habían convertido los tres ATGs internos. Todo el marco de lectura abierto de puromicina se amplificó entonces con PUROBamHI TTG IF (SEQ. ID. NO. 133):

GAATTCGGATCCACCTTGGTTACCGAGTACAAGCCCACGGTG y

50 PURO600RLEU (SEQ. ID. NO. 132). Este cebador introduce un codón extra (GTT) directamente después del codón de inicio TTG, debido a que se introduce el "G" en el nucleótido +4 para un contexto óptimo, y por tanto se introducen dos nucleótidos más para conservar el marco de lectura.

Resultados

Se transfectaron células CHO-K1 con el constructo que contiene el gen TTG Puro (Fig. 38) como gen de selección, clonado en dirección 5' del gen informador d2EGFP. La selección se realizó en 10 µg/ml de puromicina. El constructo carecía de elementos STAR (Control), o poseía elementos STAR 7 y 67 en dirección 5' del promotor de CMV, y STAR 7 en dirección 3' del gen de d2EGFP (Fig. 38). La Fig. 38 muestra que la señal media de fluorescencia de d2EGFP de 24 colonias de control de TTG Puro fue 37,9, de 24 colonias TTG Puro con STARS 7/67/7 75,5. Además, cuando se toma la media de los cinco valores más elevados, la señal de fluorescencia de d2EGFP de colonias de Control de TTG Puro fue 69,5, y de colonias TTG Puro con STARS 7/67/7 186,1, un incremento de casi tres veces en la señal de fluorescencia de d2EGFP. Esto muestra que la eficacia de traducción modificada, descrita, del ARNm de resistencia a puromicina da como resultado mayores niveles de expresión de la proteína d2EGFP, esto en combinación con elementos STAR.

Este experimento demuestra que el gen de resistencia a puromicina se puede mutar para eliminar las secuencias ATG de él, a la vez que sigue siendo funcional. Además, se concluye que el método de selección de la invención también funciona con todavía otro marcador de selección, puromicina.

Ejemplo 17: Creación y ensayo de constructos de neomicina con eficacias de traducción diferenciales.

Hay dieciséis ATGs internos en el gen de resistencia a neomicina, cinco de los cuales codifican una metionina en el marco de lectura abierto de neomicina (Fig. 31, Fig. 39A). Estos dieciséis ATGs se han de eliminar (Fig. 39B, C), puesto que servirán como codón de inicio cuando se ha modificado el codón de inicio ATG (o su contexto), y esto dará como resultado péptidos que no se asemejan a la proteína de resistencia a neomicina, y esto disminuirá la traducción desde el marco de lectura abierto en dirección 3' que codifica el péptido de interés en las unidades de transcripción de la invención. Para eliminar los ATGs internos, el marco de lectura abierto de la proteína de resistencia a neomicina se sintetizó completamente por un proveedor comercial (GeneArt, Alemania), en el que todos los ATGs codificantes internos (para Met) se sustituyeron por CTGs (que codifican Leu), y los ATGs no codificantes se sustituyeron de manera que se usó un codón degenerado y por tanto no se produjeron mutaciones en la secuencia proteica; la secuencia sintetizada de la neomicina se da en SEQ. ID. NO. 134. A fin de sustituir el codón de inicio ATG por GTG (Fig. 39B) o TTG (Fig. 39C), se amplificó el gen de neomicina sintetizado con pares de cebadores NEO-F-HindIII (SEQ. ID. NO. 136):

GATCAAGCTTTTGGATCGGCCATTGAAACAAGACGGATTG y

NEO EcoRI 800R (SEQ. ID. NO. 137): AAGCTTGAATTCTCAGAAGAAGACTCGTCAAGAAGGCG.

Resultados

Se usaron bacterias de *E. coli* para ensayar la funcionalidad de la proteína de resistencia a neomicina de la que se eliminaron todos los ATGs. Las bacterias de *E. coli* se transformaron con los constructos que contienen el gen GTG Neo (Fig. 39B) o TTG Neo (Fig. 39C) como gen de selección. La selección tuvo lugar haciendo crecer las bacterias en kanamicina. Sólo un gen de resistencia a neomicina funcional puede dar resistencia frente a kanamicina. La transformación con cualquier gen Neo modificado dio como resultado la formación de colonias de *E. coli*, a partir de las cuales se pudo aislar el plásmido que contiene el gen. Esto muestra que las eficacias de traducción modificadas, descritas, de los ARNm de resistencia a Neomicina, así como la eliminación de todos los ATGs del marco de lectura abierto de Neo, da como resultado la producción de proteína de resistencia a neomicina funcional.

Los genes de resistencia a neomicina mutados se incorporan en una unidad de transcripción multicistrónica de la invención, y se usan para la selección con G418 o neomicina en células hospedantes eucariotas.

Ejemplo 18: Creación y ensayo de constructos de dhfr con eficacias de traducción diferenciales.

Hay ocho ATGs internos en el gen de dhfr, seis de los cuales codifican una metionina en el marco de lectura abierto de dhfr (Fig. 29, Fig. 40A). Todos estos ATGs se han de eliminar (Fig. 40B, C), puesto que servirán como codón de inicio cuando se ha modificado el codón de inicio ATG (o su contexto), y esto dará como resultado péptidos que no se asemejan a la proteína dhfr, y disminuirá la traducción a partir del marco de lectura abierto en dirección 3' que codifica el polipéptido de interés en las unidades de transcripción de la invención. Para eliminar los ATGs internos, se sintetizó completamente el marco de lectura abierto de la proteína dhfr (SEQ. ID. NO. 138), como se describe anteriormente para neomicina. A fin de sustituir el codón de inicio ATG por GTG (Fig. 40B) o TTG (Fig. 40C), se amplificó el gen de DHFR sintetizado con los cebadores DHFR-F-HindIII (SEQ. ID. NO. 140): GATCAAGCTTTTGTTCGACCATTGAACTGCATCGTC y DHFR-EcoRI-600-R (SEQ. ID. NO. 141): AGCTTGAATTCTTAGTCTTTCTTCTCGTAGACTTC.

Resultados

Se usaron bacterias de *E. coli* para ensayar la funcionalidad de la proteína dhfr, de la que se eliminaron todos los ATGs. *E. coli* se transformó con los constructos que contienen el gen GTG dhfr (Fig. 40B) o TTG dhfr (Fig. 40C). La selección tuvo lugar haciendo crecer las bacterias en trimetoprima (Sigma T7883-56). Sólo un gen de dhfr funcional puede dar resistencia frente a trimetoprima. La transformación con cualquier gen de dhfr modificado dio como resultado la formación de colonias de *E. coli*, a partir de las cuales se pudo aislar el plásmido que contiene el gen.

Esto muestra que las eficacias de traducción modificadas, descritas, de los ARNm de dhfr, así como la eliminación de todos los ATGs del marco de lectura abierto de dhfr, dan como resultado la producción de proteína dhfr funcional.

Los genes de dhfr mutados se incorporan en una unidad de transcripción multicistónica de la invención, y se usan para la selección con metotrexato en células hospedantes eucariotas.

5 **Ejemplo 20: Ensayo de constructos de zeocina y blastidina con eficacias de traducción diferenciales en células PER.C6.**

Se ensayaron en la línea celular PER.C6 diversos genes de zeocina y de blastidina con codones de inicio mutados, todos clonados en dirección 5' del gen de d2EGFP.

Resultados

10 Las modificaciones del gen de resistencia GTG Zeocina y GTGspace Zeocina (véase también el Ejemplo 14; Fig. 41) y las modificaciones del gen de resistencia GTG Blastidina y TTG Blastidina (véase también el Ejemplo 11; Fig. 42), todos clonados en dirección 5' del gen d2EGFP, se transfirieron a células PER.C6. Como se muestra en la Fig. 41, la transfección con el gen tanto de GTG Zeocina como de GTGspace Zeocina dio como resultado colonias que expresaron d2EGFP. La señal media de fluorescencia de d2EGFP de 20 colonias GTG Zeo fue 63,8, mientras que
15 la señal media de d2EGFP de 20 colonias GTGspace Zeo fue 185, demostrando que también en células PER.C6 el GTGspace Zeo tiene una restricción de traducción mayor que el ARNm de GTG Zeo.

20 Como se muestra en la Fig. 42, la transfección con el gen tanto GTG Blastidina como TTG Blastidina dio como resultado colonias que expresaron d2EGFP. La señal media de fluorescencia de d2EGFP de 20 colonias GTG Blastidina fue 71,4, mientras que la señal media de d2EGFP de 20 colonias TTG Blastidina fue 135, demostrando que también en células PER.C6 el TTG Blastidina tiene una restricción de traducción mayor que el ARNm de GTG Blastidina.

Este ejemplo demuestra que el sistema de selección de la invención también se puede usar en otras células distintas de las células CHO.

25 **Ejemplo 21: Ensayo de la adición de una señal de pausa transcripcional a un constructo TTG Zeocina-d2EGFP.**

Se piensa que una secuencia de Pausa de la TRANscripción (TRAP) evita, al menos en parte, la formación de ARN antisentido, o, al menos en parte, evita que entre la transcripción en dicha unidad de expresión proteica (véase el documento WO 2004/055215). Una secuencia TRAP se define funcionalmente como una secuencia que, cuando se
30 coloca en una unidad de transcripción, da como resultado un nivel reducido de transcripción del ácido nucleico presente en el lado 3' de la TRAP, cuando se compara con el nivel de transcripción observado en el ácido nucleico en el lado 5' de la TRAP, y los ejemplos no limitantes de secuencias TRAP son las señales de terminación de la transcripción. A fin de que funcione para evitar o disminuir que la transcripción entre en la unidad de transcripción, la TRAP se ha de colocar en dirección 5' de un promotor que dirige la expresión de la unidad de transcripción, y la TRAP debería estar en una dirección 5' a 3'. A fin de evitar al menos en parte la formación de ARN antisentido, la
35 TRAP debería estar localizada en dirección 3' del marco de lectura abierto en una unidad de transcripción, y presente en una dirección 3' a 5' (esto es, en una orientación opuesta a la de la orientación normal de una secuencia de terminación de la transcripción que está presente habitualmente detrás del marco de lectura abierto en una unidad de transcripción). Se prefiere una combinación de una TRAP en dirección 5' del promotor en una orientación 5' a 3' y una TRAP en dirección 3' del marco de lectura abierto en una orientación 3' a 5'. La adición de una
40 secuencia TRAP a un elemento STAR mejora los efectos de los elementos STAR sobre la expresión transgénica (véase el documento WO 2004/055215). Aquí, se ensayaron los efectos de la secuencia TRAP en el contexto del gen de resistencia a TTG Zeo.

Resultados

45 El casete TTG Zeocina-d2EGFP, que estaba flanqueado con elementos STAR7 (Fig. 43), se modificó mediante adición de la secuencia TRAP SPA/pause (véase el documento WO 2004/055215); SEQ. ID. NO. 142), ambos en dirección 5' del 5' STAR7 (en dirección 5' a 3') y en dirección 3' del 3' STAR 7 (en dirección 3' a 5') (Fig. 43). Los vectores que contienen tanto STAR 7/7 como TRAP-STAR 7/7-TRAP se transfirieron a CHO-K1. Se aislaron colonias estables, y se midieron las intensidades de fluorescencia de d2EGFP. Como se muestra en la Fig. 43, la
50 señal media de fluorescencia de d2EGFP de 23 colonias TTG Zeo STAR 7/7 fue 455,1, mientras que la señal media de fluorescencia de d2EGFP de 23 colonias TTG Zeo TRAP-STAR 7/7-TRAP fue 642,3. La señal media de fluorescencia de d2EGFP en las 5 colonias TTG Zeo STAR 7/7 más elevadas fue 705,1, mientras que la señal media de fluorescencia de d2EGFP de 5 colonias TTG Zeo TRAP-STAR 7/7-TRAP fue 784,7.

55 Este resultado indica que la adición de las TRAPs no potencia la señal de fluorescencia de d2EGFP en las colonias más elevadas, sino que hay una elevación significativa en el número de colonias de expresión elevada. Mientras que sólo 5 colonias TTG Zeo STAR 7/7 tuvieron una señal de d2EGFP por encima de 600, 17 colonias TTG Zeo TRAP-STAR 7/7-TRAP tuvieron una señal de fluorescencia de d2EGFP por encima de 600.

En el experimento, se transfectaron 3 μg de ADN de cada plásmido. Sin embargo, mientras que la eficacia de la transfección fue similar, el número total de colonias con el plásmido TTG Zeo STAR 7/7 fue 62, mientras que el número total de colonias con el plásmido TTG Zeo TRAP-STAR 7/7-TRAP fue 116, casi el doble.

5 Se concluye que la adición de elementos TRAP a los plásmidos que contienen STAR con codones de traducción del gen de resistencia a zeocina modificado da como resultado un número global de colonias significativamente mayor, y que hay más colonias con los niveles de expresión más elevados.

Ejemplo 22: Dependencia de la expresión con el número de copias.

Se analizaron los niveles de expresión del anticuerpo EpCAM con relación al número de copias de ADN de EpCAM integrado.

10 Resultados

El constructo que se ensayó fue TTG-Zeo-Cadena Ligera (LC)-TTG-Blas-Cadena Pesada (HC), estando ambas unidades de expresión bajo el control de promotor de CMV (véase la Fig. 44). Este constructo contenía STAR 7 y 67 (véase la Fig. 44). Las condiciones de selección fueron tales que con 200 $\mu\text{g}/\text{ml}$ de zeocina y 20 $\mu\text{g}/\text{ml}$ de blasticidina en el medio de cultivo no sobrevivieron colonias de control (no STARs), y sólo sobrevivieron colonias STAR 7/67/7.

15 Se aisló ADN cuando las colonias tuvieron 60 días bajo presión de selección con zeocina y blasticidina (véase la Fig. 44). Se calculó el valor de R^2 , y se mostró. En todo el intervalo de 5 a 40 $\text{pg}/\text{célula}/\text{día}$ de EpCAM, hubo un grado elevado de dependencia del número de copias, como se señala mediante un R^2 relativamente elevado de 0,5978 (Fig. 44). Los datos muestran que, en el nuevo sistema de selección, en colonias que contienen constructos TTG Zeo-TTG Blas EpCAM STAR 7/67/7, hay una expresión de EpCAM dependiente del número de copias.

Ejemplo 23: Inducción de una mayor expresión de EpCAM debido a metotrexato

25 Se analizaron los niveles de expresión del anticuerpo EpCAM tras incubar los clones con metotrexato (MTX). El objetivo de este experimento fue determinar si la amplificación de un constructo que contiene STAR daría como resultado una mayor expresión de EpCAM. MTX actúa mediante la inhibición del producto génico de dhfr. Aunque se han descrito algunas cepas de CHO que son deficientes en dhfr, CHO-K1 es dhfr⁺. Por lo tanto, deben existir concentraciones relativamente elevadas de MTX en el medio de cultivo para seleccionar la amplificación mediante concentraciones aumentadas de MTX en células CHO-K1.

Resultados

30 El constructo que se ensayó fue TTG-Zeo-Cadena Pesada (HC)-TTG-Blas-Cadena Ligera (LC), estando ambas unidades de expresión bajo el control del promotor de CMV. En dirección 5' de cada promotor de CMV, se colocó STAR67, y se usó STAR7 para flanquear todo el casete (véase también el Ejemplo 13, Fig. 22 para tal constructo). Este constructo se modificó adicionalmente colocando un casete SV40-dhfr (un gen de dhfr de ratón bajo el control de un promotor de SV40) entre los casetes HC y LC, en dirección 5' del segundo STAR67 (Fig. 45). Se transfectaron células CHO-K1. La selección se realizó con 100 $\mu\text{g}/\text{ml}$ de zeocina y 10 $\mu\text{g}/\text{ml}$ de blasticidina en el medio de cultivo. Ninguna de las colonias de control (sin elementos STAR) sobrevivieron, y sólo sobrevivieron las colonias con constructos que contienen los STAR. Las colonias se aislaron y se propagaron antes de medir los niveles de expresión de EpCAM. Seis colonias que produjeron entre 20 y 35 $\text{pg}/\text{célula}/\text{día}$ se transfirieron a medio que contiene 100 nM de MTX. Esta concentración se elevó hasta 500 nM, 1000 nM y finalmente hasta 2000 nM, con períodos de dos semanas entre cada etapa. Después de dos semanas en 2000 nM de MTX, se midieron las concentraciones de EpCAM. Como se muestra en la Fig. 45, cuatro colonias mostraron una producción aumentada de EpCAM. Colonia 13: de 22 a 30; colonia 14: de 28 a 42; colonia 17: de 20 a 67; y colonia 19: de 37 a 67 $\text{pg}/\text{célula}/\text{día}$. Las colonias 4 y 16 no mostraron expresión potenciada de EpCAM. Se concluye que la adición de metotrexato al medio de cultivo de colonias CHO-K1 creadas con el sistema de selección de la invención puede dar como resultado una expresión potenciada de la proteína. Por tanto, los elementos STAR y el método de selección de la invención se pueden combinar con y son compatibles con la potenciación de los niveles de expresión proteica inducida por MTX.

Ejemplo 24: La selección TTG-Zeo opera en el contexto de diferentes promotores

Se analizaron los niveles de expresión de d2EGFP en el contexto del marcador de selección TTG Zeo y diferentes promotores. Se comparó la acción de elementos STAR en el contexto del potenciador/promotor de CMV, el potenciador/promotor de SV40 y el potenciador de CMV/promotor de β -actina.

Resultados

50 En la Fig. 46 se indican los promotores que se ensayaron en el contexto del marcador de selección TTG-Zeo. Los plásmidos ensayados consistieron en los constructos de control indicados con tres promotores diferentes y los constructos STAR que estuvieron flanqueados con STAR 7 y STAR 67 en el extremo 5' y STAR 7 en el extremo 3'. Los constructos se transfectaron a célula CHO-K1, y la selección se llevó a cabo con 200 $\mu\text{g}/\text{ml}$ en el medio de cultivo. Se aislaron y propagaron hasta 23 colonias independientes antes del análisis de los niveles de expresión de

d2EGFP. Como se muestra en la Fig. 46, la incorporación de elementos STAR en constructos con el potenciador/promotor de CMV, el potenciador/promotor de SV40 o el potenciador de CMV/promotor de β -actina dio como resultado en todos ellos la formación de colonias con niveles de expresión de d2EGFP mayores que con los constructos de control correspondientes. Esto muestra que el sistema de selección de la invención, en combinación con elementos STAR, opera bien en el contexto de diferentes promotores. Un análisis posterior demostró que la media de los valores de d2EGFP dirigida por CMV fue significativamente mayor que la media de los valores de d2EGFP dirigida por SV40 ($p < 0,05$). Por el contrario, la media de los valores de d2EGFP dirigida por CMV no fue significativamente diferente de los valores de d2EGFP dirigida por CMV/ β -actina ($p = 0,2$).

Ejemplo 25: Comparación de diferentes elementos STAR en el sistema de selección TTG-Zeo.

Se analizaron los niveles de expresión de d2EGFP en el contexto del marcador de selección TTG Zeo con el promotor de CMV y 53 elementos STAR diferentes, para obtener una comprensión más profunda de qué elementos STAR dan los mejores resultados en este contexto.

Resultados

Se clonaron 53 elementos STAR en dirección 5' y en dirección 3' del casete promotor de CMV-TTG Zeo-d2EGFP. Los siguientes elementos STAR se ensayaron en tales constructos: STAR2-12, 14, 15, 17-20, 26-34, 36, 37, 39, 40, 42-49, 51, 52, 54, 55, 57-62, 64, 65, 67. Los constructos se transfectaron a células CHO-K1, y la selección se llevó a cabo con 200 μ g/ml de zeocina en el medio de cultivo. Se aislaron y se propagaron hasta 24 colonias independientes antes del análisis de los niveles de expresión de d2EGFP. La incorporación de elementos STAR en los constructos dio como resultado diferentes grados de expresión potenciada de d2EGFP, en comparación con el control. La incorporación de elementos STAR 14, 18 y 55 en este experimento no dio como resultado un incremento de la expresión media de d2EGFP con respecto al control (sin elemento STAR). Aunque algunos constructos (con elementos STAR 2, 3, 10, 42, 48 y 49) en este experimento dieron lugar a solamente unas pocas colonias, todos los elementos STAR ensayados, excepto 14, 18 y 55, dieron como resultado niveles medios de expresión de d2EGFP mayores que para el control. Se debería señalar que algunos elementos STAR pueden actuar de una manera más específica del tipo celular, y que es muy posible que STAR 14, 18 y 55 funcionen mejor en otros tipos celulares, con otros promotores, otros marcadores de selección, o en diferente contexto o configuración que en el conjunto particular de condiciones ensayadas aquí. La adición de 10 elementos STAR, a saber, elementos STAR 7, 9, 17, 27, 29, 43, 44, 45, 47 y 61, indujo niveles medios de expresión de d2EGFP mayores de 5 veces el nivel medio de expresión de d2EGFP del control. Se retransformó el control y los 7 constructos con elementos STAR, y se repitió el experimento. Los resultados se muestran en la Fig. 47. La incorporación de elementos STAR en los constructos dio como resultado diferentes grados de expresión potenciada de d2EGFP, según se compara con el control (Fig. 47). El nivel medio de expresión de d2EGFP en colonias transfectadas con el constructo de control fue 29. Las medias de los niveles de expresión de d2EGFP en colonias con los 7 constructos STAR diferentes osciló entre 151 (STAR 67) y 297 (STAR 29). Esto es un factor de 5 a 10 veces mayor que la media en las colonias de control.

Se concluye que a) la inmensa mayoría de elementos STAR tiene un efecto positivo sobre los niveles de expresión génica, b) hay una variación en el grado de efectos positivos inducidos por los diferentes elementos STAR, y c) 10 de 53 elementos STAR ensayados indujeron niveles medios de expresión de d2EGFP de más de 5 veces, en comparación con el control, y que los elementos STAR pueden inducir un nivel medio de expresión de d2EGFP de 10 veces mayor, en comparación con el control.

Ejemplo 26: Otros elementos de control de cromatina en el contexto de un sistema de selección de la invención

Se ha dado a conocer que los elementos de ADN tales como el sitio hipersensible HS4 en la región de control del locus del gen de β -globina de pollo (Chung et al, 1997), las regiones de unión de la matriz (MAR) (Stief et al, 1989) y un elemento de apertura de cromatina ubicuo (UCOE) (Williams et al, 2005) tienen efectos beneficiosos sobre la expresión génica cuando se incorporan en un vector estos elementos de ADN. Se combinaron estos elementos de ADN con el sistema de selección de la invención.

Resultados

El elemento HS4 de 1,25 kb se clonó en el casete que engloba el promotor de CMV, TTG Zeo y d2EGFP mediante una etapa de ligación de tres vías, para obtener un constructo con un tándem de 2 elementos HS4 (Chung et al, 1997). Esta etapa se realizó tanto para el 5' como para el 3' del casete que engloba el promotor de CMV, TTG Zeo y d2EGFP. La MAR de lisozima de pollo de una longitud de 2959 pb (Stief et al, 1989) se clonó 5' y 3' del casete que engloba el promotor de CMV, TTG Zeo y d2EGFP.

El UCOE de 2614 pb de longitud (Williams et al, 2005) fue un fragmento NotI-KpnI, cortado de un clon de BAC humano (RP11-93D5), que corresponde al nucleótido 29449 a 32063. Este fragmento se clonó 5' del promotor de CMV.

El constructo STAR contenía STAR7 y STAR67 5' del promotor de CMV, y STAR7 3' del casete. Estos cuatro constructos, así como el constructo de control sin elementos de ADN de control de cromatina flanqueantes, se

transfectaron a células CHO-K1. La selección se llevó a cabo mediante 200 µg/ml de zeocina en el medio de cultivo. Las colonias se aislaron, se propagaron, y se midieron los niveles de expresión de d2EGFP. Como se muestra en la Fig. 48, los constructos con todos los elementos de ADN dieron como resultado la formación de colonias que expresan d2EGFP. Sin embargo, la incorporación de 2x elementos HS4 y el UCOE no dio como resultado la formación de colonias que presentaron mayores niveles de expresión de d2EGFP, en comparación con las colonias de control. Por el contrario, la incorporación de la MAR de lisozima dio como resultado la formación de colonias que expresaron d2EGFP de forma significativamente mayor. El nivel medio de expresión inducido por constructos que contienen MAR fue cuatro veces mayor que en las colonias de control. Sin embargo, los mejores resultados se obtuvieron incorporando STAR 7 y 67 en el constructo. Se observó un incremento de casi diez veces en el nivel medio de expresión de d2EGFP, en comparación con las colonias de control. Se concluye que se pueden usar otros elementos de ADN de control de cromatina, tales como MARs en el contexto del sistema de selección de la invención. Sin embargo, los mejores resultados se obtuvieron cuando se usaron elementos STAR como elementos de control de cromatina.

DESCRIPCIÓN DE LAS FIGURAS

- 15 Fig. 1. Diagrama esquemático de unidades de transcripción sin y con elementos STAR.
- A) Gen bicistrónico que contiene (de 5' a 3') un transgén (que codifica por ejemplo el gen de d2EGFP), un IRES, un marcador seleccionable (zeo, que confiere resistencia a zeocina) bajo el control del promotor de CMV. La unidad de expresión tiene el terminador transcripcional de SV40 en su extremo 3' (t). El nombre de constructo es CMV-d2EGFP-ires-Zeo (Control de CMV).
- 20 B) Constructo como en A, pero ahora STAR 67 se clonó en dirección 5' del promotor de CMV. El nombre del constructo es STAR67-CMV-d2EGFP-ires-Zeo (CMV-STAR67).
- C) Constructo como en B, pero los elementos STAR7 en dirección 5' y en dirección 3' se clonan para flanquear todo el constructo. El nombre de constructo es STAR7-STAR67-CMV-d2EGFP-ires-Zeo-STAR7 (CMV-STAR67 7/7).
- 25 Fig. 2. STAR67 mejora en la expresión de d2EGFP dirigida por CMV en células CHO. La media de la señal de d2EGFP para 10 colonias estables independientes se representa gráficamente para los constructos indicados en células CHO. A) Control de CMV; B) STAR67-CMV. X(10): niveles medios de expresión de d2EGFP de las 10 colonias. Véase el ejemplo 2 para detalles.
- 30 Fig. 3. STAR67 mejora la expresión de d2EGFP dirigida por EF1α en células CHO. Igual que Fig. 2. pero ahora con el promotor de EF1α. A) Control de EF1α; B) STAR67-EF1α.
- Fig. 4. STAR67 mejora la expresión de d2EGFP dirigida por UB6 en células CHO. Igual que Figs. 2 y 3, pero ahora con el promotor de UB6. A) Control de UB6; B) STAR67-UB6.
- Fig. 5. STAR67 mejora la expresión de d2EGFP dirigida por CMV en células PER.C6. Igual que Fig. 2, pero ahora en células PER.C6. A) Control de CMV; B) STAR67-CMV.
- 35 Fig. 6. STAR67 mejora la expresión de d2EGFP dirigida por EF1α en células PER.C6. Igual que Fig. 3, pero ahora en células PER.C6. A) Control de EF1α; B) STAR67-EF1α.
- Fig. 7. STAR67 mejora la expresión de d2EGFP dirigida por UB6 en células PER.C6. Igual que Fig. 4, pero ahora en células PER.C6. A) Control de UB6; B) STAR67-UB6.
- 40 Fig. 8. STAR67 mejora la expresión de d2EGFP dirigida por SV40 en células CHO-K1 en combinación con otro elemento STAR. Similar a Fig. 2, pero usando ahora el promotor de SV40 en células CHO, y los constructos indicados. La media de la señal de d2EGFP se representa gráficamente para 18-20 colonias estables e independientes, 60 días después de la transfección. A) Control de SV40, SV40-STAR67, SV40-STAR7/7; B) Control de SV40 y SV40-STAR67 7/7. Véase el ejemplo 4 para detalles.
- 45 Fig. 9. STAR67 mejora los niveles de expresión dirigida por UB6 del anticuerpo EpCAM en células CHO-K1. Véase el ejemplo 5 para detalles. A) constructos con elementos STAR; B) constructos con STAR67 en dirección 5' del promotor y elementos STAR7 flanqueantes. La concentración de anticuerpo anti-EpCAM se presenta como pg/célula/día. X(18): nivel de producción medio de las 18 colonias.
- Fig. 10. STAR67 no es un bloqueador del potenciador, mientras que STAR 6 y 7 sí lo son. Véase el ejemplo 6 para detalles.
- 50 Fig. 11. STAR67 potencia los niveles de expresión de anticuerpo dirigida por UB6 y CMV en células CHO establemente transfectadas. Véase el ejemplo 7 para detalles. A) molécula de ADN individual que contiene la cadena pesada (HC) y la cadena ligera (LC) anti-EpCAM, cada una detrás de un promotor, y cada una ligada a un gen marcador seleccionable diferente (se usó selección simultánea para ambos marcadores); constructo

sin elementos STAR. No se encontraron colonias; B) el mismo constructo con STAR67 entre los dos promotores. La concentración de anticuerpo anti-EpCAM se presenta como pg/célula/día. X(19): nivel de producción medio de las 19 colonias.

5 Fig. 12. Representación esquemática del uso de un gen marcador de selección (gen de resistencia a zeocina) según la invención. A. gen de resistencia a zeocina de tipo salvaje, que tiene su sitio de iniciación de la traducción normal (codón de inicio ATG) y un codón ATG interno, que codifica metionina. B. gen de resistencia a zeocina mutante, en el que el ATG interno se ha mutado en un codón para leucina; este mutante es un gen de resistencia a zeocina funcional. C. igual que B, pero que comprende un sitio de iniciación de la traducción mutado, en el que el contexto del codón de inicio ATG se ha mutado para disminuir la iniciación de la traducción. D. igual que B, pero que comprende un codón de inicio mutado (GTG). E. igual que B, pero con un codón de inicio TTG. Los números en las figuras C-E indican esquemáticamente una cantidad relativa de frecuencia de iniciación (bajo el codón de inicio) y la frecuencia "mediante barrido" (bajo la secuencia codificante) mediante los ribosomas, pero sólo de una manera semicuantitativa, es decir, indican la eficiencia de la iniciación de la traducción comparada entre sí, pero los números cualitativos pueden diferir completamente: los números sólo sirven para explicar la invención. Véase el ejemplo 8 para detalles.

10 Fig. 13. Representación esquemática de una unidad de transcripción multicistónica según la invención, con eficacia de la traducción interdependiente más o menos recíproca. La explicación es igual que para la Fig. 12, pero ahora se ha colocado un gen de dEGFP (que ejemplifica aquí un gen de interés) en dirección 3' de la secuencia codificante del polipéptido marcador seleccionable. El gen de resistencia a zeocina comprende la mutación interna Met->Leu (véase Fig. 12B). Véase el ejemplo 9 para detalles.

15 Fig. 14. Resultados de sistemas de selección según la invención, con y sin elementos STAR. A. gen de resistencia a zeocina con codón de inicio ATG en mal contexto (denominado "ATGmut" en el dibujo, pero que incluye una secuencia espaciadora detrás del ATG en el mal contexto, denominado generalmente así en el texto como "ATGmut/space"). B. gen de resistencia a zeocina con codón de inicio GTG. C. gen de resistencia a zeocina con codón de inicio TTG. La señal de d2EGFP para colonias independientes se muestra en el eje vertical. Véase el ejemplo 9 para detalles.

20 Fig. 15. Resultados del sistema de selección según la invención en experimento aumentado de escala (A), y comparación con el sistema de selección según la técnica anterior usando un IRES (B). La señal de d2EGFP para colonias independientes se muestra en el eje vertical. Véase el ejemplo 10 para detalles.

25 Fig. 16. Resultados del sistema de selección con unidad de transcripción multicistónica según la invención, usando como marcador seleccionable blasticidina. A. gen de resistencia a blasticidina mutado para comprender un codón de inicio GTG. B. gen de resistencia a blasticidina mutado para comprender un codón de inicio TTG. El gen de resistencia a blasticidina se había mutado adicionalmente para eliminar todas las secuencias ATG internas. La señal de d2EGFP para colonias independientes se muestra en el eje vertical. Véase el ejemplo 11 para detalles.

30 Fig. 17. Estabilidad de la expresión de varios clones con una unidad de transcripción multicistónica según la invención (incluye una zeocina con codón de inicio TTG). Había presión de selección (100 µg/ml de zeocina) durante todo el experimento. La señal de d2EGFP para colonias independientes se muestra en el eje vertical. Véase el ejemplo 12 para detalles.

35 Fig. 18. Igual que Fig. 17, pero la concentración de zeocina se redujo hasta 20 µg/ml después del establecimiento de los clones.

Fig. 19. Igual que Fig. 17, pero la zeocina estaba ausente del medio de cultivo después del establecimiento de los clones.

40 Fig. 20. Expresión de un anticuerpo (anti-EpCAM) usando el sistema de selección sin la unidad de transcripción multicistónica según la invención. La cadena pesada (HC) y la cadena ligera (LC) son en este ejemplo el polipéptido de interés. Cada una de estas está presente en una unidad de transcripción separada, que están ambas en una única molécula de ácido nucleico en este ejemplo. La HC está precedida por el gen de resistencia a zeocina que codifica un polipéptido marcador seleccionable, mientras que la LC está precedida por el gen de resistencia a blasticidina que codifica un polipéptido marcador seleccionable. Ambos genes de resistencia se han mutado para comprender un codón de inicio ATG en un contexto no óptimo ("mutATG" en la Figura, pero incluyendo una secuencia espaciadora, y por tanto referido generalmente en el texto como "ATGmut/space"). Cada una de las unidades de transcripción multicistónicas está bajo el control de un promotor de CMV. Los constructos con secuencias STAR como se indican, se compararon con los constructos sin secuencias STAR. Los niveles de anticuerpo obtenidos cuando estos constructos se introdujeron en células hospedantes se dan en el eje vertical en pg/célula/día para diversos clones independientes. Véase el ejemplo 13 para detalles.

45 Fig. 21. Igual que Fig. 20, pero ambos genes marcadores de selección se han proporcionado con un codón de inicio GTG. Véase el ejemplo 13 para detalles.

Fig. 22. Igual que Fig. 20, pero ambos genes marcadores de selección se han proporcionado con un codón de inicio TTG. Véase el ejemplo 13 para detalles.

5

Fig. 23. Estabilidad de la expresión en subclones en ausencia de presión de selección (después de establecer colonias bajo presión de selección, algunas colonias se subclonaron en medio que no contiene zeocina). Véase el ejemplo 12 para detalles.

Fig. 24. Dependencia de los niveles de expresión con el número de copias de una realización de la invención. Véase el ejemplo 12 para detalles.

10

Fig. 25. Igual que Fig. 12, pero para el gen de resistencia a blasticidina. Ninguno de los 4 ATG internos en este gen están en marco que codifica una metionina, y por lo tanto se usó la redundancia del código genético para mutar estos ATG sin mutar la secuencia de aminoácidos interna de la proteína codificada.

Fig. 26. Secuencia codificante del gen de resistencia a zeocina de tipo salvaje (SEQ. ID. NO. 108). Los ATG en negrita codifican metionina. El primer ATG en negrita es el codón de inicio.

15

Fig. 27. Secuencia codificante del gen de resistencia a blasticidina de tipo salvaje (SEQ. ID. NO. 110). Los ATG en negrita codifican metionina. El primer ATG es el codón de inicio. Se subrayan otros ATG en la secuencia: estos ATG internos no codifican metionina, debido a que no están en el marco.

Fig. 28. Secuencia codificante del gen de resistencia a puromicina de tipo salvaje (SEQ. ID. NO. 112). Los ATG en negrita codifican metionina. El primer ATG en negrita es el codón de inicio.

20

Fig. 29. Secuencia codificante del gen de DHFR de ratón de tipo salvaje (SEQ. ID. NO. 114). Los ATG en negrita codifican metionina. El primer ATG en negrita es el codón de inicio. Se subrayan otros ATG en la secuencia: estos ATG internos no codifican metionina, debido a que no están en marco.

Fig. 30. Secuencia codificante del gen de resistencia a higromicina de tipo salvaje (SEQ. ID. NO. 116). Los ATG en negrita codifican metionina. El primer ATG es el codón de inicio. Se subrayan otros ATG en la secuencia: estos ATG internos no codifican metionina, debido a que no están en el marco.

25

Fig. 31. Secuencia codificante del gen de resistencia a neomicina de tipo salvaje (SEQ. ID. NO. 118). Los ATG en negrita codifican metionina. El primer ATG es el codón de inicio. Se subrayan otros ATG en la secuencia: estos ATG internos no codifican metionina, debido a que no están en el marco.

Fig. 32. Secuencia codificante del gen de glutamina sintasa (GS) humano de tipo salvaje (SEQ. ID. NO. 120). Los ATG en negrita codifican metionina. El primer ATG es el codón de inicio. Se subrayan otros ATG en la secuencia: estos ATG internos no codifican metionina, debido a que no están en el marco.

30

Fig. 33. Representación esquemática de algunos genes marcadores de selección de resistencia a zeocina modificados con un codón GTG según la invención, que permiten el ajuste fino adicional de la restricción de selección. Véase el ejemplo 14 para detalles.

35

Fig. 34. Resultados con sistemas de expresión que contienen los genes marcadores de selección de resistencia a zeocina modificados. Véase el ejemplo 14 para detalles. Los puntos indican puntos de datos individuales; las líneas indican los niveles medios de expresión; los constructos usados (véase también Fig. 33) se indican en el eje horizontal (la adición de 7/67/7 al final del nombre del constructo indica la presencia de secuencias STAR 7 y 67 en dirección 5' del promotor y STAR7 en dirección 3' del sitio de terminación de la transcripción), y se representan esquemáticamente encima de la gráfica; el eje vertical indica la señal de d2EGFP.

40

Fig. 35. Representación esquemática de algunos genes marcadores de selección de resistencia a zeocina modificados adicionales con un codón de inicio TTG según la invención, que permite el ajuste fino adicional de la restricción de la selección. Véase el ejemplo 15 para detalles.

45

Fig. 36. Resultados con sistemas de expresión que contienen los genes marcadores de selección de resistencia a zeocina modificados adicionales. Véase el ejemplo 15 para detalles. Los puntos indican puntos de datos individuales; las líneas indican los niveles medios de expresión; los constructos usados se indican en el eje horizontal, y se representan esquemáticamente encima de la gráfica; el eje vertical indica la señal de d2EGFP.

50

Fig. 37. Igual que Fig. 12, pero para el gen de resistencia a puromicina. Los tres ATG internos codifican metionina (panel A), y están sustituidos por secuencias CTG que codifican leucina (panel B). véase el ejemplo 16 para detalles.

Fig. 38. Resultados con constructos de expresión que contienen el gen de resistencia a puromicina con un codón de inicio TTG y codones ATG no internos. Véase el ejemplo 16 para detalles. Los puntos indican puntos de datos individuales; las líneas indican los niveles medios de expresión; los constructos usados se

indican en el eje horizontal, y se representan esquemáticamente encima de la gráfica; el eje vertical indica la señal de d2EGFP.

5 Fig. 39. Igual que Fig. 12, pero para el gen de resistencia a neomicina. Véase el Ejemplo 17 para detalles. A. gen de resistencia a neomicina de tipo salvaje; se indican las secuencias ATG; los ATG que codifican metionina se indican mediante Met encima del ATG. B. gen de resistencia a neomicina sin secuencias ATG, y con un codón de inicio GTG. C. gen de resistencia a neomicina sin secuencias ATG, y con un codón de inicio TTG.

10 Fig. 40. Igual que Fig. 12, pero para el gen dhfr. Véase el Ejemplo 18 para detalles. A. gen dhfr de tipo salvaje; se indican las secuencias ATG, los ATG que codifican metionina se indican mediante Met encima del ATG. B. gen dhfr sin secuencias ATG, y con un codón de inicio GTG. C. gen dhfr sin secuencias ATG, y con un codón de inicio TTG.

15 Fig. 41. Resultados con constructos de expresión (marcador seleccionable de zeocina) según la invención en células PER.C6. Véase el Ejemplo 20 para detalles. Los puntos indican puntos de datos individuales; las líneas indican los niveles medios de expresión; los constructos usados se indican en el eje horizontal, y se representan esquemáticamente encima de la gráfica; el eje vertical indica la señal de d2EGFP.

20 Fig. 42. Resultados con constructos de expresión (marcador seleccionable de blasticidina) según la invención en células PER.C6. Véase el Ejemplo 20 para detalles. Los puntos indican puntos de datos individuales; las líneas indican los niveles medios de expresión; los constructos usados se indican en el eje horizontal, y se representan esquemáticamente encima de la gráfica; el eje vertical indica la señal de d2EGFP.

25 Fig. 43. Resultados con constructos de expresión según la invención, que comprenden además una secuencia de pausa de la transcripción (TRAP). Véase el Ejemplo 21 para detalles. Los puntos indican puntos de datos individuales; las líneas indican los niveles medios de expresión; los constructos usados se indican en el eje horizontal, y se representan esquemáticamente encima de la gráfica; el eje vertical indica la señal de d2EGFP.

30 Fig. 44. Dependencia de la expresión de un anticuerpo con el número de copias usando unidades de transcripción según la invención. Véase el Ejemplo 22 para detalles.

35 Fig. 45. Expresión de anticuerpo a partir de colonias que contienen constructos de expresión según la invención, en las que el número de copias de los constructos de expresión está amplificado mediante metotrexato. Véase el Ejemplo 23 para detalles. Barras blancas: selección con zeocina y blasticidina; barras negras: selección con zeocina, blasticidina y metotrexato (MTX). Los números de colonias ensayadas se representan en el eje horizontal.

40 Fig. 46. Resultados con diferentes promotores. Véase el Ejemplo 24 para detalles. Los puntos indican puntos de datos individuales; las líneas indican los niveles medios de expresión; los constructos usados se indican en el eje horizontal, y se representan esquemáticamente encima de la gráfica; el eje vertical indica la señal de d2EGFP.

45 Fig. 47. Resultados con diferentes elementos STAR. Véase el ejemplo 25 para detalles. Los puntos indican puntos de datos individuales; las líneas indican los niveles medios de expresión; los constructos usados se indican en el eje horizontal, y se representan esquemáticamente encima de la gráfica; el eje vertical indica la señal de d2EGFP.

Fig. 48. Resultados con otros elementos de control de cromatina. Véase el Ejemplo 26 para detalles. Los puntos indican puntos de datos individuales; las líneas indican los niveles medios de expresión; los constructos usados se indican en el eje horizontal, y se representan esquemáticamente encima de la gráfica (los triángulos en negro indican diferentes elementos de control de cromatina ensayados); el eje vertical indica la señal de d2EGFP.

Referencias

- 50 Boshart, M, Weber, F, Jahn, G, Dorsch-Hasler, K, Fleckenstein, B, y Schaffner, W. (1985) A very strong enhancer is located upstream of an immediate early gene of human cytomegalovirus *Cell* 41, 521-530.
- Chung JH, Whiteley M, y Felsenfeld G. (1993) A 5' element of the chicken beta-globin domain serves as an insulator in human erythroid cells and protects against position effect in *Drosophila*. *Cell* 74: 505-514.
- Chung JH, Bell AC, Felsenfeld G. (1997). Characterization of the chicken beta-globin insulator. *Proc Natl Acad Sci USA* 94: 575-580.

- Das, GC, Niyogi, SK, y Salzman, NP. (1985) SV40 promoters and their regulation *Prog Nucleic Acid Res Mol Biol* 32, 217-236.
- Dumas, P, Bergdoll, M., Cagnon, C y Masson JM. 1994. Crystal structure and site-directed mutagenesis of a bleomycin resistance protein and their significance for drug sequestering. *EMBO J* 13, 2483-2492.
- 5 Gill DR, Smyth SE, Goddard CA, Pringle IA, Higgins CF, Colledge WH, y Hyde SC. (2001) Increased persistence of lung gene expression using plasmids containing the ubiquitin C or elongation factor 1 α promoter. *Gene Therapy* 8: 1539-1546.
- Gossen M, y Bujard H. (1992) Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. *Proc Natl Acad Sci USA* 89: 5547-5551.
- 10 Graham FO, Smiley J, Russell W and Nairn R. (1977). Characteristics of a human cell line transformed by DNA from human adenovirus type 5. *J. Gen. Virol.* 36, 59-72.
- Huls GA, Heijnen IAFM, Cuomo ME, Koningsberger JC, Wiegman L, Boel E, van der Vuurst-de Vries A-R, Loyson
- 15 SAJ, Helfrich W, van Berge Henegouwen GP, van Meijer M, de Kruif J, Logtenberg T. (1999). A recombinant, fully human monoclonal antibody with antitumor activity constructed from phage-displayed antibody fragments. *Nat Biotechnol.* 17, 276-281.
- Jones D, Kroos N, Anema R, Van Montfort B, Vooy's A, Van Der Kraats S, Van Der Helm E, Smits S, Schouten J, Brouwer K, Lagerwerf F, Van Berkel P, Opstelten D-J, Logtenberg T, Bout A (2003) High-level expression of recombinant IgG in the human cell line PER.C6. *Biotechnol. Prog.* 19: 163-168.
- 20 Kaufman, RJ. (2000) Overview of vector design for mammalian gene expression *Mol Biotechnol* 16, 151-160.
- Kaufman, RJ, y Sharp, PA. (1982) Construction of a modular dihydrofolate reductase cDNA gene: analysis of signals utilized for efficient expression *Mol Cell Biol* 2, 1304-1319.
- Kellum R, y Schedl P. (1991) A position-effect assay for boundaries of higher order chromosomal domains. *Cell* 64: 941-950.
- 25 Kim SJ, Kim Ns, Ryu CJ, Hong HJ, Lee GM. 1998. Characterization of chimeric antibody producing CHO cells in the course of dihydrofolate reductase-mediated gene amplification and their stability in the absence of selective pressure. *Biotechnol Bioeng* 58: 73-84.
- Kozak M. (1986) Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell* 44: 283-292.
- 30 Kozak M. (1987) An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res.* 15: 8125-8148.
- Kozak M. (1989) Context effects and inefficient initiation at non-AUG codons in eucaryotic cell-free translation systems. *Mol Cell Biol.* 9: 5073-5080.
- 35 Kozak M. (1990) Downstream secondary structure facilitates recognition of initiator codons by eukaryotic ribosomes. *Proc Natl Acad Sci USA* 87:8301-8305.
- Kozak M. (1997) Recognition of AUG and alternative initiator codons is augmented by G in position +4 but is not generally affected by the nucleotides in positions +5 and +6. *EMBO J.* 16: 2482-2492.
- Kozak M. (2002) Pushing the limits of the scanning mechanism for initiation of translation. *Gene* 299: 1-34.
- 40 Kwaks TH, Barnett P, Hemrika W, Siersma T, Sewalt RG, Satijn DP, Brons JF, van Blokland R, Kwakman P, Kruckeberg AL, Kelder A, Otte AP. (2003) Identification of anti-repressor elements that confer high and stable protein production in mammalian cells. *Nat Biotechnol* 21, 553-558. Errata en: *Nat Biotechnol* 21, 822 (2003).
- Phi-Van L, Von Kreis JP, Ostertag W, y Stratling WH. (1990) The chicken lysozyme 5' matrix attachment region increases transcription from a heterologous promoter in heterologous cells and dampens position effects on the expression of transfected genes. *Mol. Cell. Biol.* 10: 2302-2307.
- 45 McBurney, MW, Mai, T, Yang, X, y Jardine, K. (2002) Evidence for repeat-induced gene silencing in cultured Mammalian cells: inactivation of tandem repeats of transfected genes *Exp Cell Res* 274, 1-8.
- Rees, S, Coote, J, Stables, J, Goodson, S, Harris, S, y Lee, MG. (1996) Bicistronic vector for the creation of stable mammalian cell lines that predisposes all antibiotic-resistant cells to express recombinant protein *Biotechniques* 20, 102-104, 106, 108-110.

Schorpp, M, Jager, R, Schellander, K, Schenkel, J, Wagner, EF, Weiher, H, y Angel, P. (1996) The human ubiquitin C promoter directs high ubiquitous expression of transgenes in mice *Nucleic Acids Res* 24, 1787-8.

Stief A, Winter DM, Stratling WH, Sippel AE (1989) A nuclear DNA attachment element mediates elevated and position-independent gene activity. *Nature* 341: 343-345.

5 Van der Vlag, J, den Blaauwen, JL, Sewalt, RG, van Driel, R, y Otte, AP. (2000) Transcriptional repression mediated by polycomb group proteins and other chromatin-associated repressors is selectively blocked by insulators *J Biol Chem* 275, 697-704.

West AG, Gaszner M, Felsenfeld G (2002) Insulators: many functions, many mechanisms. *Genes Dev.* 16: 271-288.

10 Whitelaw, E, Sutherland, H, Kearns, M, Morgan, H, Weaving, L, and Garrick, D. (2001) Epigenetic effects on transgene expression *Methods Mol Biol* 158, 351-68.

Williams S, Mustoe T, Mulcahy T, Griffiths M, Simpson D, Antoniou M, Ivine A, Mountain A, Crombie R (2005) CpG-island fragments from the HNRPA2B1/CBX3 genomic locus reduce silencing and enhance transgene expression from the hCMV promoter/enhancer in mammalian cells. *BMC Biotechnol.* 5:17.

15 LISTADO DE SECUENCIAS

<110> ChromaGenics B.V.

Otte, Arie P.

Kwaks, Theodorus H.J.

Sewalt, Richard G.A.B.

20 van Blokland, Rik

<120> Selección de células hospedantes que expresan proteína en niveles elevados

<130> 0117 US P01 PRO

<160> 142

<170> PatentIn version 3.2

25 <210> 1

<211>749

<212> ADN

<213> Homo sapiens

<220>

30 <221> característica diversa

<223> secuencia de STAR1

<400> 1

ES 2 384 391 T3

atgcggtggg ggcgcgccag agactcgtgg gatccttggc ttggatgttt ggatctttct 60
gagttgcctg tgccgcgaaa gacaggtaca tttctgatta ggcctgtgaa gcctcctgga 120
ggaccatctc attaagacga tggatttggg gggagagtca cagaaagaac tgtggccctc 180
ccctcactgc aaaacggaag tgatthttatt ttaatgggag ttggaatatg tgagggctgc 240
aggaaccagt ctccctcctt cttggttggg aaagctgggg ctggcctcag agacaggttt 300
tttgcccccg ctgggctggg cagtctagtc gaccctttgt agactgtgca caccctaga 360
agagcaacta cccctataca ccaggctggc tcaagtgaaa ggggctctgg gctccagtct 420
ggaaaatctg gtgtcctggg gacctctggt cttgcttctc tcctccctg cactggctct 480
gggtgcttat ctctgcagaa gcttctcgct agcaaacca cattcagcgc cctgtagctg 540
aacacagcac aaaaagccct agagatcaaa agcattagta tgggcagttg agcgggaggt 600
gaatatttaa cgcttttgtt catcaataac tcggttgctt tgacctgtct gaacaagtcg 660
agcaataagg tgaaatgcag gtcacagcgt ctaacaaata tgaaaatgtg tatattcacc 720
ccggtctcca gccggcgcgc caggctccc 749

<210>2

<211>883

<212> ADN

5 <213> Homo sapiens

<220>

<221> característica diversa

<223> secuencia de STAR2

<400> 2

gggtgcttcc tgaattcttc cctgagaagg atggtggccg gtaaggtccg tgtaggtggg 60
gtgcggctcc ccaggccccg gcccggtgtg gtggccgctg cccagcggcc cggcaccccc 120
atagtccatg gcgcccgagg cagcgtgggg gaggtgagtt agaccaaaga gggctggccc 180

10

ES 2 384 391 T3

ggagttgctc atgggctcca catagctgcc cccacgaag acggggcttc cctgtatgtg 240
 tggggctcca tagctgccgt tgccctgcag gccatgagcg tgcgggcat agtcgggggt 300
 gcccctgcg cccgccctg cgcctgtga gcgcttctgt ggggggtggcg ggggtgcgca 360
 gctgggcagg gacgcagggt aggaggcggg gggcagccc taggtaccct gggggggctt 420
 ggagaagggc gggggcgact ggggctcata cgggacgctg ttgaccagcg aatgcataga 480
 gttcagatag ccaccggctc cggggggcac ggggctgcga cttggagact ggcccccca 540
 tgacgttagc atgcccttgc ccttctgatc cttttgtac ttcattgccc gattctggaa 600
 ccagatcttg atctggcgt cagtgaggtt cagcagattg gccatctcca cccggcgcgg 660
 ccggcacagg tagcggttga agtggaaactc tttctccagc tccaccagct gcgcgctcgt 720
 gtaggcgctg cgcgcgogct tggacgaagc ctgccccggc gggtcttctg cgcagcgcga 780
 gctttcgct gcgaggacag agagaggaag agcggcgtca gggctgccc cggccccgcc 840
 cagcccctga ccagcccgg cccctccttc caccaggccc caa 883

<210>3

<211>2126

<212> ADN

5 <213> Homo sapiens

<220>

<221> característica diversa

<223> secuencia de STAR3

<400>3

ES 2 384 391 T3

atctcgagta ctgaaatagg agtaaactctg aagagcaaat aagatgagcc agaaaacat	60
gaaaagaaca gggactacca gttgattcca caaggacatt cccaaggtga gaaggccata	120
tacctccact acctgaacca attctctgta tgcagattta gcaaggttat aaggtagcaa	180
aagattagac ccaagaaaat agagaacttc caatccagta aaaatcatag caaatattt	240
gatgataaca attgtctcca aaggaacaag gcagagtcgt gctagcagag gaagcacgtg	300
agctgaaaac agccaaatct gctttgtttt catgacacag gagcataaag tacacaccac	360
caactgacct attaaggctg tggtaaaccg attcatagag agaggttcta aatacattgg	420
tccctcacag gcaaaactga gttcgctccg aacgtagtcc ctggaaattt gatgtccagt	480
atagaaaagc agagcagtc aaaaatatag ataaagctga accagatggt gcctgggcaa	540
tgttagcagc accacactta agatataacc tcaggctgtg gactccctcc ctggggagcg	600
gtgctgccgg cggcggggcg gctccgcaac tccccggctc tctcgcccgc cctcccgttc	660
tcctcgggcg gcggcggggg ccgggactgc gccgctcaca gcggcggtc ttctgcgccc	720
ggcctcggag gcagtgccgg tggcggccat ggcctcctgc gttcgccgat gtcagcattt	780
cgaactgagg gtcactctct tgggactggt tagacagtgg gtgcagccca cggagggcga	840
gttgaagcag ggtgggggtg cacctcccc aggaagtcca gtgggtcagg gaactccctc	900
ccctagccaa gggaggccgt gagggactgt gcccggtgag agactgtgcc ctgaggaaag	960

ES 2 384 391 T3

gtgcaactctg gccagatac tacacttttc ccacggtcctt caaaacccgc agaccaggag 1020
 attccctcgg gttcctacac caccaggacc ctgggtttca accacaaaac cgggccattt 1080
 gggcagacac ccagctaget gcaagagttg tttttttttt tatactcctg tggcacctgg 1140
 aacgccagcg agagagcacc tttcactccc ctggaaaggg ggctgaaggc agggaccttt 1200
 agctgcgggc tagggggttt ggggttgagt gggggagggg agagggaaaa ggccctcgtca 1260
 ttggcgtcgt ctgcagccaa taaggctacg ctccctcgtc gcgagtagac ccaatccttt 1320
 cctagagggtg gagggggcgg gtaggtggaa gtagagggtg cgcggtatct aggagagaga 1380
 aaaagggtcg gaccaatagg tgcccggaa aggcggacc agcggctcgt tgattgggat 1440
 tggcagtgga ccctccccg gggtggtgcc ggaggggggg atgatgggtc gaggggtgtg 1500
 tttatgtgga agcgagatga ccggcaggaa cctgccccaa tgggctgcag agtggttagt 1560
 gagtgggtga cagacagacc cgtaggccaa cgggtggcct taagtgtctt tggctcctc 1620
 caatggagca gcggcggggc gggaccgcga ctccggttta atgagactcc attgggctgt 1680
 aatcagtgtc atgtcggatt catgtcaacg acaacaacag ggggacacaa aatggcggcg 1740
 gcttagtcct acccctggcg gcggcggcag cggtgccgga ggcgacggca ctccctccagg 1800
 cggcagccgc agtttctcag gcagcggcag cggccccggc aggcgcggtg gcggtgggcg 1860
 gcagccaggt ctgtcaccca ccccgcgct tcccaggggg aggagactgg gcgggagggg 1920
 ggaacagacg gggggggatt caggggcttg cgacgcccct cccacaggcc tctgcgcgag 1980
 ggtcaccgcg gggccgctcg gggtcaggct gcccctgagc gtgacggtag ggggcggggg 2040
 aaaggggagg agggacaggc cccgcccctc ggcagggcct ctagggcaag ggggcggggc 2100
 tcgaggagcg gaggggggcg gggcgg 2126

<210>4

<211>1625

<212> ADN

5 <213> Homo sapiens

<220>

<221> característica diversa

<223> secuencia de STAR4

<400> 4

ES 2 384 391 T3

gatctgagtc atgttttaag gggaggattc ttttggctgc tgagttgaga ttaggttgag 60
ggtagtgaag gtaaaggcag tgagaccacg taggggtcat tgcagtaate caggctggag 120
atgatggtgg ttcagttgga atagcagtgc atgtgctgta acaacctcag ctgggaagca 180
gtatatgtgg cgttatgacc tcagctggaa cagcaatgca tgtggtggtg taatgacccc 240
agctgggtag ggtgcatgtg gtgtaacgac ctcagctggg tagcagtgtg tgtgatgtaa 300
caacctcagc tgggtagcag tgtacttgat aaaatggtg catactctag atttgttatg 360
agggtagtgc cattaatatt ctccacaaat tggttgtcac gtatgagtga aaagaggaag 420
tgatggaaga cttcagtgtc tttggcctga ataaatagaa gacgtcattt ccagttaatg 480
gagacagga agactaaagg taggggtgga ttcagtagag caggtgttca gttttgaata 540

tgatgaactc tgagagagga aaaacttttt ctacctctta gtttttgtga ctggacttaa 600
gaattaaagt gacataagac agagtaacaa gacaaaaata tgcgaggta tttaatattt 660
ttacttgacg aggggaatct tcaaaagaaa aatgaagacc caaagaagcc attaggggtca 720
aaagctcata tgccttttta agtagaaaat gataaathtt aacaatgtga gaagacaaaag 780
gtgtttgagc tgagggcaat aaattgtggg acagtgatta agaaatataat gggggaaatg 840
aaatgataag ttattttagt agattttatt ttcatatcta ttttggcttc aacttccagt 900
ctctagtgat aagaatgttc ttctcttctc ggtacagaga gagcaccttt ctcatgggaa 960
attttatgac cttgctgtaa gtagaaaggg gaagatcgat ctctgtttc ccagcatcag 1020
gatgcaaaca tttcctcca ttccagttct caaccccatg gctgggcctc atggcattcc 1080
agcatcgcta tgagtgcacc tttctgcag gctgcctcgg gtagctggtg cactgctagg 1140
tcagtctatg tgaccaggag ctgggcctct gggcaatgcc agttggcagc ccccatccct 1200
cactgctgg gggcctccta tccagaaggg cttggtgtgc agaacgatgg tgcaccatca 1260
tcattcccca cttgccatct ttcaggggac agccagctgc tttgggcgcg gcaaaaaaca 1320
cccaactcac tcctcttcag gggcctctgg tctgatgcca ccacaggaca tccttgagtg 1380
ctgggcagtc tgaggacagg gaaggagtga tgaccacaaa acaggaatgg cagcagcagt 1440
gacaggagga agtcaaaggc ttgtgtgtcc tggccctgct gagggctggc gagggcctg 1500
ggatggcget cagtgcctgg tcggctgcaa gagccagcc ctctgccat gaggggagct 1560
ggcagtgacc aagctgcact gccctggtgg tgcatcttct gccccactct ttccttctaa 1620
gatcc 1625

<210>5

<211> 1571

5 <212> ADN

<213> Homo sapiens

<220>

ES 2 384 391 T3

<221> característica diversa

<223> secuencia de STAR5

<400> 5

```

cacctgattt aaatgatctg tctggtgagc tcaactgggtc tttactcgca tgctgggtcc      60
acagctccac tgtcctgcag ggtccgtgag tgtgggcccc ttatctatctt catcatcata    120
accctgcgtg tcctcaactc ctggcacata ttgggtggcc ccatccacac acggttggtg    180
agtgaatcca tgagatgaca aaggctatga tgtagactat atcatgagcc agaaccaggc    240
tttctacct ccagacaatc aagggccttg atttgggatt gagggagaaa ggagtagaag    300
ccaggaagga gaagagattg aggtttacca aggggtgcaaa gtcctggccc ctgactgtag    360
gctgaaaact atagaaatga tagaacaatt ttgcaatgaa atgcagaaga cctgcatca    420
actttagggtg ggacttcggg tatttttatg gccacagaac atcctcccat ttacctgcat    480
ggcccagaca cagacttcaa aacagttgag gccagcaggc tccaggtaag tggtaggatt    540

ccagaatgcc ctcagagtgt tgtgggaggc agcaggcgat tttcctggac ttctgagttt    600
atgagaaccc caaaccccaa ttggcattaa cattgaggtc tcaatgtatc atggcaggaa    660
gcttccgagt ggtgaaaagg aaagtgaaca tcaaagctcg gaagacaaga ggggtggagtg    720
atggcaacca agagcaagac ccttcctct cctgtgatgg ggtggctcta tgtgaagccc    780
ccaaactgga cacaggtctg gcagaatgag gaaccactg agatttagcg ccaacatcca    840
gcataaaagg gagactgaca tagaatttga gttagttaaa aataaggcac aatgcttttc    900
atgtattcct gagttttgtg gactgggtgt caatttgcag cattcttagt tgattaaatc    960
tgagatgaag aaagagtgtc caacactttc accttggaag gctctggaaa agcaaaaggg   1020
agagacaatt agcttcatcc attaaactcac ttagtcatta tgcattcatt catgtaacta   1080
ccaaacacgt actgagtgcc taacactcct gagacactga gaagtttctt gggaatacaa   1140
agatgaataa aaaccacgcc aggcaggagt tggaggaagg ttctggatgc caccacgctc   1200
tacctcctgg ctggacacca ggcaatgttg gtaaccttct gcctccaatt tctgcaata   1260
cataattaat aaacacaagg ttatcttcta aacagttctt aaaatgagtc aactttgttt   1320
aaacttgttc ttttagaga aaaatgtatt tttgaaagag ttggttagtg ctaggggaaa   1380
tgtctgggca cagctcagtc tgggtgtgaga gcaggaagca gctctgtgtg tctgggggtg   1440
gtacgtatgt aggacctgtg ggagaccagg ttgggggaag gcccctctc atcaagggtc   1500
cctttgcttt ggtttgcttt ggcgtgggag gtgctgtgcc acaagggaat acgggaaata   1560
agatctctgc t

```

5

<210>6

<211> 1173

<212> ADN

ES 2 384 391 T3

<213> Homo sapiens

<220>

<221> característica diversa

<223> secuencia de STAR6

5

<400> 6

```

tgaccaccca cagacatccc ctctggcctc ctgagtgggt tcttcagcac agcttccaga    60
gccaaattaa acgttcactc tatgtctata gacaaaaagg gttttgacta aactctgtgt    120
tttagagagg gagttaaag ctgttaactt tttaggggtg ggcgagagg atgacaaata    180
acaacttgtc tgaatgtttt acatttctcc cactgcctc aagaaggttc acaacgaggt    240
catccatgat aaggagtaag acctcccagc cggactgtcc ctgggcccc agaggacact    300
ccacagagat atgctaactg gacttggaga ctggctcaca ctccagagaa aagcatggag    360
cacgagcgca cagagcaggg ccaaggctcc agggacagaa tgtctaggag ggagattggg    420
gtgagggtaa tctgatgcaa ttactgtggc agctcaacat tcaaggagg ggaagaaag    480
aaacagtccc tgtcaagtaa gttgtgcagc agagatggta agctccaaa tttgaaactt    540
tggtgctgg aaagttagt ggggcagaga taagaagaca taagagactt tgagggttta    600
ctacacacta gacgctctat gcatttattt atttattatc tcttattat tactttgtat    660

aactcttata ataatcttat gaaaacggaa accctcatat acccatttta cagatgagaa    720
aagtgacaat tttgagagca tagctaagaa tagctagtaa gtaaaggagc tgggacctaa    780
accaaaccct atctcaccag agtacacact cttttttttt ttccagtgtg atttttttta    840
atttttattt tactttaagt tctgggatac atgtgcagaa ggtatggttt gttacatagg    900
tatatgtgtg ccatagtgga ttgctgcacc tatcaaccog tcatctaggt ttaagcccca    960
catgcattag ctatttgtcc tgatgctctc cctcccctcc ccacaccaga caggccttgg   1020
tgtgtgatgt tcccctccct gtgtccatgt gttctcactg ttcagctccc acttatgagt   1080
gagaacgtgt ggtatttggg tttctgttcc tgtgttagtt tgctgaggat gatggcttcc   1140
agcttcatcc atgtccctgc aaaggacacg atc                                     1173

```

<210>7

<211> 2101

10

<212> ADN

<213> Homo sapiens

<220>

<221> característica diversa

<223> secuencia de STAR7

15

<400> 7

ES 2 384 391 T3

aggtgggtgg atcacccgag gtcaggagtt caagaccagc ctggccaaca tggtaaaacc 60
 tcgtctctac taaaaaatac gaaaaattag ctggttgtgg tgggtcgtgc ttgtaatccc 120
 agctactcgg gaggctgagg caggagaatc acttgaatct gggaggcaga ggttgcagtg 180
 agctgagata gtgccattgc actccagcct gggcaacaga cggagactct gtctccaaaa 240
 aaaaaaaaaa aaatcttaga ggacaagaat ggctctctca aacttttgaa gaaagaataa 300
 ataaattatg cagttctaga agaagtaatg gggatatagg tgcagctcat gatgaggaag 360
 acttagctta actttcataa tgcattctgc tggcctaaga cgtggtgagc tttttatgtc 420
 tgaaaacatt ccaatataga atgataataa taatcacttc tgacccccct ttttttctct 480
 ctccctagac tgtgaagcag aaaccccata tttttcttag ggaagtggct acgcactttg 540
 tatttatatt aacaactacc ttatcaggaa attcatattg ttgccctttt atggatgggg 600
 aaactggaca agtgacagag caaaatccaa acacagctgg ggatttcctt ctttttagatg 660
 atgattttaa aagaatgctg ccagagagat tcttgcaagt ttggaggaca tatatgacct 720
 ttaagatatt ttccagctca gagatgctat gaatgtatcc tgagtgcatt gatggacctc 780
 agttttgcag attctgtagc ttatacaatt tgggtggtttt ctttagaaga aaataacaca 840
 tttataaata ttaaaatagg cccaagacct tacaagggca ttcatacaaa tgagaggctc 900
 tgaagtttga gtttgttcac tttctagtta attatctcct gcctgtttgt cataaatgcg 960
 tttagtaggg agctgctaatt gacaggttcc tccaacagag tgtggaagaa ggagatgaca 1020
 gctggettcc cctctgggac agcctcagag ctagtgggga aactatgtta gcagagtgat 1080
 gcagtgacca agaaaatagc actaggagaa agctgggtcca tgagcagctg gtgagaaaag 1140

ES 2 384 391 T3

```

gggtggaat catgtatgcc ctttctgtt ttatTTTTta ttgggtttcc ttttgcctct 1200
caattccttc tgacaataca aaatggttgg tggaacatgg agcacctgga agtctggttc 1260
atTTTctctc agtctcttga tgttctctcg ggttcaactgc ctattgttct cagttctaca 1320
cttgagcaat ctctcaata gctaaagctt ccacaatgca gatTTTgtga tgacaaattc 1380
agcatcaccc agcagaactt aggtTTTTtt ctgtctctcg tttctgacc ttttcttct 1440
gagtgcTTta tgtcacctcg tgaaccatcc tttccttagt catctaccta gcagtcctga 1500
ttctTTtgac ttgtctccct acaccacaat aaatcactaa ttactatgga ttcaatcct 1560
aaaatttgca caaacttgca aatagattac gggTTgaaac ttagagattt caaacttgag 1620
aaaaaagttt aaatcaagaa aaatgacctt taccttgaga gtagaggcaa tgtcatttcc 1680
aggaataatt ataataatat tgtgtTTaat atttgtatgt aacatttgaa taccttcaat 1740
gttcttattt gtgttatttt aatctcttga tgttactaac tcatttggtg ggaagaaaa 1800
catgctaaaa taggcatgag tgtcttatta aatgtgacaa gtgaatagat ggcagaaggt 1860
ggattcatat tcagttttcc atcacctgg aaatcatgcg gagatgattt ctgcttgcaa 1920
ataaaactaa cccaatgagg ggaacagctg ttcttaggtg aaaacaaaac aaacacgcca 1980
aaaaccttta ttctctttat tatgaatcaa atTTTctctc tcagataatt gttttattta 2040
tttattTTta ttattattgt tattatgtcc agtctcactc tgtcgcttaa gctggcatga 2100
t 2101

```

<210>8

<211> 1821

<212> ADN

5 <213> Homo sapiens

<220>

<221> característica diversa

<223> secuencia de STAR8

<400> 8

ES 2 384 391 T3

gagatcacct cgaagagagt ctaacgtccg taggaacgct ctcggttca caaggattga 60
ccgaaccca ggatacgtcg ctctccatct gaggcttgc ccaaatggcc ctccactatt 120
ccaggcacgt ggggtgtctc cctaactctc cctgctctcc tgagcccatg ctgcctatca 180
cccatcggtg caggtccttt ctgaagagct egggtggatt ctctccatcc cacttccttt 240
cccaagaaag aagccaccgt tccaagacac ccaatgggac attccccttc cacctccttc 300
tccaaagttag ccaggtggt catcacaggt tagggagaga agccccagg tttcagttac 360
aaggcatagg acgctggcat gaacacacac acacacacac acacacacac acacacacac 420
acacgactcg aagaggtagc cacaagggtc attaaacact tgacgactgt tttccaaaaa 480
cgtggatgca gttcatccac gccaaagcca aggggtgcaa gcaaacacgg aatgggtggag 540
agattccaga ggctcaccaa accctctcag gaatattttc ctgaccctgg gggcagaggt 600
tggaaacatt gaggacattt cttgggacac acggagaagc tgaccgacca ggcattttcc 660
ttccactgc aatgaccta tggcgggggc atttacttt cccctgcaa tcacctatgg 720

cgaggtacct cccaagccc ccacccccac ttccgcgaat cggcatggct cggcctctat 780
ccgggtgtca ctccaggtag gcttctcaac gctctcggct caaagaagga caatcacagg 840
tccaagccca aagcccacac ctcttccttt tgttataccc acagaagtta gagaaaacgc 900
cacactttga gacaaattaa ggtccttta ttttaagccg cggccaaaga gatggctaac 960
gctcaaaatt ctctgggccc cgaggaaggg gcttgactaa cttctatacc ttggtttagg 1020
aaggggaggg gaactcaaat gcggtaatc tacagaagta aaaacatgca ggaatcaaaa 1080
gaagcaaatg gttatagaga gataaacagt tttaaaaggc aatgggttac aaaaggcaac 1140
ggtaccaggt gcggggctct aatccttca tgacacttag atataggtgc tatgctggac 1200
acgaactcaa ggctttatgt tgttatctct tcgagaaaaa tcctgggaac ttcactgact 1260
gtttgtgcca gtatcttate agttgattgg gctcccttga aatgctgagt atctgcttac 1320
acaggtcaac tccttgcgga agggggttg gtaaggagcc cttcgtgtct cgtaaattaa 1380
ggggtcgatt ggagtttgtc cagcattccc agctacagag agccttattt acatgagaag 1440
caaggctagg tgattaaaga gaccaacagg gaagattcaa agtagcgact tagagtaaaa 1500
acaaggttag gcatttcact tcccagaga acgcgcaaac attcaatggg agagaggtcc 1560
cgagtcgtca aagtcccaga tgtggcgagc ccccgggagg aaaaaccgtg tcttccttag 1620
gatgcccga acaagagcta ggcttccgga gctaggcagc catctatgtc cgtgagccgg 1680
cgggagggag accgccggga ggcgaagtgg ggcggggcca tccttctttc tgctctgctg 1740
ctgccgggga gctcctggct ggcgtccaag cggcaggagg ccgcctcct gcagggcgcc 1800
gtagagtttg cggtcgagag t 1821

<210>9

<211>1929

ES 2 384 391 T3

<212> ADN

<213> Homo sapiens

<220>

<221> característica diversa

5 <223> secuencia de STAR9

<400> 9

```
cacttcctgg gagtggagca gaggctctgc gtggagcacc catgtgcagt actcttaggt      60
acggaaggga ttgggctaaa ccatggatgg gagctgggaa ggaaggaggac caacttcagg      120
ccccactggg aactggagc tgccaccctt tagagccctc ctaaccctac accagaggct      180
gagggggacc tcagacatca cacacatgct ttcccatggt ttcagaaatc tggaaacgta      240
gaacttcagg ggtgagagtg cctagatatt gaatacaagg ctagattggg cttctgtaat      300
atcccaaagg accctccagc tttttacca gcacctaatg cccatcagat accaaagaca      360
cagcttagga gaggttcacc ctgaagctga ggaggaggca gccggattag agttgactga      420
gcaaggatga ctgccttctc cacctgacga tttcagctgc tgcccttttc tttcctggg      480
aatgcctgtc gccatggcct tctgtgtcca caggagagtt tgaccagat actcatggac      540
```

ES 2 384 391 T3

caggcaaagg tgctgttctt cccagcccag ggcccacat gaagcatgcc tgggagcctg 600
 gtaaggaccc agccactcct gggctgttga cattggcttc tcttgcccag cattgtagcc 660
 acgccactgc attgtactgt gagataagtc aaggtgggct caccaggacc tgcactaaat 720
 tgtgaaattc agctccaaag aactttggaa attacccatg catttaagca aatgaatga 780
 tacctgagca aaccctttca cattggcaca agttacaatc ctgtctcacc ctcttgatta 840
 caaattccat ccaggcaaga gctgtatcac cctgaggtct cccattcat gttttgggtca 900
 ataatattta gtttcctttt gaaaatagat ttttgtgtta ctccattatg atgggcagag 960
 gccagatgct tatattctat ttaaatagact atgtttttct atctgtaact gggtttgtgt 1020
 tcaggtggta aatgcttttt ttttgagtc agaagattcc tgggaaggcga ccagaaatta 1080
 gctggccgct gtcagacctg aagttacttc taaagggcct ttagaaatga attctttttt 1140
 atgccttctc tgaattctga gaagtaggct tgacttcccc taagtgtgga gttgggagtc 1200
 aactcttctg aaaagaaagt ttcagagcat tttcaaagc catggtcagc tgtgggaagg 1260
 gaagacgatg gatagtagc ttgccggaaa aactgatgg aggcggatgc tccagctcag 1320
 ccaagacct ttgttctgcc caccacagaa atgccccttc ctcaatcgca gaaacgttgc 1380
 cccatggctc ctgatactca gaatgcagcc tctgaccagg accatctgca tcctccagga 1440
 gctcgtaaga aatgcagcat cgtgggacct gctggcacct ggtgaacca aacctgcagg 1500
 gctcctgggt gtgcttggg cggtgcaggg ggaagagggg gtcagcagcc tcctcctgac 1560
 cttcccgggg gctgcttttc tgaggggcca gaatgcaccg gttgaccttg ttgcatcact 1620
 ggccatgac tggctgcttt ggtcaggtgt aaaaaggtgt ttccagaggg tctgctctc 1680
 tcactatcgg accaggttcc catggagagc tcagcctccc agcaaggata gagaacttca 1740
 aatggctcaa agaactgaga ggccacacat gtgtgacctg aatagtctct gctgcaaac 1800
 aaagggtttc ttaatgtaaa acgttctctt cctcacagag gggttcccag ctgctagtgg 1860
 gcatgttcca ggcatttctt gggctgcacc aggttgcacc aagccagagg atcatttttg 1920
 ggggctcat 1929

<210> 10

<211> 1167

<212> ADN

5 <213> Homo sapiens

<220>

<221> característica diversa

<223> secuencia de STAR10

<220>

10 <221> característica diversa

<222> (452)..(1143)

ES 2 384 391 T3

<223> n es a, c, g, o t en diversas posiciones

<400> 10

```

aggtcaggag ttcaagacca gcctggccaa catggtgaaa ccctgtccct acaaaaaata      60
caaaaattag ccgggcgtgg tggggggcgc ctataatccc agctactcag gatgctgaga      120

caggagaatt gtttgaacct gggagggtgga ggttgacagt aactgagatc gcgccactgc      180
actccagcct ggtgacagag agagactccg tctcaacaac agacaaacaa acaaacaaac      240
aacaacaaaa atgtttactg acagctttat tgagataaaa ttcacatgcc ataaagggtca      300
ccttctacag tatacaattc agtggattta gtatgttcac aaagtgtac gttgttcacc      360
atctactcca gaacatttac atcaccoceta aaagaagctc tttagcagtc acttctcatt      420
ctccccagcc cctgcccaacc acgaatctac tntctgtctc tattctgaat atttcatata      480
aaggagtcct atcatatggg ccttttacgt ctaccttctt tcaacttagca tcatgttttt      540
aagattcatc cacagtgtag cacgtgtcag ttaattcatt tcatcttatg gctggataat      600
gctctattgt atgcatatcc ctcaactttgc ttatccattc atcaactgat tgacatttgg      660
gttatttcta ctttttgact attatgagta atgctgctat gaacattcct gtaccaatcg      720
ttacgtggac atatgctttc aattctcctg agtatgtaac tagggttgga gttgctgggt      780
catatgtaa ctcagtgttt catttttttg aagaactacc aaatggtttt ccaaagtgga      840
tgcaacactt tacattccca ccagcaagat atgaaggttc caatgtctct acatttttgc      900
caacacttgt gattttcttt tatttattta tttatttatt tatttttgag atggagtctc      960
actctgtcac ccaggctgga gtgcagtggc acaatttcag ctcaactgaa tctccacctc     1020
tcgggctcaa gcgatactcc tgcccaacc tcccagataa ctgggattac aggcgccccac     1080
caccacacca agctaatttt ttgtattttt agtagagacg gggtttcac atgtcggcca     1140
ggntgtactc gaactctgac ctcaagt                                     1167

```

<210> 11

5 <211> 1377

<212> ADN

<213> Homo sapiens

<220>

<221> característica diversa

10 <223> secuencia de STAR11

<400> 11

ES 2 384 391 T3

aggatcactt gagcccagga gttcaagacc agcctgggca acatagcgag aacatgtctc 60
 aaaaaggaaa aaaatggggg aaaaaaccct cccagggaca gatatccaca gccagtcttg 120
 ataagctcca tcattttaaa gtgcaaggcg gtgcctccca tgtggatgat tatttaatcc 180
 tcttgtactt tgtttagtcc tttgtggaaa tgcccatctt ataaattaat agaattctag 240
 aatctaatta aaatggttca actctacatt ttactttagg ataatatcag gaccatcaca 300
 gaatgtctga gatgtggatt taccctatct gtagctcact tcttcaacca ttcttttagc 360
 aaggctagtt atcttcagtg acaaccctt gctgccctct actatctcct ccctcagatg 420
 gactactctg attaagcttg agctagaata agcatgttat cccgggattt catatggaat 480
 atttataca tgagtgagcc attatgagtt gtttgaaaat ttattatgtt gagggagggt 540
 aaccgctgta acaaccatca ccaaacttaa tgcactgaat acatttgacg tttatttctt 600
 gttcacctga cagttcagtg ttacctaata ttacatgaag acccagaggc ccacgctcct 660

 tcattttggg ctccaccgac ctccaaggtt tcagggcctt ctgccccgcc tcttgcaccc 720
 acaggggaag agagtggagg atgcacacgc ccaggcctgg aagtgacgca tgtggcttcc 780
 ccgtccacag acttcaccca cagtccattg gccttcttaa gtcattggact cctgctgagc 840
 tgccagggtg catgggaaat ccatgtgact gtgtgccctg gaggaagggg agcgtttcgg 900
 tgagcacaca ggagtotttg ccactagacg ctgatgagga ttccccacag gcgatgaagc 960
 atggagactc atcttgtaac aacagatga gttgttgaca tctcttaagt ttactttgtg 1020
 tgcagttttt attcagatag gaaaggctgt taaaatctta acacctaact ggaagaaggg 1080
 ttttagagaa gtgtggtttt cagtaagcca gttctttcca caatccaaga aacgaaataa 1140
 atttccagca tggagcagtt ggcaggtaa gtttttggtg tggctctgcc caggcttgag 1200
 tgtaaccggt gtggtcatag ctactacat tctcaaactc ctggccttaa gtcatectcc 1260
 tgcctcagcc toccaaaggc aagtaagggt aagaataggg gaaagggtgaa gtttcacagc 1320
 ttttctagaa ttctttttat tcaagggact ctcatatcat caaaccacc cagaatc 1377

<210> 12

<211> 1051

5 <212> ADN

<213> Homo sapiens

<220>

<221> característica diversa

<223> secuencia de STAR12

10 <400> 12

ES 2 384 391 T3

atcctgcttc tgggaagaga gtggcctccc ttgtgcaggt gactttggca ggaccagcag 60
 aaaccaggt ttctgtcag gaggaagtgc tcagcttacc tctgtgaagg gtcgtgataa 120
 ggcacgagga ggcaggggct tgccaggatg ttgcctttct gtgccatag ggacatctca 180
 gcttacgttg ttaagaaata tttggcaaga agatgcacac agaatttctg taacgaatag 240
 gatggagttt taagggttac tacgaaaaaa agaaaactac tggagaagag ggaagccaaa 300
 caccaccaag tttgaaatcg attttattgg acgaatgtct cactttaaat ttaaattggag 360
 tccaacttcc ttttctcacc cagacgtcga gaagggtggca ttcaaatgt ttacacttgt 420
 ttcatctgcc tttttgctaa gtctgtgtcc cctacctcct ttccctcact tcacatttgt 480
 cgtttcatcg cacacatag ctcatcttta tatttacata tatataattt ttatatatgg 540
 cttgtgaaat atgccagacg agggatgaaa tagtctgaa aacagctgga aaattatgca 600
 acagtgggga gattgggcac atgtacattc tgtactgcaa agttgcacaa cagaccaagt 660
 ttgttataag tgaggctggg tggtttttat tttttctcta ggacaacagc ttgcctggtg 720
 gaggtagcct cctgcagaag gcattttctt aggagcctca acttcccaa gaagaggaga 780
 gggcgagact ggagttgtgc tggcagcaca gagacaaggg ggcacggcag gactgcagcc 840
 tgcagagggg ctggagaagc ggaggctggc acccagtggc cagcgaggcc caggtccaag 900
 tccagcgagg tcgaggtcta gagtacagca agccaaggt ccaaggtcag tgagtctaag 960

 gtccatggtc agtgaggctg agaccaggg tccaatgagg ccaaggcca gaggccagta 1020
 aggccgagat ccagggtcca gggaggtaa g 1051

<210> 13

<211> 1291

<212> ADN

5 <213> Homo sapiens

<220>

<221> característica diversa

<223> secuencia de STAR13

<400> 13

ES 2 384 391 T3

agccactgag gtcctaactg cagccaaggg gccgttctgc acatgtegct caccctctgt 60
gctctgttcc ccacagagca aacgcacatg gcaacgttg tccgctcagc cactggttct 120
gtggtggaac ggtggatgtc tgcactgtga catcagctga gtaagtaaca acgactgagg 180
atgccgctga cccagggctg ggggaagggga ctcccagctc agacaggctt ggctgtggtt 240
tgctttggga ggagagtga catcacaggg aatggctcat gtcagcccca ggaggggtggg 300
ctggcccctg gtccccgggc tccttctggc cctgcaggcg atagagagcc tcaacctgct 360
gccgcttctc cttggcccgg gtgatggccg tctggaagag cctgcagtag aggtgcacag 420
ccagcggaga gtcgtcattg ccgggtacag ggtagtgat gaggcagggg ttgcagttgg 480
tgtccacgat gccactgtg gggatgttca tcttggctgc gtctctcacg gccacgtgtg 540
gctcaaagat gttgttgagc gtgtgcagga agatgatgag gtccggcagg cggaccgtgg 600
ggccaaagag gaggcgcgcg ttggtcagca tgccgccct gaagttagcga gtgtgggctg 660
actgccaca gtcacgggcc atgttctcaa tcaggtacga gaactgccgg ttgcggctta 720
taaacaagat gatgcccttg cggtaggcca tgtggcggt gaagttcaag gccagctgga 780
ggtgcgtggc tgtctgttcc aggtcagatg tgctgtggtc caggcggctc ccaaagatgt 840
acggctccat aaacctgcca gagacccac caaggcaagg gggatgagag ttcacggggc 900
catctccact ggctccttgc aggaacacag acgcccacca gggactcccg ggctcctctg 960
tgggggcact atgggctggg aagcacaatt tgcaacgctc cccgtgtgca tggacagcag 1020
tgcagaccca tccagggcac ccctctgcat gcctcgtctc gtggcttaac ccctoctacc 1080
ctctacctct tcccgaagga atcctaatag aactgacccc atatggatgt gtggacatcc 1140
aacatgacgc caaaaggaca ttctgccccg tgcagctcac agggcagccg cctccgtcac 1200
tgtcctcttc ccgaggcttt gcggatgagg cccctctggg gttggactta gcggggtgct 1260
ctggggcaaaa agcattaagg gatcagggca g 1291

<210> 14

<211> 711

<212> ADN

5 <213> Homo sapiens

<220>

<221> característica diversa

<223> secuencia de STAR14

<400> 14

ES 2 384 391 T3

```

ccttggacca gggccgtgg tcttgggtgg cactggcttc ttcttgctgg gtgttttct 60
gtgggtctct ggcaaggcac tttttgtggc gctgcttgct ctgtgtgcgg gaggggcagg 120
tgctctttcc tcttggagct ggaccctctg gggcgggtcc cegtgcgcct ccttgtgtgt 180
tttctgcacc tggtagagct ggatggcctc ctcaatgccg tcgtcgtctc tggagtcgga 240
cgcctcgggc gcctgtacgg cgctcgtgac tcgctttccc ctcttgccg tgctggcggt 300
ccttttaate ccacttttat tctgtactgc ttctgaaggg cgggtgggggt tgctggcttt 360
gtgtgcctt ccttctcctg cgtggctcgtg gtcgtgacct tggacctgag gcttctgggc 420
tgcacgtttg tctttgctaa ccgggggagg tctgcagaag gcgaactcct tctggacgcc 480
catcaggccc tgccgggtgca ccacctttgt agccggctct tgggtgggatt tcgagagtga 540
cttcgccgaa ttttcatgtg tgtctgggtt ctctccact gacctcac atttttgggt 600
ctcatgctgt cttttctcat tcagaaactg ttctatttct gccctgatgc tctgctcaaa 660
ggagtctgct ctgctcatgc tgactgggga ggagagccc tggtccttgc t 711

```

<210> 15

<211> 1876

<212> ADN

5 <213> Homo sapiens

<220>

<221> característica diversa

<223> secuencia de STAR15

<400> 15

ES 2 384 391 T3

gagtccaaga tcaaggtgcc agcatcttgt gagggccttc ttgttacgtc actccctagc 60
 gaaagggcaa agagaggggtg agcaagagaa aggggggctg aactcgtcct tgtagaagag 120
 gccattccc gagacaatgg cattcatcca ttcaactccac cctcatggcc tcaccacctc 180
 tcatgaggct ccacctccca gccctggttt gttggggatt aaatttccaa cacatgcctt 240
 ttgggggaca tgttaaaatt atagcacccc aaatgttaca ctatcttttg atgagcggta 300
 gttctgattt taagtctagc tggcctactt tttcttgcaac gtgggatgct ttctgcctgt 360
 tccagggcag gcagctcttc tctgtccctc tgctggcccc acctcatcct ctgttgtcct 420
 cttccctcct tctgtgccct ggggtcctgg tgggggtgtg actgtcaact gcgttgggct 480
 aacttttttc cctgctgggt gcccgtaatg aaagaaagct tcttgctccc aagttcctta 540
 aatccaagct catagacaac gcggtctcac agcaggcctg gggccagcct cacgtgagcc 600
 cttccctgg tgtagtcact ggcattgggg aatgggattt cctgttgccc tactgtgtgg 660
 ctgaggtggg ggttgcttcc tggagccagg ctttgtggaa gggcagtgcc cactgcagtg 720
 gatgctgggc cctgaatctg acccagtggt tcattggctc tgtgagacct agtgagggca 780
 gggagggaaag tggagctggg gtgagaagta gaggcctgc agggcccacg tgccagccac 840
 caggcctcag actaggetca gatgacggag agctgcacac ctgcccaacc caggccctgc 900
 agtgcceaca tgccagccgc tggggcccag acttgctcca gagggcggag agctttacac 960

 cggcccaacc caggccatgg ctccaaatgc gtgacagttt tgctgttgct tcttttagtc 1020
 attgtcaagt tgatgcttgt tttgcagagg accaaggctt tatgaacctt ttaccctgtg 1080
 tgaagagttt caccaggtta tggaaatttc tttaaaacca taccacagtt ttttcattat 1140
 tcatgtatat ttttaaaaat aattactgca ctcaagtagaa taacatgaaa atgttgccctg 1200
 ttagcccttt tccagtttgc cccgagaata ctgggggcac ttgtggctgc aatgtttatc 1260
 ctgcggcagc tttgccatga agtatctcac ttttattatt atttttgcat tgctcgagta 1320
 tattgacttt ggaacaaaa gacatcattc tatttatagc attatgtttt tagtagtggt 1380
 atttccatat acaagataca gtaattttcc gtcaatgaaa atgtcaaatt ctagaaaatg 1440
 taacattcct atgcgtgggt ttaacatcgt tctctaacag ttgttggccg aagattcgtt 1500
 tgatgaatcc gatttttcca aaatagccga ttctgatgat tcagacgatt ctgatgttct 1560
 gtttagaaat aattccaaga acagttttta cattttattt tcacattgaa aatcagtcag 1620
 atttgcttca gcctcaaaga gcacgtttat gtaaaattaa atgagtgctg gcagccagct 1680
 gcgctttggt tttctaaatg ggaaaagggt taaatttcac tcagctttta aatgacagcg 1740
 cacagcctgt gtcataagagg gttggaggag atgactttaa ctgcctgtgg ttaggatccc 1800
 tttccccag gaatgtctgg gagcccactg ccgggtttgc tgtccgtctc gtttgactc 1860
 agttctgcat gtactg 1876

ES 2 384 391 T3

<210> 16
 <211> 1282
 <212> ADN
 <213> Homo sapiens

5

<220>
 <221> característica diversa
 <223> secuencia de STAR16
 <400> 16

```

cgcccacctc ggctttccaa agtgctggga ttacaggcat gagtcaactgc gcccatcctg      60
attccaagtc ttagataat aacttaactt tttegaccaa ttgccaatca ggcaatcttt      120
gaatctgcct atgacctagg acatccctct ccctacaagt tgccccgcgt ttccagacca      180
aaccaatgta catcttacat gtattgattg aagttttaca tctccctaaa acatataaaa      240
ccaagctata gtctgaccac ctcaggcacg tgtctcagg acctccctgg ggctatggca      300
tgggtcctgg tctcagatt tggtcagaa taaatctctt caaatatctt ccagaatctt      360
actcttttca tcaccattac ctatcaccca taagtcagag ttttccacaa ccccttctc      420
agattcagta atttgctaga atggccacca aactcaggaa agtattttac ttacaattac      480
caatttatta tgaagaactc aaatcaggaa tagccaaatg gaagaggcat agggaaaggt      540
atggaggaag gggcacaaag cttccatgcc ctgtgtgcac accaccctct cagcatcttc      600
atgtgtcac caactcagaa gctcttcaaa ctttgcatt taggggtttt tatggcagtt      660
ccactatgta ggcattggtg ataaatcact ggtcatcggg gatagaactc tgtctccagc      720

tctctctctc ctctcccca gaagtcctga ggtggggctg aaagtttcac aaggtagtt      780
gctctgacaa ccagccccta tctgaagct attgaggggt ccccaaaaag ttaccttagt      840
atggttgaa gaggttatt atgaataaca aaagatgctc ctatctttac cactagggag      900
catatccaag tcttgcgga acaaagcatg ttactggtag caaattcata caggtagata      960
gcaatctcaa ttcttgctt ctcagaagaa agaatttgac caagggggca taaggcagag     1020
tgagggacca agataagttt tagagcagga gtgaaagttt attaaaaagt ttaggcagg     1080
aatgaaagaa agtaaagtac atttggaaga gggccaagtg ggcgacatga gagagtcaaa     1140
caccatgcc tgtttgatgt ttggcttggg gtcttatatg atgacatgct tctgagggtt     1200
gcatecttct cccctgatte ttcccttggg gtgggctgtc cgcatgcaca atggcctgcc     1260
agcagtaggg aggggccgca tg                                             1282
    
```

10

<210> 17
 <211> 793
 <212> ADN

ES 2 384 391 T3

<213> Homo sapiens

<220>

<221> característica diversa

<223> secuencia de STAR17

5 <400> 17

```

atccgagggg aggaggagaa gaggaaggcg agcagggcgc cggagcccga ggtgtctgcg      60
agaactgttt taaatgggtg gcttgaaaat gtcactagtg ctaagtggct tttcggattg      120
tcttatttat tactttgtca ggtttcctta aggagagggg gtggtggggg tgggggagga      180
ggtggactgg ggaaacctct gcgtttctcc tcctcggtg cacaggggta gtaggaaacg      240
cctcgctgcc acttaacaat ccctctatta gtaaacttac gcgagactc tatgggaagc      300
cgagaaccag tgtcttcttc cagggcagaa gtcacctgtt gggaacggcc cccgggtccc      360
cctgctgggc tttcggctc ttctaggcgg cctgatttct cctcagccct ccaccacgcg      420
tcctcaggg acttttcaca cctccccacc ccatttcca ctacagtctc ccagggcaca      480
gcacttcatt gacagccaca cgagccttct cgttctcttc tcctctgttc cttctcttctc      540
tcttctctc tgttccttct ctttctctgt cataatttcc ttggtgcttt cgccacctta      600
aacaataaag agaaaaaat aaaataaaaa aaacccttcc tgagccaaag tattttaaga      660
tgaatccaag aaagcgacct acatagccct ccccaccac ggagtgcgcc aagacgcacc      720
caggctccat cacagggccg agagcagcgc cactctggtc gtacttttgg gtcaagagat      780
cttgcaaaag agg                                             793

```

<210> 18

<211>492

<212> ADN

10 <213> Homo sapiens

<220>

<221> característica diversa

<223> secuencia de STAR18

<400> 18

ES 2 384 391 T3

atctttttgc tctctaaatg tattgatggg ttgtgttttt tttcccacct gctaataaat 60
attacattgc aacattcttc cctcaacttc aaaactgctg aactgaaaca atatgcataa 120
aagaaaatcc tttgcagaag aaaaaaagct attttctccc actgattttg aatggcactt 180
gcggatgcag ttcgcaaate ctattgccta tcccctcatg aacattgtga aatgaaacct 240
ttggacagtc tgccgcattg cgcattgagac tgccctgcga aggcaagggt atggttccca 300
aagcaccagc tggtaaatec taacttatta tccccttaa attccaatgt aacaacgtgg 360
gccataaaag agtttctgaa caaaacatgt catctttgtg gaaagggtgt tttcgttaatt 420
aatgatggaa tcatgctcat ttcaaatgg aggtccacga tttgtggcca gctgatgcct 480
gcaaattatc ct 492

<210> 19

<211> 1840

<212> ADN

5 <213> Homo sapiens

<220>

<221> característica diversa

<223> secuencia de STAR19

<400> 19

ES 2 384 391 T3

tcacttctcg atatattaca ttcaaggcta gctttatgca tatgcaacct gtgcagttgc 60
 acagggcttt gtgttcagaa agactagctc ttggtttaat actctgttgt tgccatcttg 120
 agattcatta taatataatt tttgaatttg tgttttgaac gtgatgtcca atgggacaat 180
 ggaacattca cataacagag gagacaggtc aggtggcagc ctcaattcct tgccaccctt 240
 ttcacataca gcattggcaa tgcccatga gcacaaaatt tgggggaacc atgatgctaa 300
 gactcaaagc acatataaac atgttacctc tgtgactaaa agaagtggag gtgctgacag 360
 ccccagagg ccacagttta tgttcaaacc aaaacttgct tagggtgagc aaagaaggca 420
 atggcagggc ctaagaaaca gcccatcata tccttgttta ttcattgttac gtccctgcat 480
 gaactaatca cttacactga aaatattgac agaggaggaa atggaaagat agggcaaccc 540
 atagttcttt ttcttttag tctttcctta tcagtaaacc aaagatagta ttggtaaaat 600
 gtgtgtgagt taattaatga gttagtttta ggcagtgttt ccactgttgg ggtaagaaca 660
 aaatatatag gcttgtattg agctattaaa tgtaaattgt ggaatgtcag tgattccaag 720
 tatgaattaa atatccttgt atttgcattt aaaattggca ctgaacaaca aagattaaca 780
 gtaaaaattaa taatgtaaaa gtttaatttt tacttagaat gacattaat agcaaataaa 840
 agcaccatga taaatcaaga gagagactgt ggaaagaagg aaaacgtttt tatttttagta 900
 tatttaatgg gactttcttc ctgatgtttt gttttgtttt gagagagagg gatgtggggg 960
 cagggaggtc tcattttggt gcccaggctg gacttgaact cctgggctcc agctatcctg 1020
 ccttagcttc ttgagtagct gggactacag gcacacacca cagtgtctga cattttctgg 1080
 attttttttt tttttttatt ttttttgtga gacaggttct ggctctgtta ctcaggttgc 1140

 agtgcagtgg catgatagcg gctcactgca gcctcaacct cctcagetta agctactctc 1200
 ccacttcagc ctctgagta gccaggacta cagttgtgtg ccaccacacc tgtggctaata 1260
 tttttagtag atggggcttc tccacgttgc cgaggctggc ctccaactcc tggctctcaag 1320
 cgaacctcct gacttggcct cccgaagtgc tgggattaca ggcttgagcc actgcatcca 1380
 gcctgtcttc tgtgttaaac ctactccaat ttgtctttca tctctacata aacggctctt 1440
 ttcaaagttc ccatagacct cactgttgcct aatctaataa taaattatct gccttttctt 1500
 acatggttca tcagtagcag cattagattg ggctgctcaa ttcttcttgg tatattttct 1560
 tcatttggtc tctggggcat cacactctct ttgagttact cattcctcat tgatagcttc 1620
 ttcttagtct tctttactgg ttcttctctt tctcctgac tccttaatat tgtttttctc 1680
 cccaggcttt agttcttagt cctcttctgt tatctattta cacccaattc tttcagagtc 1740
 tcatccagag tcatgaactt aaacctgttt ctgtgcagat aattcacatt attatatctc 1800
 cagcccagac tctcccgcaa actgcagact gatcctactg 1840

<210> 20

<211> 780

ES 2 384 391 T3

<212> ADN

<213> Homo sapiens

<220>

<221> característica diversa

5 <223> secuencia de STAR20

<400> 20

```

gatctcaagt ttcaatatca tgttttggca aaacattcga tgctcccaca tccttaccta      60
aagctaccag aaaggctttg ggaactgtca acagagctac agaaaagtca gtaaagacca      120
atggaccctt caaacaaaaa cagccaagct tttctgccaa aaagatgact gagaagactg      180
ttaaagcaaa aaactctggt cctgcctcag atgatggcta tccagaaata gaaaaattat      240
ttcccttcaa tcctctaggc ttcgagagtt ttgacctgcc tgaagagcac cagattgcac      300
atctcccctt gagtgaagtg cctctcatga tacttgatga ggagagagag cttgaaaagc      360
tgtttcagct gggccccctt tcacctttga agatgccctc tccaccatgg aatccaatc      420
tgttgcagtc tcctttaagc attctgttga ccctggatgt tgaattgcca cctgtttgct      480
ctgacataga tatttaaatt tcttagtgct ttagagtttg tgtatatttc tattaataaa      540
gcattatttg tttaacagaa aaaaagatat atacttaaat cctaaaataa aataaccatt      600
aaaaggaaaa acaggagtta taactaataa ggaacaaaag gacataaaat gggataataa      660
tgcttaatcc aaaataaagc agaaaatgaa gaaaatgaa atgaagaaca gataaataga      720
aaacaaatag caatatgaaa gacaaaacttg accgggtgtg gtggctgatg cctgtaatcc      780

```

<210> 21

<211> 607

10 <212> ADN

<213> Homo sapiens

<220>

<221> característica diversa

<223> secuencia de STAR21

15 <400> 21

ES 2 384 391 T3

gatcaataat ttgtaatagt cagtgaatac aaaggggtat atactaaatg ctacagaaat 60
tccattcctg ggtataaatc ctagacatat ttatgcatat gtacaccaag atatatctgc 120
aagaatgttc acagcaaatc tctttgtagt agcaaaagggc caaaaggtct atcaacaaga 180
aaattaatac attgtggcac ataatggcat ccttatgcc aataaaatgg atgaaattat 240
agttaggttc aaaaggcaag cctccagata atttatatca tataattcca tgtacaacat 300
tcaacaacaa gcaaaaactaa acatatacaa atgtcagggg aatgatgaa caagggttaga 360
aatgattaa tataaaaata ctgcacagtg ataacattta atgagaaaa aagaaggaag 420
ggcttaggga gggacctaca gggaaactcca aagttcatgg taagtactaa atacataatc 480
aaagcactca aatagaaaa tatttttagta atgttttagc tagttaatat cttacttaaa 540
acaaggtcta ggccaggcac ggtggctcac acctgtaatc ccagcacttt gggaggctga 600
ggcgggt 607

<210> 22

<211> 1380

<212> ADN

5 <213> Homo sapiens

<220>

<221> característica diversa

<223> secuencia de STAR22

<400> 22

ES 2 384 391 T3

cccttgatgat ccacccgect tggcctccca aagtgctggg attacaggcg tgagtcacta 60
cgcccggcca ccctccctgt atattatttc taagtatact attatgttaa aaaaagttta 120
aaaatattga tttaatgaat tcccagaaac taggatttta catgtcacgt tttcttatta 180
taaaaaataaa aatcaacaat aaatatatgg taaaagtaaa aagaaaaaca aaaacaaaaa 240
gtgaaaaaaa taaacaacac tctgtcaaaa aaacaacagt tgtgataaaa ctttaagtgcc 300
tgaaaattta gaaacatcct tctaaagaag ttctgaataa aataaggaat aaaaataatca 360
catagttttg gtcattgggt ctgtttatgt gatggattat gtttattgat ttgtgtatgt 420
tgaacttate tcaatagatg cagacaaggc cttgataaaa gtttttaaca ccttttcatg 480
ttgaaaactc tcaatagact aggtattgat gaaacataatc tcaaaataat agaagctatt 540
tatgataaac ccatagccaa taccatactg agtgggcaaa agctggaagc attccctttg 600
aaaactggca caagacaagg atgccctctc tcaccactcc tattaatgt agtattggaa 660
gttctggcca gagcaatcag gcaggagaaa gaaaaggat taaaatagga agagaggaag 720
tcaaatgtc tctgtttgca gtaaacaatga ttgtatattt agaaaacccc attgtctcat 780
cctaaaaact ccttaagctg ataaacaact tcagcaaagt ctcaggatac aaaatcaatg 840
tgcaaaaatc acaagcattc ctatacaccg ataatagaca gcagagagcc aatcatgag 900

tgaagtccca ttcacaattg cttcaagaa aataaaatac ttaggaatac aactttcacg 960
ggacatgaag gacattttca aggacaacta aaaaccactg ctcaaggaaa tgagagagga 1020
caciaaagaaa tggaaaaaca ttccatgctc atggaagaat caatatcatg aaaatggcca 1080
tactgcccaa agtaatttat agattcaatg ctaaccccat caagccacca ttgactttct 1140
tcacagaact agaaaaaac tatttttaaaa ctcatatgta gtcaaaaaga gtcggtatag 1200
ccaagacaat cctaagcata aagaacaaag ctggatgcat cacgctgact tcaaaccata 1260
ctacaaggct acagtaacca aaacagcatg gtactggtag caaaacagat agatagaccg 1320
atagaacaga acagaggcct cggaaataac accacacatc tacaaccctt tgatcttcaa 1380

<210> 23

<211> 1246

5 <212> ADN

<213> Homo sapiens

<220>

<221> característica diversa

<223> secuencia de STAR23

10 <400> 23

ES 2 384 391 T3

atccccctcat ccttcagggc agctgagcag ggcctcgagc agctggggga gcctcactta 60
atgctcctgg gagggcagcc agggagcatg gggctcgagc gcatgggtcca gggctcctgca 120
ggcggcacgc accatgtgca gccgccccca cctgttgcctc tgccctccgcc acctggccat 180
gggcttcagc agccagccac aaagtctgca gctgctgtac atggacaaga agcccacaag 240
cagctagagg accttgtggt ccacgtgccc agggagcatg gcccacagcc caaagaccag 300
tcaggagcag gcaggggctt ctggcagggc cagctctacc tctgtcttca cacagatggg 360
agatttctgt tgtgattttg agtgatgtgc ccctttggtg acatccaaga tagttgctga 420
agcaccgctc taacaatgtg tgtgtattct gaaaacgaga acttctttat tctgaaataa 480
ttgatgcaaa ataaattagt ttggatttga aattctattc atgtaggcat gcacacaaaa 540
gtccaacatt gcatatgaca caaagaaaag aaaaagcttg cattccttaa atacaaatat 600
ctgttaacta tatttgcaaa tatatttgaa tacacttcta ttatgttaca tataatatta 660
tatgtatatg tatatataat atacatatat atgttacata taatatactt ctattatggt 720
acataataata tttatctata agtaaataca taaatataaa gatttgagta gctgtagaac 780
attgtcttat gtgttatcag ctactactac aaaaatatct cttccactta tgccagtttg 840
ccatataaat atgatcttct cattgatggc ccagggcaag agtgcagtgg gtacttattc 900
tctgtgagga gggaggagaa aagggaaaca ggagaaagtc acaaagggaa aactctggtg 960
ttgccaaaat gtcaagtttc acatattccg agacggaaaa tgacatgtcc cacagaagga 1020
ccctgcccag ctaatgtgtc acagatatct caggaagctt aatgatttt tttaaaagaa 1080
aagagatggc attgtcactt gtttcttgta gctgaggctg tgggatgatg cagatttctg 1140
gaaggcaag agctcctgct tttccacac cgagggactt tcaggaatga ggccagggtg 1200
ctgagcacta caccaggaaa tccctggaga gtgtttttct tactta 1246

<210> 24

<211> 939

<212> ADN

5 <213> Homo sapiens

<220>

<221> característica diversa

<223> secuencia de STAR24

<400> 24

ES 2 384 391 T3

acgagggtcac gagttcgaga ccagcctggc caagatgggtg aagccctgtc tctactaaaa 60
 atacaacaag tagccgggcg cgggtgacggg cgctgtaat cccagctact caggaggctg 120
 aagcaggaga atctctagaa cccaggagge ggaggtgacg tgagctgaga ctgccccgct 180
 gcaactctagc ctgggcaaca cagcaagact ctgtctcaaa taaataaata aataaataaa 240
 taaataaata aataaataaa tagaaagga gagttggaag tagatgaaag agaagaaaag 300
 aaatocctaga tttcctatct gaaggcacca tgaagatgaa ggccacctct tctgggccag 360
 gtcctcccgt tgcaggtgaa ccgagttctg gcctccattg gagaccaaag gagatgactt 420
 tggcctggct cctagtgagg aagccatgcc tagtccctgtt ctgtttgggc ttgatcctgt 480
 atcacttgat tgtctctcct ggactttcca tggattccag ggatgcaact gagaagtta 540
 tttttaatgc acttacttga agtaagagtt attttaaaac attttagcaa aggaaatgaa 600
 ttctgacagg ttttgactg aagacattca catgtgagga aacaggaaa accactatgc 660
 tagaaaaagc aatgctgtt gagattgtct cacaaacaca aattgctgc cagcaggtag 720
 gtttgagcct cagggtgggc acattttacc ttaagcgcac tgttggtgga acttaagggtg 780
 actgtaggac ttatatatac atacatacat ataatatata tacatattta tgtgtatata 840
 cacacacaca cacacacaca cacacagggt cttgctatct tgcccagggt ggtctccaac 900
 tctgggtctc aagcgatcct ctgcctcccc ttcccaaag 939

<210> 25

<211> 1067

<212> ADN

5 <213> Homo sapiens

<220>

<221> característica diversa

<223> secuencia de STAR25

<400> 25

cagccccctct tgtgttttctc tttatttctc gtacacacac gcagttttaa ggggtgatgtg 60
 tgtataatta aaaggaccct tggccatac tttcctaatt ctttagggac tgggattggg 120
 tttgactgaa atatgttttg gtggggatgg gacggtggac ttccattctc cctaaactgg 180
 agttttggtc ggtaatcaaa actaaaagaa acctctggga gactggaaac ctgattggag 240
 cactgaggaa caagggaatg aaaaggcaga ctctctgaac gtttgatgaa atggactctt 300
 gtgaaaatta acagtgaata ttcactgttg cactgtacga agtctctgaa atgtaattaa 360
 aagtttttat tgagcccccg agctttggct tgcgctgatt tttccggtcg cggacatccc 420

10

ES 2 384 391 T3

accgcgcaga gcctcgcctc cccgctgccc tcagcctccg atgacttccc cgcccccgcc 480
ctgctcggtg acagacgttc tactgcttcc aatcggaggc acccttcgcg ggagcggcca 540
atcgggagct ccggcaggcg gggaggcccg gccagttaga tttggaggtt caacttcaac 600
atggccgaag caagtagcgc caatctaggc agcggctgtg aggaaaaag gcatgagggg 660
tcgtcttcgg aatctgtgcc acccggcact accatttcga gggagaagct cctcgacacc 720
atggtggaca cttttcttca gaagctggtc gccgcccggca ggtaaagtgg acgcagccgc 780
ggtagggagt tttgttgcca ccgaagctca aatcccgcga ggtcaggacg gccgcaggct 840
ggcgcgcggg gacgtgggtc cgcggtgggg gcggggcagt cggacgagge gaccagtca 900
aatcctgagc cttaggagtc agggatttca cgcactgata acctgtagcg gaccgggata 960
gctagctact ccttcctaca ggaagccccg ttttactaa aatttcaggt ggttgggag 1020
aaagatagag cctttgcaaa ttagagcagg gttttttatt tttttat 1067

<210> 26

<211> 540

<212> ADN

5 <213> Homo sapiens

<220>

<221> característica diversa

<223> secuencia de STAR26

<400> 26

ccccctgaca agccccagtg tgtgatgttc cccactctgt gtccatgcat tctcattggt 60
caactcccat ctgtgagtga gaacatgcag tgtttggtt tctgtccttg agatagtttg 120
ctgagaatga tggtttccag ctatccat gtccttgcaa aggaagtga cttatccttt 180
tttatggctt catagtattc catggcacat atgtgccaca tttttttaat ccagtctatc 240
attgatggac atttgggttg gttccaagtc tttgetattg tgaatagcac cacaattaac 300
atatgtgtgc atgtatacat ctttatagta gcatgattta taatccttcg ggtatatacc 360
ctgtaatggg atcgtcgggt caaatggtat ttctagttct agatccttga ggaatcacca 420
cactgctttc cacaatgggt gaactaattt acgctcccac cagcagtgta aaagcattcc 480
tatttctcca cgtcctctcc agtatctggt gtttctgac tttttaatga tcatcattct 540

10

<210> 27

<211> 1520

<212> ADN

<213> Homo sapiens

15

<220>

<221> característica diversa

ES 2 384 391 T3

<223> secuencia de STAR27

<400> 27

```

cttggccctc acaaagcctg tggccagga acaattagcg agctgcttat ttgctttgt    60
atccccaatg ctgggcataa tgcctgcat tatgagtaat gccggtagaa gtatgtgttc    120
aaggaccaa gttgataaat accaaagaat ccagagaagg gagagaacat tgagtagagg    180

atagtgacag aagagatggg aacttctgac aagagttgtg aagatgtact aggcaggggg    240
aacagcttaa ggagagtca acaggaccga gctcttgta agccggtgc catggaggct    300
gggtggggcc atggtagctt tcccttcctt ctcaggttca gagtgtcagc cttgaacttc    360
taattcccag aggcatttat tcaatgtttt cttctagggg catacctgcc ctgctgtgga    420
agactttctt ccctgtgggt cgccccagtc ccagatgag acggtttggg tcagggccag    480
gtgcaccgtt ggggtgtgtg ttatgtctga tgacagttag ttactcagtc attagtcatt    540
gagggaggtg tggtaaagat ggagatgctg ggtcacatcc ctagagaggt gttccagtat    600
gggcacatgg gagggctgga aggataggtt actgctagac gtagagaagc cacatccttt    660
aacaccctgg cttttccac tgccaagatc cagaaagtcc ttgtggttcc gctgctttct    720
cctttttttt tttttttttt tttctgagat ggagtctggc tctgtcgccc aggctggagt    780
gcagtggcac gatttcggct cactgcaagt tccgcctcct aggttcatac cattctccca    840
cctcagcctc ccgagtagct gggactacag gcgccaccac acccagetaa ttttttgtat    900
ttttagtaga gacggcgttt caccatgtta gccaggatgg tcttgatccg cctgcctcag    960
cctcccaaag tgctgggatt acagggctga gccaccgccc ccggcctgct ttcttctttc   1020
atgaagcatt cagctggtga aaaagctcag ccaggtggtt ctggaactct tgacctcaag   1080
tgatctgcct gcctcagcct cccaaagtgc tgagattaca ggcatgagcc agtccgaatg   1140
tggtttttt tgttttggtt tgaacaagg tctcactggt gcccaggctg cagtgcagtg   1200
gcatacctca gtcactgc agcctcgacc tctgggctc aagcaatcct cccaactgag   1260
cctccccagt agctggggct acaagcgcac gccaccacgc ctggctattt tttttttttt   1320
ttttttttt gagaaggagt ttcattcttg ttgcccaggc tggagtgcaa tggcacagtc   1380
tcagctcact gcagcctccg cctcctgggt tcaagcgatt ctctgcctc agcctcccga   1440
gtagctggga ttataggcac ctgccaccat gcctggctaa ttttttgta tttttagtag   1500
ggatgggggt tcaccatggt                                     1520

```

5 <210> 28

<211> 961

<212> ADN

<213> Homo sapiens

<220>

ES 2 384 391 T3

<221> característica diversa

<223> secuencia de STAR28

<400> 28

```

aggaggttat tctgagcaa atggccagcc tagtgaactg gataaatgcc catgtaagat      60
ctgtttaccc tgagaagggc atttcctaac tctccctata aaatgccaag tggagcacc      120
cagatgaaat agctgatatg ctttctatac aagccatcta ggactggctt tatcatgacc      180
aggatattca cccactgaat atggctatta cccaagttat ggtaaagctt gtagttaagg      240
gggtcccttc cacatggaca ccccaggta taaccagaaa gggttcccaa tctagactcc      300

aagagagggt tcttagacct catgcaagaa agaacttggg gcaagtacat aaagtgaaag      360
caagtttatt aagaaagtaa agaaacaaaa aaatggctac tccataagca aagttatttc      420
tcacttatat gattaataag agatggatta ttcattgagtt ttctgggaaa ggggtgggca      480
attcctggaa ctgagggttc ctcccacttt tagaccatat aggtatctt cctgatattg      540
ccatggcatt tgtaaactgt catggcactg atgggagtgt ctttagcat tctaatgcat      600
tataattagc atataatgag cagtgaggat gaccagaggt cacttctgtt gccatattgg      660
tttcagtggg gtttggttgg cttttttttt tttttaacca caacctgttt tttatttatt      720
tatttattta tttatttatt tatatttttt attttttttt agatggagtc ttgctctgtc      780
accaggtta gagtgcagtg gcaccatctc ggctcactgc aagctctgcc tccttgggtc      840
acgccattct gctgcctcag cctcccagat agctgggact acaggtgcct gccaccatac      900
ccggctaatt ttttctattt ttcagtagag acgggggttc accgtgttag ccaggatggt      960
c                                                                              961

```

5

<210> 29

<211> 2233

<212> ADN

<213> Homo sapiens

10

<220>

<221> característica diversa

<223> secuencia de STAR29

<400> 29

ES 2 384 391 T3

agcttggaca cttgctgatg ccactttgga tgttgaaggg ccgccctctc ccacaccgct 60
 ggcactttt aaatatgtcc cctctgceca gaagggcccc agaggagggg ctggtgaggg 120
 tgacaggagt tgactgetct cacagcaggg ggttccggag ggacctttc tccccattgg 180
 gcagcataga aggacctaga agggccccct ccaagcccag ctgggcgtgc agggccagcg 240
 attcgatgcc tccccctgac tcaggtggcg ctgtcctaaa ggtgtgtgtg ttttctgttc 300
 gccagggggt ggcggataca gtggagcacc gtgcccaag tgtctgagcc cgtggttaagt 360
 ccctggaggg tgcacggctc cctccgactg tctccatcac gtcaggcctc acagcctgta 420
 ggcaccgctc ggggaagcct ctggatgagg ccattgtgtc atccccctgg agtcctggcc 480
 tggcctgaag aggaggggag gaggaggcca gcccctccct agccccaagg cctgagggc 540
 tgcaagcccg gcccacatt ctagtccagg cttggctgtg caagaagcag attgcctggc 600
 cctggccagg cttcccagct aggatgtggt atggcagggg tgggggacat tgaggggctg 660
 ctgtagcccc cacaacctcc ccaggtaggg tgggaacag taggctggac aagtggacct 720
 gttcccatct gagattcaag agcccacctc tcggaggtg cagtgagccg agatccctcc 780
 actgcactcc agcctgggca acagagcaag actctgtctc aaaaaaacag aacaacgaca 840
 aaaaaaac cacctctggc ccactgccta actttgtaaa taaagtttta ttggcacata 900
 gacacacca ttcatttaca tactgctgcg gctgcttttg cattaccctt gagtagacga 960
 cagaccacgt ggccatggaa gccaaaaata tttactgtct ggccctttac agaagtctgc 1020

ES 2 384 391 T3

tctagagggg gaccccggcc catggggcag gaccactggg cgtgggcaga agggaggcct 1080
 cgggtgcctcc acgggcctag ttgggtatct cagtgcctgt ttcttgcatt gagcaccagg 1140
 ggtcagggca agtacctgga ggagggcaggc tgttgcccgc ccagcactgg gaccaggagg 1200
 acctgagag gctcttaacg aatgggagac aagcaggacc agggctccca ttggctgggc 1260
 ctcagtttcc ctgcctgtaa gtgagggagg gcagctgtga aggtgaactg tgaggcagag 1320
 cctctgctca gccattgcag gggcggetct gccccactcc tgttgtgcac ccagagttag 1380
 gggcacgggg tgagatgtca ccatcagccc ataggggtgt cctcctggtg ccagggtccc 1440
 aagggatgtc ccatcccccc tggctgtgtg gggacagcag agtcctggg gctgggaggg 1500
 ctccacactg ttttgtcagt ggtttttctg aactgttaa tttcagtgga aaattctctt 1560
 tcccctttta ctgaaggaac ctccaaagga agacctgact gtgtctgaga agttccagct 1620
 ggtgctggac gtcgcccaga aagcccaggc actgccacgg gcgcccggca ggggtgtgtc 1680
 tgcgccagcc atgggcacca gccaggggtg tgtctacgcc ggccaggggt aggtctccgc 1740
 cggcctccgc tgctgcctgg ggagggccgt gcctgacct gcaggcccgg tttgtccgcg 1800
 gtcagctgac ttgtagtac cctgcccttg gatggtcgtt acagcaactc tgggtggttg 1860
 ggaaggggcc tcctgattca gcctctgcgg acggtgcgcg aggggtggagc tcccctccct 1920
 ccccaccgcc cctggccagg gttgaacgcc cctgggaagg actcaggccc gggctctgctg 1980
 ttgctgtgag cgtggccacc tctgccctag accagagctg ggccttcccc ggcctaggag 2040
 cagccgggca ggaccacagg gctccgagtg acctcagggc tgcccgaact ggaggccctc 2100
 ctggcgtcgc ggtgtgactg acagcccagg agcgggggct gttgtaattg ctgtttctcc 2160
 ttcacacaga accttttcgg gaagatggct gacatcctgg agaagatcaa gaagtaagtc 2220
 ccgcccccca ccc 2233

<210> 30

<211> 1851

<212> ADN

5 <213> Homo sapiens

<220>

<221> característica diversa

<223> secuencia de STAR30

<400> 30

ES 2 384 391 T3

ggggtgcattt ccacccaggg gacacttggc aatgggtggga gacattgctt gttgtcacaa 60
 ctgggcatgg gagtgcctgct gcgtctagtg ggtagaggcc agagatgctc ctaatatcct 120
 acaaggcaca gaacagcccc ccacaacaga gaattatcca gcctgaaaat gtccacagtg 180
 ctgaggttgg gaaaccctat tctagagcca acaggctgtg aagcttgact catggttcca 240
 tcaccaatag ctgcgtgacc ttggtgagtt ccttagctgc tctgtgcctc ggattcatgg 300
 taggttttcc ttgttaggtt taaatgagtg aagttataca gagggcctga agtctcatgg 360
 tattttacta gagcctcatt gtgttttagt tataattaga aattgggtaa ggtaaggaca 420
 cagaagaagc catctgatct gggggcttca cacttagaag tgacctcgga gcaattgtat 480
 tggggtggaa agggactaac agccaggagc agagggcaca ttggaattgg ggccagaggg 540
 cacagactgc cttgtccatc aggcatagca atggacagag gaaggggaat gactagtatt 600
 ggctgcaagg ccaagtacag gggacttatt tctcatatct atctatctat ctacctaccg 660
 tctatttate tatcatctat ctacttattt atctatctat ttatgcatgt gtaccaaccg 720
 aaagttttag taaatgcaca aactgcgata taatgaaaat ggaaattttc aaaagaagag 780
 aatcacctg ccacctgact accttaacaa atgagtggtt ttcatctctc cttccagggc 840
 tgtcattttt acagtgcttt agtcataaaa caggctcctc attctattgt tttatgtcac 900
 atgaaattgt accataagca ttttccatga tgtgactcca ctgtttcatt ttccattttt 960
 ttccagaatg aagataacct cattgttttt ttccctgattg taaaaatgct ctgtgctctt 1020
 tttttttttt ttaacaatg caggcagtac caaaaagtat gaagaagaat gtaatagttc 1080
 ccatttccca tctcactctt taaggccagc attttgggtga acatccatcc gaacaaatct 1140
 ccacgcgttt atcaatttgt tgacttactc cttcttttat gtaaatatga acatgattta 1200
 actgccagtc catttggaac cttaaagtga aggtttttta ttggtggggt ttgctatggg 1260
 ctgaatatgt gtgtcccccc aaaatttatg ttgaatccta acgcccattg cgattaggag 1320
 gtggggccat taggaggtga ttaagtcag aagtcacag ccctaatgaa tgggatttgt 1380
 ggcttgaaa agggaccca gagagctgcc ttgccccttc tgccatgtaa ggacacagtg 1440
 aggagctagg aagggggcct cagcagagac caaatgtgat ggtgcctcga tattggactt 1500
 cccagcctcc agaatgtgag aatgaattt ctgttgttta taagtcaccc agtctatagt 1560
 attttgttct agcagcccaa acagactaag tcagggttgt tgttttagga agtggggaat 1620
 ggggccatgc atgggtgtac gccagaacaa aggaagccag caagtccctga aagatactgg 1680
 aaaagggaat agtgggcacg tgcagtgtgt tagtttctg aggctgctat aacaaagcac 1740
 cacaggttgg gtggcttaa taacagaaat tcattctccc atcattctgg ggaccagacg 1800
 tctgaaatca agactcctat gccatgctcc ttctgaaggc tccaggggag g 1851

<210>31

<211> 1701

ES 2 384 391 T3

<212> ADN

<213> Homo sapiens

<220>

<221> característica diversa

5 <223> secuencia de STAR31

<220>

<221> característica diversa

<222> (159)..(1696)

<223> n es a, c, g, o t en diversas posiciones

10 <400> 31

cacccgcctt ggccccccag agtgctggga ttacaagtgt aaaccacat tcctggctag 60

atthaatttt ttaaaaaata aagagaagta ggaatagttc attttaggga gagcccctta 120

ES 2 384 391 T3

actgggacag gggcaggaca ggggtgaggc ttcccttant tcaagctcac ctcaaaccca 180
cccaggactg tgtgtcacat tctccaataa aggaaagggt gctgcccccg cctgtgagtg 240
ctgcagtgga gggtagaggg ccgtgggcag agtgcttcat ggactgctca tcaagaaagg 300
cttcatgaca atcggcccag ctgctgtcat cccacattct acttccagct aggagaaggc 360
ggcttgccca cagtacacca gccggcaagt gtcaccctg ggttgaccc agagctatga 420
tcctgccag ggtccagct gagaatcagg cccacgttct aggcagaggg gctcacctac 480
tgggactcca gtagctgtag tgcattggag catcatggct gcagcagcct ggacctggtc 540
tcacactggc tgtccctgtg ggcaggccat cctcaatgcc aggtcaggcc caagcatgta 600
tcccagacaa tgacaatggg gtggaatcct ctcttgtccc agaagccact cctcactgtt 660
ctacctgagg aaggcagggg catggtggaa tcctgaagcc tgcctgtgagg gtctccagcg 720
aacttgacaa tggtcagccc tgccttctcc tcctgaact agattgagcg agagcaagaa 780
ggacattgaa ccagcaccca aagaattttg gggaaaggcc tctcatccag gtcaggctca 840
cctccttttt aaaatttaat taattaatta attaattttt ttttagagac agagtcttac 900
tgtgtggccc aggtctgagt gcagtggcac aatcatagtt cactgcagcc tcaaactccc 960
cacctcagcc tctggattag ctgagactac aggtgcacca ccaccacacc cagctaatat 1020
ttttattttt gtagagagag ggtttcacca tcttgcccag gctggtctca aactcctggg 1080
ctcaagtgat cccgccaggg tctgaaagcc cccaggctgg cctcagactg tggggttttc 1140
catgcagcca cccgagggcg cccccaagcc agttcatctc ggagtccagg cctggccctg 1200
ggagacagag tgaaaccagt ggtttttatg aacttaactt agagttaaaa agatttctac 1260
tcgatcactt gtcaagatgc gccctctctg gggagaaggg aacgtgactg gattccctca 1320
ctgttgatc ttgaataaac gctgctgctt catcctgtgg gggccgtggc cctgtccctg 1380
tgtgggtggg gcctcttcca tttccctgac ttagaaacca cagtccacct agaacagggt 1440
ttgagaggct tagtcagcac tgggtagcgt tttgactcca ttctcgctt tcttctttt 1500
ctttccagga ttttgtgca gaaatggttc tttgttgcc gtgtagtcc tccttggaag 1560
gcagctcaga aggcccgtga aatgtcgggg gacaggacc ccagggaggg aaccccaggc 1620
tacgcacttt agggttcgtt ctccaggagg ggcgacctga cccccgnatc cgtcggngcg 1680
cgnngnacn aannnttcc c 1701

<210> 32

<211> 771

<212> ADN

5 <213> Homo sapiens

<220>

<221> característica diversa

<223> secuencia de STAR32

ES 2 384 391 T3

<400> 32

```

gatcacacag cttgtatgtg ggagctagga ttggaacccc agaagtctgg cccaggttc      60
atgctctcac ccaactgata caatggcctc tcataaatca atccagtata aacattaga      120

atctgcttta aaaccataga attagtagcg taagtaataa atgcagagac catgcagtga      180
atggcattcc tggaaaaagc cccagaagg aattttaaat cagctttcgt ctaatcttga      240
gcagctagtt agcaaatatg agaatacagt tgttcccaga taatgcttta tgtctgacca      300
tcttaaactg gcgctgtttt tcaaaaactt aaaaacaaaa tccatgactc ttttaattat      360
aaaagtgata catgtctact tgggaggctg aggtggtggg aggatggctt gagtttgagg      420
ctgcagtatg ctactatcat gcctataaat agccgctgca ttccagcttg ggcaacatac      480
ccaggcccta tctcaaaaaa ataaaaagta atacatctac attgaagaaa attaatttta      540
ttgggttttt ttgcattttt attatacaca gcacacacag cacatatgaa aaaatgggta      600
tgaactcagg cattcaactg gaagaacagt actaaatcaa tgtccatgta gtcagcgtga      660
ctgaggttgg tttgtttttt cttttttctt ctctttctt ctcttttctt tttttttgag      720
acggagcttt gctctttttg cccaggcttg attgcaatgg cgtgatctca g              771

```

<210> 33

5 <211> 1368

<212> ADN

<213> Homo sapiens

<220>

<221> característica diversa

10 <223> secuencia de STAR33

<400> 33

ES 2 384 391 T3

gcttttatcc tccattcaca gctagcctgg cccccagagt acccaattct ccctaaaaaa 60
 cggtcacgct gtatagatgt gtgtggcttg gtagtgctaa agtggccaca tacagagctc 120
 tgacaccaaa cctcaggacc atgttcatgc cttctcactg agttctggct tgttcgtgac 180
 acattatgac attatgatta tgatgacttg tgagagcctc agtcttctat agcactttta 240
 gaatgcttta taaaaccat ggggatgtca ttatattcta acctgtagc acttctgttc 300
 gtattacca tcacatccca acatcaattc tcatatatgc aggtacctct tgtcacgcgc 360
 gtccatgtaa ggagaccaca aaacaggctt tgtttgagca acaaggtttt tatttcacct 420
 gggcagcagg gggctgagtc tgaaaagaga gtcagtgaag ggagacaggg gtgggtccac 480
 ttataagat ttgggtaggt agtggaaaat tacaatcaaa gggggttgtt ctctggctgg 540
 ccagggtggg ggtcacaagg tgctcagtgg gagagccttt gagccaggat gagccagaag 600
 gaatttcaca aggtaatgct atcagttaag gcagggactg gccattttca cttcttttgt 660
 ggtggaatgt catcagttaa ggcaggaacc ggccattttc acttcttttg tgattcttca 720
 cttgcttcag gccatctgga cgtataggtg caggtcacag tcacagggga taagatggca 780
 atggcatagc ttgggctcag aggcctgaca cctctgagaa actaaagatt ataaaaatga 840
 tggtcgcttc tattgcaaat ctgtgtttat tgtcaagagg cacttatttg tcaattaaga 900
 acccagtggt agaatcgaat gtccgaatgt aaaacaaaat acaaacctc tgtgtgtgtg 960
 tgtgtgtgag tgtgtgtgta tgtgtgtgtg tgtgtattag agaggaaaag cctgtatttg 1020
 gaggtgtgat tcttagatcc taggttcttt cctgcccacc ccatatgcac ccaccccaca 1080
 aaagaacaaa caacaaatcc caggacatct tagcgcaca tttcagtttg catattttac 1140
 atatttactt ttcttacata ttaaaaaact gaaaatttta tgaacacgct aagttagatt 1200
 ttaaattaag tttgttttta cactgaaaat aatttaatat ttgtgaagaa tactaataca 1260
 ttggtatatt tcattttctt aaaattctga acccctcttc ccttatttcc ttttgacccg 1320
 attggtgtat tggtcacgct actcatggat ttgccttaag gcaggagg 1368

<210> 34

<211> 755

5 <212> ADN

<213> Homo sapiens

<220>

<221> característica diversa

<223> secuencia de STAR34

10 <400> 34

ES 2 384 391 T3

actgggcacc ctctaggca ggggaatgtg agaactgccg ctgctctggg gctgggcgcc 60
 atgtcacagc aggagggagg acggtgttac accacgtggg aaggactcag ggtggtcagc 120
 cacaaagctg ctgggtgatga ccaggggctt gtgtettcac tctgcagccc taacacccag 180
 gctgggttcg ctaggctcca tcttgggggt gcagaccctg agagtgatgc cagtgggagc 240
 ctcccgcccc tccccttctc cgaaggccca ggggtcaaac agtgtagact cagaggcctg 300
 agggcacatg tttatttagc agacaaggtg gggctccatc agcgggggtg cctggggagc 360
 agctgcatgg gtggcactgt ggggagggtc tcccagctcc ctcaatggtg ttcgggctgg 420
 tgcgpcagct ggcggcaccc tggacagagg tggatatgag ggtgatgggt ggggaaatgg 480
 gaggcacccg agatggggac agcagaataa agacagcagc agtgctgggg ggccaggggga 540
 tgagcaaagg caggcccaag acccccagcc cactgcaccc tggcctcca caagccccct 600
 cgcagccgcc cagccacact cactgtgcac tcagccgtcg atacactggt ctgttaggga 660
 gaaagtccgt cagaacaggc agctgtgtgt gtgtgtgcgt gtatgagtgt gtgtgtgtga 720
 tcctgactg ccaggtcctc tgcactgccc ctggg 755

<210> 35

<211> 1193

<212> ADN

5 <213> Homo sapiens

<220>

<221> característica diversa

<223> secuencia de STAR35

<220>

10 <221> característica diversa

<222> (312)..(1191)

<223> n es a, c, g, o t en diversas posiciones

<400> 35

cgacttgggtg atgcgggctc ttttttgggt ccatatgaac tttaaagtag tcttttccaa 60
 ttctgtgaag aaagtcattg gtaggtgat ggggatggca ttgaatctgt aaattacctt 120

ES 2 384 391 T3

gggcagtatg gccattttca caatggtgat tcttcctatc catgatgatg gaatgttctt 180
 ccattagttt gtatcctctt ttatttcctt gagcagtggt ttgtagttct ccttgaagag 240
 gtccttcaca tcccttgtaa gttggattcc taggtatgtt attctctttg aagcaaattg 300
 tgaatgggag tncactcacg atttggctct ctgtttgtct gctgggtgta taaanaatgt 360
 ngtgatnttn gtacattgat ttngtatccn tgagacttng ctgaatttgc ttnatcnget 420
 tnnnggaacc ttttgggctg aaacnatggg attttctaaa tatacaatca tgtcgtctgc 480
 aaacagggaa caatttgact tctcttttc ctaattgaat acactttatc tccttctcct 540
 gcctaattgc cctgggcaaa acttccaaca ctatgntngn aataggagnt ggtgagagag 600
 ggcacccctg ttcttgttgc cagnttttca aagggaaatgc ttccagtttt ggcccattca 660
 gtatgatatg ggctgtgggt ngtgtcataa atagctctta tnattttgaa atgtgtccca 720
 tcaataccta atttattgaa agtttttagc atgaangcat ngttgaattt ggtcaaaggc 780
 tttttctgca tctatggaaa taatcatgtg gtttttgtct ttggctcntg tttatatgct 840
 ggatnacatt tattgatttg tgtatatnga acccagcctn ncatcccagg gatgaagccc 900
 acttgatcca agcttggcgc gcngnctagc tcgaggcagg caaaagtatg caaagcatgc 960
 atctcaatta gtcagacccc atagtccgcc cctacctccg cccatccgcc cctaactcng 1020
 nccgttcgcc cattctcgcc catggetgac taatnttttt annatccaag cggngccgcc 1080
 ctgcttganc attcagagtn nagagnnttg gagccnagc cttgcaaac tccggacnng 1140
 ttctnnggat tgaccccnnt taaatatttg gtttttgn tttcannng nga 1193

<210> 36

<211> 1712

<212> ADN

5 <213> Homo sapiens

<220>

<221> característica diversa

<223> secuencia de STAR36

<400> 36

ES 2 384 391 T3

gatcccatcc ttagcctcat cgatacctcc tgctcacctg tcagtgcctc tggagtgtgt 60
gtctagccca ggcccatccc ctggaactca ggggactcag gactagtggg catgtacact 120
tggcctcagg ggactcagga ttagtgagcc ccacatgtac acttggcctc agtgagactca 180
ggactagtga gccccacatg tacacttggc ctcaggggac tcaggattag tgagccccc 240
catgtacact tggcctcagg ggactcagga ttagtgagcc ccacatgtac acttggcctc 300
aggggactca ggactagtga gccccacatg tacacttggc ctcaggggac tcagaactag 360
tgagcccccac atgtacactt ggcttcaggg gactcaggat tagtgagccc cacatgtaca 420
cttggacacg tgaaccacat cgatgtgctg cagagctcag ccctctgcag atgaaatgtg 480
gtcatggcat tccttcacag tggcaccctc cgttccctcc ccacctcacc tcccattctt 540
gtctgtcttc agcacctgcc atgtccagcc ggcagattcc accgcagcat cttctgcagc 600

acccccgacc acacacctcc ccagcgcctg cttggcctc cagcccagct cccgcctttc 660
ttccttgggg aagctccctg gacagacacc cctcctccc agccatggct ttttctgct 720
ctgccccacg cgggaccctg ccttgatgt gctacaatag acacatcaga tacagtcctt 780
cctcagcagc cggcagacc aggttgact gctcggggcc tgctgtgag gtcacacagg 840
tgtcgtaac ttgccatctc agcaactagt gaatatggc agatgctacc ttccttcgg 900
ttccctggtg agaggtactg gtggatgtcc tgtgttgccg gccacctttt gtccttggat 960
gccatttatt tttttccaca aatatttccc aggtctcttc tgtgtgcaag gtattagggc 1020
tgcagcgggg gccaggccac agatctctgt cctgagaaga cttggattct agtgaggag 1080
actgaagtgt atcacaccaa tcagtgtaaa ttgttaactg ccacaaggag aaaggccagg 1140
aaggagtggg gcatggtggt gttctagtgt tacaagaaga agccaggag ggcttctg 1200
atgaagtggc atctgacctg ggatctggag gaggagaaaa atgtcccaa agagcagaga 1260
gccacccta ggctctgcac caggaggcaa cttgctggc ttatggaatt cagagggcaa 1320
gtgataagca gaaagtcctt gggggccaca attaggattt ctgtcttcta aaggcctct 1380
gccctctgct gtgtgacctt gggcaagtta cttcacctct agtgctttgg ttgcctcacc 1440
tgtaaagtgg tgaggataat gctatcacac tggttgagaa ttgaagtaat tattgtgca 1500
aagggttat aagggtgtct aatactagta ctagtaggta cttcatgtgt cttgacaatt 1560
ttaatcatta ttatttgtc atcaccgtca ctcttcagg ggactaatgt cctgtctgtt 1620
ctgtccaaat taaacattgt ttatccctgt gggcatctgg cgagggtggct aggaaagcct 1680
ggagctgttt cctgttgacg tgccagacta gt 1712

<210> 37

<211> 1321

5 <212> ADN

<213> Homo sapiens

ES 2 384 391 T3

<220>

<221> característica diversa

<223> secuencia de STAR37

<400> 37

```

aggatcacat ttaaggaagt gtgtggggtc cctggatgac accagcacc accagcacc agtgcggctc    60
tgtctggcaa ccgctcccaa ggtggcagga gtgggtgtcc cctgtgtgtc agtgggcagc    120
tcctgctgag cctacagctc actggggagc ctgacagcgg ggccatgtgc ctgacactcc    180
tctctgcttg tggacctggc aaggcagga gcagaaaaca gagccacttg aaggctttct    240
gtctgcgtct gtgtgcagtg tggatttagt tgtgcttttt tcttgctggg agagcacagc    300
caccatttac aagcagtgtc accctcatgg gtggcgagga cagaacagga gcctctgctc    360
tctgtacctt tctgggcccg gtgggctccc ttgtcctggc ttccatctct gtctcagcga    420
ccattcagcc ctgctcagga acacatgctt cttagaaaag ccaaattcag cccttgctc    480
tgctcctctt ggtctcatga tgtgcatctg ttaccttgaa actggaaacc agtctatcaa    540
tgtctgtgcc aattttttat tccctcccca acctccttcc ccatacgaact ttttatttat    600

```

5

```

gtaggatgtg tgctgtctaa tgatgggatg accacatttt tccatgttct aaaagtgtc    660
ctctcccgca gggctccagg gctgggtggtt gctttgggtc tacagctacg tcttaccgcg    720
ctctcgctc aacagcctgt gtgggtggca agccgggtgtg gggctgggga acgcagcgtt    780
ctccaggagg gggacctggc tctccttctg cagtgcaggc gaaggcctag atgccagtgt    840
gacctccac aaggcgtggc ttccagactc cccggetgga agtgatgctt ttttgcctcc    900
ggccctgggt ttgaagcagc ctggctttct cttggtaagt ggctgggtgc ttagcagctg    960
caatctgagc tcagccacct acacaccacc gtggccgaca ctttcattaa aaagtttct    1020
gagacgactt gcgtgcatgt tgacttcatg atcagcgccg ctgggaagaa cccctgagcc    1080
ggtgggggtg ggtggaagc agcaggtgca gtgatggggc tgggtgcca ggaggcctca    1140
gtgctcaatc aggccaaagt ggccaagccc aggctgcagg gaaggccggc ctgggggttg    1200
tgggtgagca caggcaggca ccagctgggc agtgttagga tgctggagca gcatccgtaa    1260
ccccactgag tgggtagtc tggttggggc agggaccgct gttgctttgg cagagagaga    1320
t                                                                                   1321

```

<210> 38

<211> 1445

<212> ADN

<213> Homo sapiens

<220>

<221> característica diversa

<223> secuencia de STAR38

10

ES 2 384 391 T3

<220>

<221> característica diversa

<222> (348)..(949)

<223> n es a, c, g, o t en diversas posiciones

5

<400> 38

```

gatctatggg agtagettcc ttagtgagct ttccttcaa atactttgca accaggtaga      60
gaatthttgga gtgaagggtt tgtttcttctg ttcttcacaa tatggatatg catcttcttt    120
tgaaaatggt aaagtaaatt acctctcttt tcagatactg tcttcatgcy aacttggtat    180
cctgthttcca tcccagcctt ctataacca gtaacatctt ttttgaaacc agtgggtgag    240
aaagacacct ggtcaggaac gcggaccaca ggacaactca ggctcaccca cggcatcaga    300
ctaaaggcaa acaaggactc tgtataaagt accggtggca tgtgtatnag tggagatgca    360
gcctgtgctc tgcagacagg gagtcacaca gacacttttc tataatttct taagtgcttt    420
gaatgttcaa gtagaaagtc taacattaaa tttgattgaa caattgtata ttcattggaat    480
atthttggaac ggaataccaa aaaatggcaa tagtggttct ttctggatgg aagacaaact    540
thttctgttt aaaataaatt ttatthttata tthttgaggt tgaccacatg accttaagga    600
tacatataga cagtaaactg gttactacag tgaagcaaat taacatatct accatcgtac    660
atagttacat thttttgtgt gacaggaaca gctaaaaatc acgtatthta caaaaatcct    720
aaagacaata cattthttatt aactatagcc ctcatgatgt acattagatc gtgtggttgt    780

thtttccgtc cccgccacgc ctctctctctg ggatggggat tcattcccta gcaggtgtcg    840
gagaactggc gcccttgacg ggtaggtgcc ccggagcctg aggcgggnac thtaanatca    900
gacgcttggg ggccggctgg gaaaaactgg cggaaaatat tataactgna ctctcaatgc    960
cagctgttgt agaagctcct gggacaagcc gtggaagtcc cctcaggagg ctctccgat    1020
gtcctaggtg gctgctccgc ccgccacggt catttccatt gactcacacg cgcgcctgg    1080
aggaggaggc tgcgctggac acgccggtg cgcctttgcc tgggggagcg cagcctggag    1140
ctctggcggc agcgtggga gcggggcctc ggaggctggg cctggggacc caaggttggg    1200
cggggcgag gaggtgggct cagggttctc cagagaatcc ccatgagctg acccgcaggg    1260
cggccgggcc agtaggcacc gggccccgc ggtgacctgc ggaccgaag ctggagcagc    1320
cactgcaaat gctgcgctga ccccaaatgc tgtgtccttt aatgtthta attaagaata    1380
attaataggt ccgggtgtgg aggctcaagc cttaatcccc agcacctggc gaggccgagg    1440
agggg                                           1445

```

<210> 39

<211> 2331

10

<212> ADN

ES 2 384 391 T3

<213> Homo sapiens

<220>

<221> característica diversa

<223> secuencia de STAR39

5

<400> 39

```

gtgaaataga tcactaaagc tgattcctct tgtctaaatg aaactttcta ccctttgatg      60
gacagctatg ctttccccat cctctcccgt cccccagccc ttgtaacca tcctcctact      120
ctctacttgt aggagtcaa cttgtttaga ttttgtgagt gagaacatgt ggtatttgcc      180
tttagagtcc tctaggttta tccatattgt gttaaagac aggattccct gcctttttaa      240
ggctgaatag tatttcattg taatatatat acatacacac acacataac acacacatat      300
atatacatat atacatatat gtacatagat acatatatat gtacatatat acacacacat      360
atacacacat atatacacat atatacatat acatatatac acatatatgt acatatatat      420
aacttttttt catttatcca ttcacttaat acatatgatg gagggcttta tatatgccag      480
gctctgtgat gaatgctgga aattcaatag tgagaaagac tcagtctctg cctccaaaga      540
gcatcatggg ctaggtgctg caacgaggaa ttgccactg ttgtcatgag agcacagaga      600
agggactcaa ccagccttga agaatcaggg gaggcttcta agctaattgg gtgtgcctgg      660
ggatcacatt gtttcaagca gcagtaacag gatgtgctca ggtccagatg tgagagagag      720
agagagcata tgtcttcaag aaactaacag tagctcccta tagctgaagc aggagtacaa      780
aatagtgagt ttaagtgatg aggcaagaga tatgaagaag cttgaccatg cagctacacc      840
gggcagcatg ccctctgaga catctcatgg aagccggaaa tgggagtgcc ttgataccaa      900
gccagagaaa ttataaact aagtagatag actgagcagc actcctcctg ggaagaatga      960

```

ES 2 384 391 T3

gacaagccct gaatttggag gtaagttgtg gatttggtgat tagaggagag gtaacaggca 1020
ccaagcaag aaatagtatt gatgcaaagc tgaggttaat tggatgacaa aatgaagagc 1080
ataaggggct cagacacaga ctgagcagaa aacgagtagc atctgaacct agattgagtt 1140
actaatggat gagaaagagt tcttaaagtt gatgaccacg ggatccatat ataagaatgt 1200
ccaatctccc caaattgatc cacgagttca gtgcaatgcc aatcaaaatc ccactaacia 1260
gtttatttta aaatgtaaat gaaaatacaa aatttttaa aagcaaagca atattgaaaa 1320
cccaggaaaa attaggagga cttacacaac ctgatctcaa aacttaccat tatcaagaca 1380
gagtgttatt gacacaagga gagacaaata gataaacgga atgtggtagt ctggagatgc 1440
accacatgt atgtgggtcaa ttgatttttg gccaaaggcac caagtcaatt caaaggagca 1500
aggaaagtag tacagaaaca accaaatatt gttttggaaa ataatgacaa agggcttata 1560
accagaatat aagcatataa atataattct ttcaaatcaa taataagaag gcaaatatct 1620
aataaaaatg agcaaagact tgaaaagtca cttaaaaagg cttattaatt agaaatatgc 1680
aaatgttatt agtcttcagt ggaatttaca ttaaaccaca agggatacta ttatatctta 1740
tgcccactag aataaccaa ggaaaaaaga cagacaaaac aaaatgctgg tgaggatgtg 1800
aagcaactgg aactctcata cattattggt ggtaatgtaa aatttataca accattatga 1860
ataaaggttt ggcagtttct tacaaagttg aatgcacttc tccacgatga ctaggctttt 1920
cactcatagg cgtctggctc cctagaactg aaaacatatg ttcacaagaa gacttgcaaa 1980
tatatattct cccacgtcag gagatatttg ctatgcattt aactgacata agattagtgc 2040
tagagtttat aatgaggttc ttcaaatcta aaagaaaatg caaagcatat aatagtaagg 2100
ggtgcaggcc aggcgcagtg gctcactctg taatcccagc actttgggag gccgaggtgg 2160
gcggatcaca aggtcaggag ttcgagacca acctggccaa catagtgaaa ccctgtctct 2220
actaaaaata caaaaactag ccaggtgctg tgtcatgcac ctgtagtccc agctactcgg 2280
gaggccgagg caggagaatc acttgaacct gggaggtgga ggttgcagtg a 2331

<210> 40

<211> 1071

<212> ADN

5 <213> Homo sapiens

<220>

<221> característica diversa

<223> secuencia de STAR40

<400> 40

ES 2 384 391 T3

gctgtgattc aaactgtcag cgagataagg cagcagatca agaaagcact ccgggctcca 60
 gaaggagcct tccaggccag ctttgagcat aagctgctga tgagcagtga gtgtcttgag 120
 tagtgttcag ggcagcatgt taccattcat gcttgacttc tagccagtgt gacgagagggc 180
 tggagtcagg tctctagaga gttgagcagc tccagcctta gatctcccag tcttatgcgg 240
 tgtgcccatt cgetttgtgt ctgcagtccc ctggccacac ccagtaacag ttctgggatc 300
 tatgggagta gcttccttag tgagctttcc cttcaaatc tttgcaacca ggtagagaat 360

 tttggagtga aggttttgtt cttcgtttct tcacaatag gatatgcac ttcttttgaa 420
 aatgttaaag taaattacct ctcttttcag atactgtctt catgcgaact tggatcctg 480
 tttccatccc agccttctat aaccagtaa catctttttt gaaaccagtg ggtgagaaaag 540
 acacctggtc aggaacgcgg accacaggac aactcaggct caccacggc atcagactaa 600
 aggcaaaaaa ggactctgta taaagtaccg gtggcatgtg tattagtgga gatgcagcct 660
 gtgctctgca gacagggagt cacacagaca cttttctata atttcttaag tgctttgaat 720
 gttcaagtag aaagtctaac attaaatttg attgaacaat tgtatattca tggaaatatt 780
 tggaaacggaa taccaaaaaa tggcaatagt ggttctttct ggatggaaga caaacttttc 840
 ttgtttaaaa taaattttat tttatatatt tgaggttgac cacatgacct taaggataca 900
 tatagacagt aaactgggta ctacagtga gcaaattaac atatctacca tcgtacatag 960
 ttacattttt ttgtgtgaca ggaacagcta aaatctacgt atttaacaaa aatcctaaag 1020
 acaatacatt tttattaact atagccctca tgatgtacat tagatctcta a 1071

<210> 41

<211> 1135

5 <212> ADN

<213> Homo sapiens

<220>

<221> característica diversa

<223> secuencia de STAR41

10 <400> 41

ES 2 384 391 T3

cgtgtgcagt ccacggagag tgtgttctcc tcctcctcgt tccggtgggt gtggcgggaa 60
 acgtggcgct gcaggacacc aacatcagtc acgtatttca ttctggaaaa aaaagtagca 120
 caagcctcgg ctggttccct ccagctctta ccaggcagcc taagcctagg ctccattccc 180
 gctcaaggcc ttectcaggg gcttgcctcac cacaggagct gttcccatgc agggactaag 240
 gacatgcagc ctgcatagaa accaagcacc caggaaaaca tgattggatg gagcgggggg 300
 gtgtggctct tagccttgtc cacctccggt cctcatgggt ctcacacctc ctgagaatgg 360
 gcaccgcaga ggccacagcc catacagcca agatgcagaga ctccgtaagt gacagggatc 420
 cacagcagag tgggtgaaat gttccctata aactttacaa aattaatgag ggcaggggga 480
 ggggagaaat gaaaatgaac ccagctcgca gcacatcagc atcagtcact aggtcggcgt 540
 gctctctgac tgcttccctc tagctgcttg gtgtctcatt gcctcagaag catgtagacc 600
 ctgtcacaag attgtagttc ccctaactgc tccgtagatc acaacttgaa ccttaggaaa 660
 tgctgttttc cctttgagat attcctttgg gtcctgtata ctgatggagc tactgactga 720
 gctgctccga aggacccac gaggagctga ctaaaccaag agtgcagttt gtacaccctg 780
 atgattacat ccccttgcc ccaccaatca actctcccaa tttccagcc cctcaccctc 840
 cagtcccctt aaaagcccca gcccaggccg ggcacagtgg ctcctgcctg taatcccagc 900
 actttgggag gcccaagggtg gcagatcacc tgagggcagg aatttgagac cagcctgacc 960

 aacatgaaga aaccccgctc ctattacaaa tacaaaatta gccgggctg ttgctgcata 1020
 ctggtaatcc cagctacttg ggaggggtgag gcaggagaat cacttgaatc tgggagggcg 1080
 aggttgcgat gagccgagac agcgcattg cactgcagcc tgggcaacaa gagca 1135

<210> 42

<211> 735

5 <212> ADN

<213> Homo sapiens

<220>

<221> característica diversa

<223> secuencia de STAR42

10 <400> 42

ES 2 384 391 T3

aagggtgaga tcactagga gggaggaagg agctataaaa gaaagaggtc actcatcaca 60
 tcttacacac tttttaaacc cttgggtttt taatgtccgt gttcctcatt agcagtaagc 120
 cctgtggaag caggagtctt tctcattgac caccatgaca agaccctatt tatgaaacat 180
 aatagacaca caaatgttta tcggatattt attgaaatat aggaattttt cccctcacac 240
 ctcatgacca cattctggtt cattgtatga atgaatatac cataatttta cctatggctg 300
 tatatttagg tcttttcgtg caggctataa aaatatgtat gggccggtca cagtgactta 360
 cgcccgtagt cccagaactt tgggaggccg aggcgggtgg atcacctgag gtcgggagtt 420
 caaaaccagc ctgaccaaca tggagaaacc ccgtctctgc taaaaataca aaaattaact 480
 ggacacggtg gcgtatgcct gtaatcccag ctactcggga agctgaggca ggagaactgc 540
 ttgaaccagc gaggcggagg ttgtggtgag tcgagattgc gccattgcac tccagcctgg 600
 gcaacaagag cgaaattcca tctcaaaaaa aagaaaaaag tatgactgta tttagagtag 660
 tatgtggatt tgaaaaatta ataagtgttg ccaacttacc ttagggttta taccatttat 720
 gaggggtgctg gtttc 735

<210> 43

<211> 1227

<212> ADN

5 <213> Homo sapiens

<220>

<221> característica diversa

<223> secuencia de STAR43

<400> 43

caaatagatc tacacaaaac aagataatgt ctgcccattt ttccaaagat aatgtggtga 60
 agtgggtaga gagaaatgca tccattctcc ccaccaacc tctgctaaat tgtccatgct 120
 acagtactga gaccaggggg cttattccca gcgggcagaa tgtgcaccaa gcacctcttg 180
 tctcaatttg cagtctaggc cctgctattt gatggtgtga aggcttgac ctggcatgga 240
 aggtccgttt tgtacttctt gctttagcag ttcaaagagc agggagagct gcgagggcct 300
 ctgcagcttc agatggatgt ggtcagcttg ttggaggggc cttctgtggt ccattatctc 360
 cagccccct gcggtgttgc tgtttgcttg gcttgtctgg ctctccatgc cttgttggt 420

10

ES 2 384 391 T3

ccaaaatgtc atcatgctgc accccaggaa gaatgtgcag gcccattctct tttatgtgct 480
 ttgggctatt ttgattcccc gttgggtata ttccctaggt aagaccaga agacacagga 540
 ggtagttgct ttgggagagt ttggacctat gggtaggagg taatagacac agtatcttct 600
 ctttcatttg gtgagactgt tagctctggc cgcggactga attccacaca gctcacttgg 660
 gaaaacttta ttccaaaaca tagtcacatt gaacattgtg gagaatgagg gacagagaag 720
 aggccctaga tttgtacatc tgggtgttat gtctataaat agaatgcttt ggtggccaac 780
 tagacttggt catggtgaca tttagtcttg ccttttcggg ggtgatttaa aaattatgta 840
 tatcttgttt ggaatatagt ggagctatgg tgtggcattt tcatctggct ttttgtttag 900
 ctccagccgt cctgttatgg gcagccttga agctcagtag ctaatgaaga ggtatcctca 960
 ctccctccag agagcgggcc cctcacggct cattgagagt ttgtcagcac cttgaaatga 1020
 gtttaaactt gtttattttt aaaacattct tggttatgaa tgtgcctata ttgaattact 1080
 gaacaacctt atggttgtga agaattgatt tggtgctaag gtgtataaat ttcaggacca 1140
 gtgtctctga agagttcatt tagcatgaag tcagcctgtg gcaggttggg tggagccagg 1200
 gaacaatgga gaagctttca tgggtgg 1227

<210> 44

<211> 1586

<212> ADN

5 <213> Homo sapiens

<220>

<221> característica diversa

<223> secuencia de STAR44

<400> 44

ES 2 384 391 T3

cacctgcctc agcctcccaa agtgctgaga ttcaaagaaa ttttcatgga gaggggacag 60
atggagtcaa ttcttgtggg gtgaacatga gtaccacagt tagactgagg ttgggaaaga 120
ttttccagac aattggaaga gcatgtgaaa gacacagatt ttgagaaatg ttaagtctag 180
ggaactgcaa ggcttttggc acaagaaagc cactgtagac tatagaggca ggatgcctag 240
attcaaatcc caactgctac acttctaagc tttgtaattt tggcaagttt ttaccctcta 300
ttttcttata tataaaatat agattttata tatatagata tagatatata gatagataat 360
aattgtgcat gcctaataaa gttgtcaaag attaaatggt atatgtgaag tattttgtac 420
ggatgatagga acccaggaag ggctctatga atattatgta ttattattat tctaaagtag 480
ctggaataca atgttcaaag gagatagtgg caggagataa gtttgaattg aaagattgag 540
gccagaacat aaagtgcctc ctatattata ttttacataa ttggaacatc attgaaaaat 600
ttaagtatta tttatgtgtg tatgtgtgtt ttatataatt aattctagtt catcatttta 660
aaatatcttt ctgatgtcac tgtgaacaac agatgagaag aagtgaatcc tgagttaagg 720
agaccagctc tctgattact gccataatcc agggagggta ccataaggat ttcaactgga 780
agtgaatcca tcatgatgga gaggaaggac agggctgaaa aatacttagg aagtagtacc 840
agtaggactg gttaagagag agcagaggca ggctacaggg gttggagggtg tcaatcacag 900

agatagggaa aatgggagga gaagcaggct ttgaaaaagt ggcttgtctt gtaaaattat 960
gtgctgttaa aacagtacaa gaaattaata tattcaatcc caaaatacag ggacaattct 1020
ttttgaaaga gttaccaga tagtcttctt tgaagttttc agttaaagaa atttcttggt 1080
aacaataat gtagtcatag aagaaaacac ttaaaacttt attgaataaa gctaataaat 1140
catttaatat aatttatagg aaattgttac ataacacaca cattcaatac tttttgctaa 1200
agtataaatt aatggaagga gagcacgcac acagagggtg aattatggtt atgactttat 1260
tagtcaagaa taaaaattg agtagctaca tcaagcagaa gcacatgctt tacaatccag 1320
cacagaatcc cttgacatcc aaactccoga aacagacatg taaatacaga tgacattgtc 1380
agaacaaaat agggctctac ccgacctata atgttctttt cttgatataa atatgcacat 1440
gaattgcata cggtcatatg gttccaatta ccattatttc ctctgggctt agctatccat 1500
ctaaggggaa tttacaccaa cactgtactt ctacttgcaa gaatatatga aagcatagtt 1560
aacttctggc ttaggacccc aactca 1586

<210> 45

<211> 1981

5 <212> ADN

<213> Homo sapiens

<220>

<221> característica diversa

ES 2 384 391 T3

<223> secuencia de STAR45

<400> 45

atggatcata	gggtaaataa	atataaatt	tcttgagaaa	gcttcgtact	gttttccaag	60
atggctgtac	taatttccat	tcctaccaac	agtgtacagg	gtttcttttt	ctccacatcc	120
tcaccaacac	ttatcttcca	tcttttttta	taatagccct	agtaaaatgt	gtgaggtgat	180
atctcattgt	ggcattgatt	tgcacttctc	tgataattag	gaatgtttat	gattttttca	240
tgtacctggt	tggccttttg	tatgatgtag	gaaatgtcta	ttctgattct	ttgcttattt	300
tttaataagc	atagtttttt	tcttattttt	gagtaggttg	agttgcttat	atattattat	360
atgagccct	tacctgatgt	atggtttaaa	aatattatcc	catttgtggg	ttctcttaat	420
tctatcattg	cttcttttcc	tgtggaaaag	ttttaagttt	tatgcagtct	catttgtgtg	480
ttttgctttt	gttgcctttt	ggaataatct	acagaaaatc	atagctcagg	ccaatgtcat	540
acagtctcct	tctatatttc	cttgtagtag	ttttacattt	aaactttaat	tttgatttga	600
tgcttgata	aagagcaaaa	taaaagtcaa	attttattct	tctgtatgtg	gatagtcagt	660
tttgtctaca	ccatttattg	aaaataattt	tctttcttca	ctgtgtattt	ttagttattt	720
tatcaaaaaa	tcaattgacc	acagacacac	ggatttattt	acaggttcta	tatccctttg	780
tactgtttta	catgtctggt	tttatgccat	tgctatgctg	ttttaattcc	tatagctttg	840
taatagagtt	tggagtcagg	tagtctgatg	cctccagctt	tgttcttttt	gttcaagatt	900
gctttgggtg	gtccaggctc	tttgtggttc	catacaaatt	ttagcagtaa	tttttctatt	960

ES 2 384 391 T3

tctgtgaaga atgacattgg aatttgatag tggttgcatt taatctgtag attgctttgg 1020
 gtagcattga cacttttaca atactaattt ttgaatccat caatgaagga tgtttctcca 1080
 tttatttatg ccattttaat ttttttcac aatgtgctat agttttcagt atgtaaactc 1140
 tttatggttt tgattaaatt tactcctgtc ttttatatat ttatatatct gttttgatcc 1200
 tattataaat tgaattgcct ttatttttca ggtaatagtt tgcattagtt taatagaaac 1260
 aataatgata tttgtatggt gattttgtaa ctattaactt tattgaattt cttcatcagc 1320
 tataaccatt ttttttggty gaacttttaa gattttctct atcttaagat tataatttca 1380
 aaaaacagaa acaatcttac ctcttccttc cctatgtgga tttcttttac gtctttgtct 1440
 tgtgtaactg ttctggctag gcaattacac ataatgtttt catcatttat aattttacat 1500
 cacatccatc tattgtggca cattgattgc tacttttcaa gttgtaaacc tggacattta 1560
 tcactactct tcctccaata caggagtcca tggcgtggtg tgggccctac tgtgccacag 1620
 tccagggcac ggctgggctg aggttctctt gtgcaagagt ccgtggctct gcggagcaag 1680
 agttctccag tgccttagtc cagggtagg caggggtggg gctccttcag tagcttagtc 1740
 cagtgcgccg ccctgcgagg gtccctctga gcaggagtac acgatgagge aggtcctac 1800
 tgtgccttag cccaggaage ggggggctgg gtccctctgt gccatagtcc aggctgccgg 1860
 gagctgggtc ctctggtgcc atagctcagg ccggcgggag ctgggtcctc tggtgccgta 1920
 gtccagggty cagcagaaca ggagtcctgc ggagcagtag tccagggcac gctggggcgt 1980
 g 1981

<210> 46

<211> 1859

<212> ADN

5 <213> Homo sapiens

<220>

<221> característica diversa

<223> secuencia de STAR46

<400> 46

ES 2 384 391 T3

attgttttc tgcacctct gcattttctg caaattctgt tgaatcattg cagtactta 60
 ggtttgcttc gtctcccca ttacaaacta cttactgggt tttcaacc tagttccctc 120
 atttttatga tttatgctca tttctttgta cacttcgtct tgcctcatct cccaactcat 180
 ggcccctggc tttggattat tgttttggtc ttttattttt tgtcttcttc tacctcaaca 240
 cttatcttcc tctcccagtc tccggtacc tatcaccaag gttgtcatta acctttcata 300
 ttattctca ttatccatgt attcatttgc aaataagcgt atattaacaa aatcacaggt 360
 ttatggagat ataattcaca taccttaaaa ttcaggcttt taaagtgtac ctttcatgtg 420
 gtttttggtat tattcacaaa gttatgcatt gatcaccacc atctgattcc ataacatgtt 480
 caatacctca aaaagaagtc tgtactcatt agtagtcatt tcacattcac cactccctct 540
 ggctctgggc agtactgat ctttgtgtct ctatggattt gcctagtcta ggtattttta 600
 tgtaaagtc atcatacaac atgtgacctt ttgtttggct ttttccattt agcaaatgt 660

 tatcaagtc tgtccctgtt gtagcatgta ttagcacttc atttcttata tgctgaatga 720
 tatactttat ttgtccatca gttgttcatt ctttatttgg ccatcagttg atgaacattt 780
 gcgtttttgc cactttgggc tattaagaat aatgctactg tgaacaagtg tgtacaagtt 840
 cctctacaaa tttttgtgtg gacatatcct ttcagttctc tcaggtgtat atctgggaat 900
 tgaattgctg ggtcgtgtag tagctatgtt aaacactttg agaaactgct ataatgttct 960
 ccagagctgt accattttaa attctgtgta tgaggattcc acgttctcca ctctctcacc 1020
 agtgtatgga tttgggggta tactttttta aaagtgggat taggctgggc acagtggctc 1080
 acacctgtaa tcccaacct tcaggaagct gaggtgggag gatcacttga gcctagtagt 1140
 ttgagaccag cctgggcaac atagggagac cctgtctcta caaaaaataa tttaaaataa 1200
 attagctggg cgttgtggca cacacctgta gtcccagcta catgggaggc tgaggtggaa 1260
 ggattccctg agcccagaag tttgaggttg cagttagcca tgatggcagc actatactgt 1320
 agctgggtg tcagagcaag actccgttcc agggaagaaa aaaaaaagtg ggatgatatt 1380
 tttgacactt ttcttcttgt tttcttaatt tcatacttct ggaaattcca ttaaattagc 1440
 tggtaacct ctaactcatt gtgtttcatg gctgcatagt aatattgcat aatataaata 1500
 taccattcat tcatcaaagt tagcagatat tgactgttag gtgccaggca ctgctctaag 1560
 cgtaaagaa aaacacacaa aaacttttgc attcttagag tttattttcc aatggagggg 1620
 gtggagggag gtaagaattt aggaaataaa ttaattacat atatagcata gggtttcacc 1680
 agtgagtgca gcttgaatcg ttggcagctt tcttagtagt ataaatacag tactaaagat 1740
 gaaattactc taaatggtgt tacttaaaatt actggaatag gtattactat tagtcacttt 1800
 gcaggtgaaa gtggaaacac catcgtaaaa tgtaaaatag gaaacagctg gttaatgtt 1859

<210> 47

<211> 1082

ES 2 384 391 T3

<212> ADN

<213> Homo sapiens

<220>

<221> característica diversa

5 <223> secuencia de STAR47

<400> 47

```

atcattagtc attagggaaa tgcaaatgaa aacacaagc agccaccaat atacacctac      60
taggatgatt taaaggaaaa taagtgtgaa gaaggacgta aagaaattgt aaccctgata    120
cattgatggt agaaatggat aaagttgcag ccaactgtgaa aacagtctg cagtggctca    180
gaaggttaaa tatagaacct ctgttggacc caggaactct actcttaggc accccaaaga    240
atagagaaca gaaatcaaac agatgtttgt atactaatgt ttgtagcatc acttttcaca    300
ggagccaaaa ggtggaaata atccaacat cagtgaacaa atgaatgtaa taaaagcaag    360
gtggtctgca tgcaatgcta catcatccat ctgtaaaaaa cgaacatcat tttgatagat    420
gatacaacat ggggtggacat tgagaacatt atgcttagtg aaataagcca gacacaaaag    480

gaatatattg tataattgta attaccatgaa gtgcctagaa tagtcaaatt catacaagag    540
aaagtgggat aggaatcacc atgggctgga aataggggga aggtgctata ctgcttattg    600
tggacaaggt ttcgtaagaa atcatcaaaa ttgtgggtgt agatagtggg gttggttatg    660
caaccctgtg aatatattga atgcatgga gtgcacactt tggttaaaag gttcaaatga    720
taaattattg gttatatata tttccccacg atagaaaaca cgcacagcca agcccacatg    780
ccagtcttgt tagctgcctt cctttacctt caagagtggg ctgaagcttg tccaatcttt    840
caaggttgc tgaagactgta tgatggaagt catctgcatt gggaaagaaa ttaatggaga    900
gaggagaaaa cttgagaatc cacactactc accctgcagg gccaagaact ctgtctccca    960
tgctttgctg tctctgtctca gtatttcctg tgaccacctc ctttttcaac tgaagacttt  1020
gtacctgaag gggttcccag gtttttcacc tcggcccttg tcaggactga tcctctcaac  1080
ta                                     . . . . .                          1082

```

<210> 48

10 <211> 1242

<212> ADN

<213> Homo sapiens

<220>

<221> característica diversa

15 <223> secuencia de STAR48

<400> 48

ES 2 384 391 T3

atcatgtatt tgttttctga attaattcctt agatacatta atgttttatg ttaccatgaa 60
 tgtgatatta taatataata tttttaattg gttgctactg tttataagaa tttcattttc 120
 tgtttacttt gccttcatat ctgaaaacct tgctgatttg attagtgcac ccacaaattt 180
 tcttgatttt tctatgggta attacaaatc tccacacaat gaggttgacag tgagccaaga 240
 tcacaccact gtactccagc ctgggagaca gagtgagaca ccatctcaca aaaacacata 300
 aacaaacaaa cagaaactcc acacaatgac aacgtatgtg ctttcttttt ttcttctctc 360
 ttctataata tttctttgtc ctatcttaac tgaactggcc agaaacccca ggacaatgat 420
 aaatacgagc agtgtcaaca gacatctcat tccctttcct agcttttata aaaataacga 480
 ttatgcttca acattacata tgggtggtgc gatggttttg ttatagataa gcttatcagg 540
 ttaagaaatt tgtctgcggt tccatgattg gtataaagat tttaatataa atgaatggtg 600
 tattttatca tcttattttt ttectacatc tgctaaggta atcctgtggt tteccctttt 660
 caatctccta atgtgggtgaa tgacattaaa ataccttcta ttgttaaaat attcttgcaa 720
 cgctgtatag aaccaatgcc tttattctgt attgctgatg gatttttgaa aaatatgtag 780
 gtggacttag ttttctaagg ggaatagaat ttctaataa tttaaaatat tttgcatgta 840
 tgttctgaag gacattggtg tgtcatttct ataccatctg gctactagag gagccgactg 900
 aaagtacacac tgccggagga ggggagaggt gctcttccgt ttctgggtgc tgtagccatc 960
 tccagtggta gctgcagtga taataatgct gcagtgccga cagttctgga aggagcaaca 1020
 acagtgattt cagcagcagc agtattgcgg gatccccacg atggagcaag ggaaataatt 1080

 ctggaagcaa tgacaatatc agctgtggct atagcagctg agatgtgagt tctcacggtg 1140
 gcagcttcaa ggacagtagt gatggtccaa tggcgcccag acctagaaat gcacatttcc 1200
 tcagcaccgg ctccagatgc tgagcttggg cagctgacgc ct 1242

<210> 49

<211> 1015

5 <212> ADN

<213> Homo sapiens

<220>

<221> característica diversa

<223> secuencia de STAR49

10 <400> 49

ES 2 384 391 T3

aaaccagaaa cccaaaacaa tgggagtgac atgctaaaac cagaaacca aaacaatggg 60
 agggctctgc taaaccagaa acccaaaaca atgggagtgga agtgctaaaa ccagaaaccc 120
 aaaacaatgg gagtgtcctg ctacaccaga aaccctaaaac gatgggagtg acgtgataaa 180
 accagacacc caaaacaatg ggagtgacgt gctaaaccag aaaccctaaa caatgggagt 240
 gacgtgctaa aacctggaaa cctaaaacaa tgcgagtgag gtgctaacac cagaatccat 300
 aacaatgtga gtgacgtgct aaaccagaac ccaaaacaat gggagtgacg tgctaaaaca 360
 ggaaccctaaa acaatgagag tgacgtgcta aaccagaaac ccaaaacaat gggaatgacg 420
 tgctaaaacc ggaaccctaaa acaatgggag tgatgtgcta aaccagaaac ccaaaacaat 480
 gggaatgaca tgctaaaact ggaaccctaaa acaatggtaa ctaagagtga tgctaaggcc 540
 ctacattttg gtcacactct caactaagtg agaacttgac tgaaaaggag gatttttttt 600
 tctaagacag agttttggtc tgtccccag agtggagtgac agtggcatga tctcggctca 660
 ctgcaagctc tgcctcccgg gttcaggcca ttctctgcc tcagcctcct gagtagctgg 720
 gaatacaggc acccgccacc aacttggtt aattttttgt atttttagta gagatggggg 780
 ttcaccatat tagcaaggat ggtctcaatc tctgacctc gtgatctgcc cacctcaggc 840
 tcccaaagtg ctgggattac aggtgtgagc caccacacc agcaaaaagg aggaattttt 900
 aaagcaaat tatgggaggc cattgttttg aactaagctc atgcaatagg tcccaacaga 960
 ccaaaccctaaa ccaaaccctaaa atggagtcac tcatgctaaa tgtagcataa tcaaa 1015

<210> 50

<211> 2355

<212> ADN

5 <213> Homo sapiens

<220>

<221> característica diversa

<223> secuencia de STAR50

<400> 50

caaccatcgt tccgcaagag cggttggtt attaaacatg aatgagggga aaagcctagt 60
 agctccattg gattgggaag aatggcaag agagacaggc gtcattttct agaaagcaat 120
 cttcacacct gttggtcctc acccattgaa tgcctcacc caatctcaa cacagaaatg 180

10

ES 2 384 391 T3

agtgactgtg tgtgcacatg cgtgtgcatg tgtgaaagta tgagtgtgaa tgtgtctata 240
 tgggaacata tatgtgattg tatgtgtgta actatgtgtg actggcagcg tggggagtgc 300
 tggttggagt gtggtgtgat gtgagtatgc atgagtggct gtgtgtatga ctgtggcggg 360
 aggcggaagg ggagaagcag caggctcagg tgtcgcagga gaggtggga ggaaactata 420
 aacctgggca atttcctcct catcagcgag cctttcttgg gcaatagggg cagagctcaa 480
 agttcacaga gatagtgcct gggaggcatg aggcaaggcg gaagtactgc gaggaggggc 540
 agagggctcg acacttgagg ggttctaata ggaaaggaaa gaccacact gaattccact 600
 tagccccaga ccctgggcc agcggtgccg gcttccaacc ataccaacca tttccaagtg 660
 ttgccggcag aagttaacct ctcttagcct cagtttcccc acctgtaaaa tggcagaagt 720
 aaccaagctt accttcccgg cagtgtgtga ggatgaaaag agctatgtac gtgatgcact 780
 tagaagaagg tctaggggtg gagtggact cgtctgggtg gtgtggagaa gacattctag 840
 gcaatgagga ctggggagag cctggcccat ggcttccact cagcaaggtc agtctcttgt 900
 cctctgcact ccagccttc cagagaggac cttccaacc agcactcccc acgctgccag 960
 tcacacatag ttacacacat acaatcacat atatgtccc atatagacac attcacactc 1020
 atacctcac acatgcacac gcatgtgac acacagtcac tcatttctgt gttggagatt 1080
 gggtgaggac attcaatggg tgaggaccaa cagggtgtgaa gattgctttc tagaaaatga 1140
 ctctgtctc tctttgcat tcttccaat ccgatggagc tactaggett ttccctcatt 1200
 tcatgtttaa taaaccttc caatggcgaa atgggctttc tcaagaagtg gtgagtgtcc 1260
 catectgog gtggggacag ggggtggcagc ggacaagcct gcctggaggg aactgtcagg 1320
 ctgattcca gtccaactcc agcttccaac acctcactcc ccaggcagtc ttcattcttg 1380
 gctetaattt cgctcttgtt ttctttttta tttttatcga gaactgggtg gagagctttt 1440
 ggtgtcattg gggattgctt tgaaacctt ctctgcctca cactgggagc tggcttgagt 1500
 caactggtct ccatggaatt tctttttta gtgtgtaaac agctaagttt taggcagctg 1560
 ttgtgccgtc cagggtgaa agcagcctgt tgatgtgaa ctgcttggt cagatttctt 1620
 gggcaaacag atgccgtgtc tctcaactca ccaattaaga agcccagaaa atgtggcttg 1680
 gagaccacat gtctggttat gtctagtaat tcagatggct tcacctggga agccctttct 1740
 gaatgtcaaa gccatgagat aaaggacata tatatagtag ctagggtggt ccacttctta 1800
 ggggccatct ccggagggtg tgagcactaa gtgccaggaa gagaggaaac tctgttttgg 1860
 agccaaagca taaaaaac ttagccacaa accactgaac atttgttttg tgcaggttct 1920
 gagtccaggg agggcttctg aggagagggg cagctggagc tggtaggagt tatgtgagat 1980
 ggagcaaggg cccttaaga ggtgggagca gcatgagcaa aggcagagag gtggtaatgt 2040
 ataaggtatg tcatgggaaa gagtttggt ggaacagagt ttacagaata gaaaaattca 2100
 acactattaa ttgacctct actacgtgct cgacattgtt ctagtactg agataggttt 2160
 ggtatacaaa acaaaatcca tcctctatgg acattttagt gactaacaac aatataaata 2220

ES 2 384 391 T3

ataaaagtga acaaaagctc aaaacatgcc aggcactatt atttatttat ttatttattt 2280
atattattat tttttgaaac agagtctcgc tctgttgccc aggctggagt gtagtggtgc 2340
gatctcggct cactg 2355

<210>51

<211> 2289

<212> ADN

5 <213> Homo sapiens

<220>

<221> característica diversa

<223> secuencia de STAR51

<400> 51

ES 2 384 391 T3

tcacaggtga caccaatccc ctgaccacgc ttgagaagc actgtactag attgactttc 60
 taatgtcagt cttcattttc tagctctggt acagccatgg tctccatatt atctagtaca 120
 acacacatac aaatatgtgt gatacagtat gaatataata taaaaatatg tgttataata 180
 taaatataat attaaaaatg gtctttatac tagataataa tacttaataa cgttgagtgt 240
 ttaactgctc taagcacttt acctgcagga aacagttttt tttttatttt ggtgaaatac 300
 aactaacata aatttattta caattttaag catttttaag tgtatagttt agtggagtta 360
 atatatcaa aatgttgtgc agccgtcacc atcatcagtc ttcataactc ttttcatatt 420
 gtaaaattaa aagtttatgc tcatttaaaa atgactceca attcccccc tctcaacct 480
 ctgaaaacta ccattctatt ttctgcctcc gtagttttgc ccactetaag tacctcacat 540
 aagtggaatt tgtcttattt gcctgtttgt gaccggctga tttcatttag tataatgtcc 600
 tcaagtttta ttcacgttat atagcatatg tcataatttt cttcactttt aagcttgagt 660
 aatatttcat cgtatgtatc tcacattttg cttatccatt catctctcag tggacacttg 720
 agttgcttct acattttagc tgttgtgaat actgctgcta tgaacatggg tgtataaata 780
 tctcaagacc tttttatcag ttttttaaaa tatatactca gtagtagttt agctggatta 840
 tatggtaatt ttatttttaa tttttgagga actgtcctac cttttattc aatagtagct 900
 ataccaattg acaattggca ttcctaccaa cagggcataa gggttctcaa ttctccacat 960
 attccctgat acttgttatt ttcaggtgtt tttttttttt tttttttttt atgggagcca 1020
 tgttaatggg tgtaagggtga tatttcatta tagttttgat ttgcatttcc ctaatgatta 1080
 gtgatgttaa gcatctcttc atgtgcctat tggccatttg tatatcttct ttaaaaatat 1140
 atatatactc attcctttgc ccatttttga attatgttta ttttttgta ttgagtttca 1200
 atacttttct atataaccta ggtattaatc ctttatcaga ctttaagattt gcaaatattc 1260
 tctttcattc cacaggttgc taattctctc tgttggtaat atcttttgat gctgttgtgt 1320
 ccagaattga ttcattctctg tgggttcttg gtctcactga cttcaagaat aaagctgcgg 1380
 acctagtgg tgagtgttac acttcttata gatggtgttt ccggagttag ttccttcaga 1440
 tgtgtccaga gtttcttctt tccaatgggt tcatggtctt gctgacttca ggaatgaagc 1500

ES 2 384 391 T3

cgcagacctt cgcagtgagg ttacagctc ttaaagggtg cgtgtccaga gttgtttggt 1560
 ccccttggtg ggttcgtggt cttgctgact tcaggaatga agccgcagac cctcgcagtg 1620
 agtgttacag ctcataaagg tagtgcgagc acagagtgag ctgcagcaag atttactgtg 1680
 aagagcaaaa gaacaaagct tccacagcat agaaggacac cccagcgggt tcctgctgct 1740
 ggctcaggtg gccagttatt attcccttat ttgcctgcc cacatcctgc tgattggtcc 1800
 attttacaga gtactgattg gtccatTTta cagagtgtg attggtgcat ttacaatcct 1860
 ttagctagac acagagtgtc gattgctgca ttcttacaga gtgctgattg gtgcatttac 1920
 agtcctttag ctagatacag aacgctgatt gctgcgtttt ttacagagtg ctgattggtg 1980
 catttacaat cctttagcta gacacagtgc tgattggtgg gtttttacag agtgctgatt 2040
 ggtgctcctt tacagagtgc tgattggtgc atttacaatc ctttagctag acacagagtg 2100
 ctgattggtg cgtttataat cctctagcta gacagaaaag tttccaagt cccacctga 2160
 ccgagaagcc ccaactggctt cacctctcac tgttatactt tggacatttg tcccccaaa 2220
 atctcatggt gaaatgtaac ccctaagtgtt ggaactgagg ccagactgga tgtggtggg 2280
 ccatgggga 2289

<210> 52

<211> 1184

<212> ADN

5 <213> Homo sapiens

<220>

<221> característica diversa

<223> secuencia de STAR52

<400> 52

ES 2 384 391 T3

ctcttctttg tttttttatt ttggggtgtg tgggtacgtg taagatgaga aatgtacaaa 60
 cacaagtatt tcagaaactc caagtaatat tctgtctgtg agttcacggt aaataaataa 120
 aaagggcaaa gtgacagaaa tacaggatta ttaaaagcaa aataatgttc tttgaaatcc 180
 cccccttggg gtatttttta tcttaggatg cagcactttc agcatgccca agtattgaaa 240
 gcagtgtttt tacgctacca cggtaatttt atttagaaac cccatgttca cttttagttt 300
 taaaatggtc tttatgacat aaaattatca gcattcatat ttttgtgttt taatattcct 360
 ttggctactt attgaaacag taaacattac gaaaattagt aaacaaatct ttgatagttg 420
 cttatttttg ttttaattgaa tgtttatttt attaggtaaa tatacaatca aatttattta 480
 aaaataatga ggaaaagaat acttttcttt cgctttgcga aagcaaagtg atttttcatt 540
 cttctccgtc cgattccttc tcttccagct gccacagccg actgacaggc tcccggcggc 600
 ctgaggagta gtatgcaaat tttggatgat tgacacctac agtagaagcc aatcacgtca 660
 aagtaggatg ctgattgggt gacaacaata ggcgtaaacc ttgacgtttt aaaaacctga 720
 cacccaatcc aggcgattca tgcaataaaa ggaagggagt cacattacca ggggccagag 780
 agacttgagt acgacctcac gtgttcagtg gtggatattg cacagacgtc tgcaaggtct 840
 atataaacgc tacataatgt tcaactcaat tgcttgctt ggcctttccc aaacttgtca 900

ctggaatata aattatccct tttttaaaaa taaaaaata agaattatgt agtgcacata 960
 tatgatggtt catgtagaaa tctaaatgga cttccaacgc atggaatttt cctatttccc 1020
 cttttcttta aattaatcct cagtgaagga ggctgttttc ccctagattt caaaaggagc 1080
 agatttacag agcctttcct tggagaaacc cgctctagge acagatggtc agtaaattta 1140
 gcttcttcag cgaagttcca catggcaccg ccagatggca taag 1184

<210> 53

<211> 1431

5 <212> ADN

<213> Homo sapiens

<220>

<221> característica diversa

<223> secuencia de STAR53

10 <400> 53

ES 2 384 391 T3

```

ccctgaggaa gatgacgagt aactcogtaa gagaaccttc cactcatccc ccacatccct      60
gcagacgtgc tattctgtta tgatactggg atcccatctg tcacttgctc cccaaatcat      120
tccttcttta caattttcta ctgtacagca ttgaggctga acgatgagag atttcccatg      180
ctctttctac tccttgccct gtatatatcc ggggatcctc cctaccagg atgctgtggg      240
gtcccaaacc ccaagtaagc cctgatatgc gggccacacc tttctctagc ctaggaattg      300
ataaccagg cgaggaagtc actgtggcat gaacagatgg ttcacttcca ggaaccgtgg      360
aaggcgtgtg caggtcctga gatagggcag aatcggagtg tgcagggtct gcaggtcagg      420
aggagttagg attgcgttgc cacgtgggtg gaactcactg ccacttattt ccttctctct      480
tcttgccca gcctcagga tacgacacat gcccatgatg agaagcagaa cgtggtgacc      540
ttcacgaac atgggcatgg ctgcggacc ctcgtcatca ggtgcatagc aagtgaaagc      600
aagtgtcac aacagtgaag agttgagcgt ctttttctt agtgtgcaa gagttcagtg      660
ttagcgttta cgttgtattt tcttacctg tgcattctg ttagatacta acattttcat      720
tgatgagcaa gacatactta atgcatattt tggtttgtgt atccatgcac ctacctaga      780
aaacaagtat tgcggttac ctctgcatgg aacagcatta ccctccttc tcccagatg      840
tgactactga gggcagttct gagtgtttaa tttcagattt tttcctctgc atttacacac      900
acacgcacac aaaccacacc acacacacac acacacacac acacacacac acacacacac      960
acacaccaag taccagtata agcatctgcc atctgctttt cccattgcca tgcgtcctgg     1020
tcaagctccc ctactctgt ttctgggtca gcatgtaact ccctcatccg attcccctgt     1080
agcagtcact gacagttaat aaacctttgc aaacgttccc cagttgtttg ctctgcatc     1140
tattgtgcac acagctctgt gcacgtgtgt gcatatttct ttaggaaaga ttcttagaag     1200
tggaattgct gtgtcaaagg agtcatttat tcaacaaaac actaatgagt gcgtcctcgt     1260
gctgagcgtc gttctaggtg ctggagcgac gtcagggaac aaggcagaca ggagttcctg     1320
acccccgttc tagaggagga tgtttccagt tgttgggttt tgtttgtttg tttcttctag     1380

agatggtggt cttgctctgt ccaggctaga gtgcagtggc atgatcatag c              1431

```

<210> 54

<211> 975

<212> ADN

5 <213> Homo sapiens

<220>

<221> característica diversa

<223> secuencia de STAR54

<400> 54

ES 2 384 391 T3

ccataaaagt gtttctaaac tgcagaaaaa tccccctaca gtcttacagt tcaagaattt 60
 tcagcatgaa atgcctggta gattacctga ctttttttgc caaaaataag gcacagcagc 120
 tctctcctga ctctgacttt ctatagtctt tactgaatta tagtccttac tgaattcatt 180
 cttcagtgtt gcagtctgaa ggacaccac attttctctt tgtctttgtc aattctttgt 240
 gttgtaaggg caggatgttt aaaagttgaa gtcattgact tgcaaaatga gaaatttcag 300
 agggcatttt gttctctaga ccatgtagct tagagcagtg ttcacactga ggttgctgct 360
 aatgtttctg cagttcttac caatagtatc atttaccag caacaggata tgatagagga 420
 cttcgaaaac cccagaaaat gttttgcat atatccaaag ccctttggga aatggaaagg 480
 aattgcgggc tcccattttt atatatggat agatagagac caagaaagac caaggcaact 540
 ccatgtgctt tacattaata aagtacaaaa tgtaacatg taggaagtct aggcgaagtt 600
 tatgtgagaa ttctttacac taattttgca acattttaat gcaagtctga aattatgtca 660
 aaataagtaa aaatttttac aagttaagca gagaataaca atgattagtc agagaaataa 720
 gtagcaaaat cttcttctca gtattgactt ggttgctttt caatctctga ggacacagca 780
 gtcttcgctt ccaaatccac aagtcacatc agtgaggaga ctgagctgag actttggcta 840
 atgttggggg gtccctcctg tgtctcccca ggcgagtgga gcctgcaggc cgacctcact 900
 cgtggcacac aactaaatct ggggagaagc aaccgatgc cagcatgatg cagatatctc 960
 aggtatgat cggcc 975

<210> 55

<211> 501

<212> ADN

5 <213> Homo sapiens

<220>

<221> característica diversa

<223> secuencia de STAR55

<400> 55

cctgaactca tgatccgccc acctcagcct cctgaagtgc tgggattaca ggtgtgagcc 60
 accacaccca gccgcaacac actcttgagc aaccaatgtg tcataaaaga aataaaatgg 120
 aatcagaaa gtatcttgag acagacaaaa atggaaacac aacataccaa aatttatggg 180
 acacagcaaa agcagtttta ggagggaggt ttatagtgat gaatacctac ctcaaaatca 240
 ttagcctgat tggatgacac tacagtgat aatgaattg aaaaccacat tgtgccccat 300

10

ES 2 384 391 T3

acatatatac aatTTTTatt tGTTaattaa aaataaaata aaactTTaaa aaagaagaaa 360
 gagctcaaat aaacaaccta actTTataacc tcaaggaat agaagagcca gctaagccca 420
 aagttgacag aaggaaaaaa atattggcag aaagaatga aacagagact agaaagacaa 480
 ttgaagagat cagcaaaaact a 501

<210> 56

<211> 741

<212> ADN

5 <213> Homo sapiens

<220>

<221> característica diversa

<223> secuencia de STAR56

<400> 56

acacagaaa agatcgcaat tGttcagcag agctttgaac cggggatgac ggtctccctc 60
 gttgcccggc aacatggtgt agcagccagc cagttatttc tctggcgtaa gcaataccag 120
 gaaggaagtc ttactgctgt cgcgccgga gaacaggttG ttctgcctc tgaacttGct 180
 gccgccatga agcagattaa agaactccag cgcctgctcg gcaagaaaac gatggaaaat 240
 gaactcctca aagaagccgt tgaatatgga cgggcaaaaa agtggatagc gcacgcgccc 300
 ttattgcccG gggatgggga gtaagcttag tcagccgttg tctccgggtg tcgcgtgcgc 360
 agttgcacgt cattctcaga cgaaccgatg actggatgga tggccgccgc agtcgtcaca 420
 ctgatgatac ggatgtgctt ctccgtatac accatgttat cggagagctg ccaacgtatg 480
 gttatcgtcg ggtatggcg ctgcttcgca gacaggcaga acttgatggT atgcctgcga 540
 tcaatgcaa acgtgtttac cggatcatgc gccagaatgc gctgttgctt gagcgaaaac 600
 ctgctgtacc gccatcgaaa cgggcacata caggcagagt ggccgtgaaa gaaagcaatc 660
 agcgatgggtg ctctgacggg ttcgagttct gctgtgataa cggagagaga ctgcgtgtca 720
 cgttcgcgct ggactgctgt g 741

10

<210> 57

<211> 1365

<212> ADN

<213> Homo sapiens

15

<220>

<221> característica diversa

<223> secuencia de STAR57

<400> 57

ES 2 384 391 T3

tccttctgta aataggcaaa atgtatttta gtttccacca cacatgttct tttctgtagg 60
gcttgtatgt tggaaathtt atccaattat tcaattaaca ctataccaac aatctgctaa 120
ttctggagat gtggcagtga ataaaaagt tatagtttct gattttgtgg agcttgact 180
ttaatgatgg acaaaacaac acattcttaa atatatattt catcaaaatt atagtgggtg 240
aattatttat atgtgcattt acatgtgtat gtatacataa atgggcggtt actggctgca 300
ctgagaatgt acacgtggcg cgaacgaggc tgggcggtca gagaaggcct cccaaggagg 360

tggtttgaa gctgagtggg gcttccacgt gaaaaggctg gaaagggcat tccaagaaaa 420
ggctgaggcc agcgggaaag aggttccagt gcgctctggg aacggaaagc gcacctgcct 480
gaaacgaaaa tgagtgtgct gaaataggac gctagaaagg gaggcagagg ctggcaaaaag 540
cgaccgagga ggagctcaaa ggagcgagcg ggggaaggccg ctgtggagcc tggaggaagc 600
acttcggaag cgcttctgag cgggtaaggc cgctgggagc atgaactgct gaggcaggtg 660
gtccagaatt cgtgggttct tggctcact gacttcaaga atgaagaggg accgcggacc 720
ctcgcggtga gtgttacagc tcttaagggtg gcgctctggg agtttggtcc ttctgatgtt 780
cggatgtgtt cagagtttct tccttctggt gggttcgtgg tctcgtctggc tcaggagtga 840
agctgcagac cttcgcggtg agtggtacag ctcataaaag cagggaggac tcaaagagtg 900
agcagcagca agatttattg caaagaatga aagaacaaag cttccacact gtggaagggg 960
aceccagcgg gttgccactg ctggctccgc agcctgcttt tattctotta tctggcccca 1020
cccacatcct gctgattggg agagccgaat ggtctgtttt gacggcgctg attggtgcgt 1080
ttacaatccc tgcgctagat acaaaggttc tccacgtccc caccagatta gctagataga 1140
gtctccacac aaaggttctc caaggcccca ccagagtagc tagatacaga gtggtgattg 1200
gtgcattcac aaaccctgag ctagacacag ggtgatgact ggtgtgttta caaaccttgc 1260
ggtagataca gagtatcaat tggcgtattt acaatcactg agctaggcat aaaggttctc 1320
caggtcccca ccagactcag gagcccagct ggcttcaccc agtgg 1365

<210> 58

<211> 1401

5 <212> ADN

<213> Homo sapiens

<220>

<221> característica diversa

<223> secuencia de STAR58

10 <400> 58

ES 2 384 391 T3

aagtttacct tagccctaaa ttatttcatt gtgattggca ttttaggaaa tatgtattaa 60
 ggaatgtctc ttaggagata aggataacat atgtctaaga aaattatatt gaaatattat 120
 tacatgaact aaaatgtag aactgaaaaa aaattattgt aactccttcc agcgtaggca 180
 ggagtatcta gataccaact ttaacaactc aactttaaca acttcgaacc aaccagatgg 240
 ctaggagatt cacctattta gcatgatatc ttttattgat aaaaaaatat aaaacttcca 300
 ttaaattttt aagctactac aatcctatta aattttaact taccagtgtt ctcaatgcta 360
 cataatntaa aatcattgaa atottctgat ttttaactct cagtcttgaa atctacttat 420
 ttttagttac atatatatcc aatctactgc cgctagtaga agaagcttgg aatttgagaa 480
 aaaaatcaga cgttttgtat attctcatat tcaactaattt attttttaa tgagtttctg 540
 caatgcatca agcagtggca aaacaggaga aaaatntaaa ttggttgaaa agatatgtgt 600
 gccaaaacat cccttgaaat ttgatgaagt gactaatctt gagttattgt ttcaaatgtg 660
 tacctgttta tacaagggtt tcaaccttga aatctcaaca ttaaatgaaa ttttataagc 720

 aatttgttgt aacatgatta ttataaaatt ctgatataac attttttatt acctgtttag 780
 agtttaaaga gagaaaagga gttaagaata attacatttt cattagcatt gtccgggtgc 840
 aaaaacttct aacactatct tcaaatcttt ttctccattg ccttctgaac ataccactt 900
 gggatctca ttagcactgc aaattcaaca ttttcgattg ctaatttttc tccctaaata 960
 tttatttggt ttctcagctt tagccaatgt ttcaactattg accatttgct caagtatagt 1020
 gacgcttcaa tgacctcag agagctggtt cagtccttcc tggactactt gcatgcttcc 1080
 aacaaaatga agcactcttg atgtcagtca ctcaaataaa tggaaatggg cccatttact 1140
 aggaatgta acagaataaa aagatagacg tgacaccagt tgcttcagtc catctccatt 1200
 tacttgctta aggcctggcc atatttctca cagttgatat ggcgcagggc acatgtntaa 1260
 atggctgttc ttgtaggatg gtttgactgt tggattcctc atcttcctc tcttaggaa 1320
 ggaaggttac agtagtactg ttggctcctg gaatatagat tcataaagaa ctaatggagt 1380
 atcatctccc actgctcttg t 1401

<210> 59

<211> 866

5 <212> ADN

<213> Homo sapiens

<220>

<221> característica diversa

<223> secuencia de STAR59

10 <400> 59

ES 2 384 391 T3

gagatcacgc cactgcaetc cagcctgggg gacagagcaa gactccatct cagaaacaaa 60
 caaacacaca aagccagtca aggtgtttaa ttcgacgggtg tcaggctcag gtctcttgac 120
 aggatacadc cagcaccocgg gggaaacgtc gatgggtggg gtggaatcta tttgtggcc 180
 tcaagggagg gtttgagagg tagtcccgca agcgggtgat gcctaaggaa gcccctccgc 240
 ccaagaagcg atattcattt ctagcctgta gccaccaag agggagaatc gggctcgcca 300
 cagacccac aacccccaac ccaccccacc cccaccctc ccacctcgtg aaatgggctc 360
 tcgctccgtc aggctctagt cacaccgtgt ggttttggaa cctccagcgt gtgtgcgtgg 420
 gttgcgtggt ggggtggggc cggctgtgga cagaggaggg gataaagcgg cgggtgcccg 480
 cgggtgcccc ggacgtgggg cgtggggcgt ggggtggggtg gccagagcct tgggaactcg 540
 tcgcctgtcg ggacgtctcc cctcctggtc cectctctga cctacgctcc acatcttcgc 600
 cgttcagtgg ggacctgtg ggtggaagtc accatccctt tggactttag ccgacgaagg 660
 ccgggctccc aagagtctcc ccggaggcgg ggccttgggc aggctcacia ggatgctgac 720
 ggtgacgggt ggtgacgggt atgtacttcg gaggcctcgg gccaatgcag aggtatccat 780
 ttgacctcgg tgggacaggt cagctttgcg gagtcccgtg cgtccttcca gagactcatc 840
 cagcgctagc aagcatggtc ccgagg 866

<210> 60

<211> 2067

<212> ADN

5 <213> Homo sapiens

<220>

<221> característica diversa

<223> secuencia de STAR60

<220>

10 <221> característica diversa

<222> (92)..(1777)

<223> n es a, c, g, o t en diversas posiciones

<400> 60

ES 2 384 391 T3

agcagtgacg aactggggaa gaagaagagt ccctacacca cctaatactc aaaagtactc 60
 gcaaaaaata acaccctca ccaggtggca tnattactct ccttcattga gaaaattagg 120
 aaactggact tcgtagaagc taattgcttt atccagagcc acctgcatac aaacctgcag 180
 cgccacctgc atacaaacct gtcagccgac cccaaagccc tcagtcgcac caagcctctg 240
 ctgcacaccc tcgtgccttc aactggccg tcccccaagc ctggggcata ctncccagct 300
 ctgagaaatg tattcatcct tcaaagccct gctcatgtgt cctnntcaac aggaaaatct 360
 cccatgagat gctctgctat ccccatctct cctgccccat agcttaggca nacttctgtg 420
 gtggtgagtc ctgggctgtg ctgtgatgtg ttcgctgcn atgntgttc tccccacaa 480
 tgatgggccc ctgaattctc tatctctagc acctgtgctc agtaaaggct tgggaaacca 540
 ggctcaaagc ctggcccaga tgccacctt tccaggtgct tccgggggc caccaaccag 600
 agtgcagcct tctcctccac caggaactct tgcagcccca ccctgagca cctgcacccc 660
 attacccatc tttgtttctc cgtgtgatcg tattattaca gaattatata ctgtattctt 720
 aatacagtat ataattgtat aattattctt aatacagtat ataattatac aaatacaaaa 780
 tatgtgttaa tggaccgttt atgttactgg taaagcttta agtcaacagt gggacattag 840
 ttaggttttt gggaagtca aaagttatat gtgcatttct aacttcttga ggggtcggta 900
 cntctnacc ccatgttgtt caanggtcaa ctgtctacac atatcatagc taattcacta 960
 cagaaatggt agcttgtgct actagtatct ccccttctca taagcttaat acacatacct 1020
 tgagagagct cttggccatc tetactaatg actgaagttt ttatttatta tagatgtcat 1080
 aataggcata aaactacatt acatcattcg agtgccaatt ttgccacctt gaccctcttt 1140
 tgcaaaacac caacgtcagt acacatatga agaggaaact gcccgagaac tgaagttcct 1200
 gagaccagga gctgcaggcg ttagatagaa tatggtgacg agagtacga ggatgacgag 1260
 agtaaatact tcatactcag tacgtgcaa gcaactgctat aagcgtctctg tatgtgtgaa 1320
 gtcatttaat cctcacagca tcccacggtg taattatfff cattatcccc atgagggaac 1380
 agaaactcag aacggtcaa cacatatgcg agaagtcgca gccggtcagt gagagagcag 1440
 gttccogtcc aagcagtcag accccgagtg cacactctcg acccctgtcc agcagactca 1500
 ctcgtcataa ggcggggagt gntctgtttc agccagatgc tttatgcac tcagagtacc 1560
 caaacatga aagaatgagg cagtattcan gagcagatgg ngctgggcag taaggctggg 1620

ES 2 384 391 T3

cttcagaata gctggaaagc tcaagtnatg ggacctgcaa gaaaaatcca ttgtttngat 1680
 aatatagccaa agtccctagg ctgtaagggg aaggtgtgcc aggtgcaagt ggagctctaa 1740
 tgtaaaatcg cacctgagtc tcctggctct atgagtnctg ggtgtacccc agtgaaaggt 1800
 cctgctgcca ccaagtgggc catggttcag ctgtgtaagt gctgagcggc agccggaccg 1860
 cttcctctaa cttcacctcc aaaggcacag tgcacctggt tcctccagca ctcagctgcg 1920
 agggccctag ccaggggtccc ggcccccgcc ccccggcagc tgcctccagct tccttcccca 1980
 cagcattcag gatgggtctgc gttcatgtag acctttgttt tcagtctgtg ctccgaggtc 2040
 actggcagca ctagccccgg ctctctgt 2067

<210>61

<211> 1470

<212> ADN

5 <213> Homo sapiens

<220>

<221> característica diversa

<223> secuencia de STAR61

<220>

10 <221> característica diversa

<222> (130)..(976)

<223> n es a, c, g, o t en diversas posiciones

<400> 61

ES 2 384 391 T3

cagccccac atgccagcc ctgtgctcag ctctgcagcg gggcatggtg ggcagagaca 60
 cagaggccaa ggccctgctt cggggacggt gggcctggga tgagcatggc cttggcette 120
 gccgagagtn ctcttgtaga ggaggggtca ggaggggctg ctgcagctgg ggaggagggc 180
 gatggcactg tggcangaag tgaantagtg tgggtgcctn gcaccccagg cacggccagc 240
 ctggggtatg gacccggggc cntctgttct agagcaggaa ggtatggtga ggacctcaaa 300
 aggacagcca ctggagagct ccaggcagag gnacttgaga ggccctgggg ccatcctgtc 360
 tcttttctgg gtctgtgtgc tctgggcctg ggcccttctt ctgctcccc gggcttgag 420
 agggctggcc ttgcctcgtg caaaggacca ctctagactg gtaccaagtc tggcccatgg 480
 cctcctgtgg gtgcaggcct gtgcgggtga cctgagagcc agggctggca ggtcagagtc 540
 aggagagggga tggcagtga tgccctgtgc aggatctgcc taatcatggt gaggctggag 600
 gaatccaaag tgggcatgca ctctgcactc atttctttat tcatgtgtgc ccatcccaac 660
 aagcagggag cctggccagg agggcccctg ggagaaggca ctgatgggct gtgttccatt 720
 taggaagat ggacggttgt gagacgggta agtcagaacg ggctgcccac ctcggccgag 780
 agggccccgt ggtgggttgg caccatctgg gcctggagag ctgctcagga ggctctctag 840
 ggctgggtga ccaggnctgg ggtacagtag ccatgggagc aggtgcttac ctggggctgt 900
 ccctgagcag gggctgcatt ggggtgctctg tgagcacaca cttctctatt cacctgagtc 960
 ccnctgagtg atgagnacac ccttgttttg cagatgaatc tgagcatgga gatgttaagt 1020
 ggcttgctg agccacacag cagatggatg gtgtagctgg gacctgaggg caggcagtcc 1080

 cagcccagag acttcccaag gttgtggcaa actctgacag catgaccca gggaacaccc 1140
 atctcagctc tggtcagaca ctgcggagtt gtgttgtaac ccacacagct ggagacagcc 1200
 accctagccc cacccttacc ctctcccaaa ggaacctgcc ctttcccttc attttctct 1260
 tactgcattg agggaccaca cagtgtggca gaaggaacat gggttcagga cccagatgga 1320
 cttgcttcac agtgcagccc tctgtcctc ttgcagagtg cgtcttccac tgtgaagttg 1380
 ggacagtcac accaactcaa tactgctggg cccgtcacac ggtgggcagg caacggatgg 1440
 cagtcactgg ctgtgggtct gcagaggtgg 1470

<210> 62

<211> 1011

5 <212> ADN

<213> Homo sapiens

<220>

<221> característica diversa

<223> secuencia de STAR62

10 <400> 62

ES 2 384 391 T3

```

agtgtcaaat agatctacac aaaacaagat aatgtctgcc catttttcca aagataatgt      60
ggtgaagtgg gtagagagaa atgcatccat tctccccacc caacctctgc taaattgtcc      120
atgtcacagt actgagacca gggggcttat tcccagcggg cagaatgtgc accaagcacc      180
tcttgtctca atttgcagtc taggccctgc tatttgatgg tgtgaaggct tgcacctggc      240
atggaaggtc cgttttgtac ttcttgcttt agcagttcaa agagcagggg gagctgcgag      300
ggcctctgca gcttcagatg gatgtggtca gcttggttga ggcgccttct gtggtccatt      360
atctccagcc cccctgcggt gttgctgttt gcttggttgg tctggctctc catgccttgt      420
tggctccaaa atgtcatcat gctgcacccc aggaagaatg tgcaggccca tctcttttat      480
gtgctttggg ctattttgat tccccgttgg gtatattecc taggtaagac ccagaagaca      540
caggaggtag ttgctttggg agagtttggg cctatgggta tgaggttaata gacacagtat      600
cttctctttc atttggtgag actgtagct ctggccgctg actgaattcc acacagctca      660
cttgggaaaa ctttattcca aaacatagtc acattgaaca ttgtggagaa tgagggacag      720
agaagagccc ctagatttgt acatctgggt gttatgtcta taaatagaat gctttggtgg      780
tcaactagac ttgttcatgt tgacatttag tcttgccttt tcggtggtga tttaaaaatt      840
atgtatatct tgtttggaat atagtggagc tatggtgtgg cattttcatc tggctttttg      900
tttagctcag cccgtcctgt tatgggcagc cttgaagctc agtagctaat gaagaggtat      960
cctcactccc tccagagagc ggtccccctca cggctcattg agagtttgtc a          1011

```

<210> 63

<211> 1410

<212> ADN

5 <213> Homo sapiens

<220>

<221> característica diversa

<223> secuencia de STAR63

<400> 63

ES 2 384 391 T3

ccacagcctg atcgtgctgt cgatgagagg aatctgctct aagggctctga gcggagggag 60
atgccgaage tttgagcttt ttgtttctgg cttaaccttg gtggattttc accctctggg 120
cattacctct tgtccagggg aggggctggg ggagtgcctg gagctgtagg gacagagggc 180
tgagtggggg ggactgcttg ggctgaccac ataataattct gctgcgtatt aatttttttt 240
tgagacagtc tttctctggt gcccaggetg gagtgtaatg gcttgatagc tcaactgccac 300
ctccgcctcc tgggttcaag tgattctcct gcttcagctt ccggagtagc tgggactgca 360
ggtgccccgc accatggctg gctaattttt gtatttttat tagcaatggg gttttgctat 420
gttgcccagg ccgggtccga actcctgccc tcaagtgata cacctgcctc ggcctcccaa 480
agtgctggga ttagaggctt gagccaactgc gcctggccag ctgcatattg ttaattagac 540
ataaaatgca aaataagatg atataaacac aaaggtgtga aataagatgg acacctgctg 600
agcgcgcctg tcttgaagca tcgccccctc gcaaaagcag gggtcagcat gtgttctccg 660
gtccttgctc ttacagagga gtgagctgcc tatgcgtctt ccagccactt cctgggctgc 720
tcagaggcct ctacaggggtg ttctggggtg ctgccacttg caggggtgct gaggcggggc 780
tcctcccgtg cggggcatgt ccaggccgcc ctctctgaag gcttggcagg tacaggtggg 840
agtgggggtc tetgggctgc tgtggggact gggcaggctc ctggaagacc tccctgtggt 900
tgggctgaaa gcgcagcccc aggggaggtc cccagggagg ccgctgtcgg gggtaggggc 960
ttggaggagg gaggggccga ggagccggcg aactcctg acggcccagg aacgtcccta 1020
aacaaggcgc cgcgttctcg atgggggtgg gtccgcttc ttttctcaa agctgcagtt 1080
actccatgct cggaggactg gcgtccgcgc cctgttcaa tgctgccccg gggccctggc 1140
cttggggaat cggggccttg gactggaccc tgggggcttc gcggagccgg gcctggcggg 1200
gcgagcggag cagaggctgg gcagccccgg ggaagcgtc gccaaagccg ggcgctgctc 1260
ccagagcgcg aggtgcagaa ccagaggctg gtcccgcggc gctaacgaga gaagaggaag 1320
cgcgctgtgt agagggcgcc caccctgtgg ggcgaacccc ctctctcaac tccatggacg 1380
gggctcatgg gttcccagcg gctcagacgc 1410

<210> 64

<211> 1414

<212> ADN

5 <213> Homo sapiens

<220>

<221> característica diversa

<223> secuencia de STAR64

<400> 64

ES 2 384 391 T3

tggatcagat ttgttttata ccctcccttc tactgctctg agagttgtac atcacagtct 60
 actgtatctg tttcccatta ttataatfff ttgcaactgt gcttgccctga agggagcctc 120
 aagttcatga gtctccctac cctcctccca aatgagacat ggacctttga atgctttcct 180
 gggaccacca ccccaccttt catgctgctg ttatccagga ttttagttca acagtgtttt 240

 aaccccccaa atgagtcatt tttattgttt cgtatagtga atgtgtatft gggtttgett 300
 atatggtgac ctgtttatft gctcctcatt gtacctcatg ctctgctctt tccttctaga 360
 ttcagtctct ttcetaatga ggtgtctcgc agcaattctt tacaagacag ccaagatagg 420
 ccagctctca gagcaactgt tgtctgaaaa agtcttgtct tatttaatft ctttttctta 480
 gagatggggg ctcattatgt taccacact ggtctcaaac ttctggctta aagcggctct 540
 cccaccttgg cctcccaaag tgetaggatt acaggcgtga ggcacctcgt ccagcctgtc 600
 tgagaaagcg tttgttttgc ccttgctctc agatgacagt ttggggatag aattctaggt 660
 ggacggtttt tttccttcag ccctttgaag agtctgtatt ttcattatct ccttgcatta 720
 gatgttcttt tgcaagtaac gtgtcttttc tctctgggta ttcttaaggt tttctctttg 780
 cctttggtga gctgcagtgg atttgctttt ttcaagaggt caagagaaag gaaagtgtga 840
 ggtttctgtt ttttactgac aatttgtttg ttgatttgtt tccccacca gaggttctct 900
 gccactttgc caggctggaa ggcagacttc ttctggtgtc ctgttcacag acggggcagc 960
 ctgoggaagg ccctgccaca tgcagggcct cggctctcat tcccttgcat gtggaccggg 1020
 gcgtgactcc tgttcaggct ggcacttccc agagctgagc cccagcctga ccttctctcc 1080
 atactgtctt cacacccctt cctttcttct gatacctgga ggttttctt tctttctgt 1140
 cacctccact tggattttaa atcctctgtc tgtggaattg tatteggcac aggaagatgc 1200
 ttgcaagggc caggctcctc agccctgtcc ctgctgctgg aagcagcaca gcagagcctc 1260
 atgctcaggc tgagatggag cagaggcctg cagacgagca cccagctcag ctggggttgg 1320
 cgccgatggt ggagggctct cgaaagctct ggggacgatg gcagagctat tggcagggga 1380
 gccgcagggt cttttgagcc cttaaaagat ctct 1414

<210> 65

<211> 1310

5 <212> ADN

<213> Homo sapiens

<220>

<221> característica diversa

<223> secuencia de STAR65

10 <400> 65

ES 2 384 391 T3

gtgaatggtg atggatcaaa tatctttctg tgttgtttat caaagttaaa ataatgtgg 60
tcatttaaag gacaaaagat gaggggttg agtctgttca agcaaagggt atattaggag 120
aaaagcagaa ttctctccct gtgaaggac agtgactcct atttccacc tcatttttac 180
taactctcct aactatctgc ttaggtagag atatatccat gtacatttat aaaccacagt 240
gaatcatttg attttggat aaagatagta taaaatgtgt cccagtgttg atatacatca 300
tacattaaat atgtctggca gtgttctaata tttacagttg tccaaagata atggttagggc 360
atactggcta tggatgaagc tccaatgttc agattgcaaa gaaacttaga attttactaa 420
tgaaaccaaa tacatcccaa gaaattttc agaagaaaaa aagagaaact agtagcaaaag 480

taaagaatca ccacaatatc atcagatattt ttttatatgt agaatattta ttcagttctt 540
ttttcaagta caccttgtct tcattcattg tactttattt tttgtgaagg tttaaattta 600
tttcttctat gtgtttagtg atatttaaaa tttttattta atcaagttta tcagaaagt 660
ctggttagaaa atatgacgag gctttaattc cgccatctat attttccgct attatataaa 720
gataattggt ttctcttttt aaaacaactt gaattgggat tttatatcat aattttttaa 780
tgtctttttt tattataactt taagttctgg gatacatgtg cagaacgtgc aggtgtgtta 840
catagatata cacgtgccat ggtggtttgc tgcaccact aacctgttat cgacattagg 900
tatttctcct aatgctatca cccctattt cccaccccc cgagaggccc cagtgtgtga 960
tgttctctc cctgtgtcca tgtgttctca ttgttcatct cccaettatg gtatctacca 1020
taaccttgaa attgtcttat gcattcactt gtttggttgt tatatagcct ccatcaggac 1080
agggatattt gctgctgctt cttttttttt tctttttgag acagtcttgc tccgtcatcc 1140
aggctggagt gcttctcggc tcaatgcaac ctccacctcc caggtttaag cgattctcca 1200
acttcagcct cccaaatggc tgggactgca ggcattgcacc actacacctg gctaattttt 1260
gtatttgtaa tagagacaat gtttcacat gttggccagg ctggtctcga 1310

<210> 66

<211> 1917

5 <212> ADN

<213> Homo sapiens

<220>

<221> característica diversa

<223> secuencia de STAR67

10 <400> 66

ES 2 384 391 T3

aggatcctaa aattttgtga ccctagagca agtactaact atgaaagtga aatagagaat 60
 gaaggaatta ttaattaag tccagcaaaa cccaacaaa tcatctgtaa aatatatttg 120
 ttttcaacat ccaggtattt tctgtgtaaa aggttgagtt gtatgctgac ttattgggaa 180
 aaataattga gttttcccct tcaacttgcc agtgagagga aatcagtact gtaattgtta 240
 aaggttacco atacctacct ctactaccgt cttagcatagg taaagtaatg tacactgtga 300
 agtttctctgc ttgactgtaa tgttttcagt ttcattccat tgattcaaca gctatttatt 360
 cagcacttac tacaacctg ctggaaaccc aagagtaaat aggetgtggt actcaacagg 420
 actgaggtac agccgaactg tcaggcaagg ttgctgtcct ttggacttgc ctgctttctc 480
 tctatgtagg aagaagaaat ggacataccg tccaggaaat agatataatg tacatttccet 540
 tattccataa ttaatattaa taaccctgga cagaaactac caagtttcta gaccettata 600
 gtaccacett accctttctg gatgaatcct tcacatgttg atacatttta tccaaatgaa 660
 aattttggtg ctgtaggtat aacagacaaa gagagaacag aaaactagag atgaagtttg 720
 ggaaaaggtc aagaaagtaa ataatgcttc tagaagacac aaaaagaaaa atgaaatggt 780
 aatggtggga aagttttaat acattttgcc ctaaggaaaa aaactacttg ttgaaattct 840
 acttaagact ggaccttttc tctaaaaatt gtgcttgatg tgaattaaag caacacaggg 900

 aaatttatgg gctccttcta agttctaccc aactcaccgc aaaactgttc ctagttagtg 960
 tggataactc tttcagattc tttgtgtgta tgtatatgtg tgtgtgtgtg tgtgtttgta 1020
 tgtgtacagt ctatatacat atgtgtacct acatgtgtgt atatataaat atatatttac 1080
 ctggatgaaa tagcatatta tagaatattc ttttttcttt aatatatat gtgcatacat 1140
 atgtatatgc acatatatac ataatgtag atatagctag gtaggcattc atgtgaaaca 1200
 aagaagccta ttacttttta atggttgcat gatattccat cataggagta tagtacaact 1260
 tatgtaacac acatttggtc tgttgtaaaa ttttggtatt aataaaatag cacatatcat 1320
 gcaaagacac cttgcatag gtctattcat tctttgattt ttaccttagg acaaaattta 1380
 aaagtagaat ttctgggtca agcagtatgc tcatttaaaa tgcattgca tttttccaaa 1440
 ttgtcctcca gaaaagtagt aacagtaaca attgatggac tgcgtgtttt ctaaaacttg 1500
 catttttttc cttattggtg aggtttggca ttttccatat gtttattggc attttaattt 1560
 tttttgggtc atgtctttta ttcccttccct gcaaatttgt ggtgtgtctc aactttattt 1620
 atactotcat tttcataatt ttctaaagga atttgacttt aaaaaataa gacagccaat 1680
 gctttggttt aatttcattg ctgctttttg aagtgactgc tgtgttttta tatactttta 1740
 tattttggtg ttttagcaaa ttcttctata ttataattgt gtatgctgga acaaaaagtt 1800
 atattttcta atctagataa aatatttcaa gatgttgtaa ttacagtccc ctctaaaatc 1860
 atataaatag acgcatagct gtgtgatttg taattagtta tgtccattga tagatcc 1917

<210> 67

ES 2 384 391 T3

	<211> 26	
	<212> ADN	
	<213> Artificial	
	<220>	
5	<223> oligo para obtener el ligador que contiene MCSII de pd2EGFP-link	
	<400> 67	
	gtacggatat cagatcttta attaag	26
	<210> 68	
	<211> 24	
10	<212> ADN	
	<213> Artificial	
	<220>	
	<223> oligo para obtener el ligador que contiene MCSII de pd2EGFP-link	
	<400> 68	
15	gtaccttaat taaagatctg atat	24
	<210> 69	
	<211> 52	
	<212> ADN	
	<213> Artificial	
20	<220>	
	<223> cebador para la amplificación de 0,37 kb de pd2EGFP	
	<400> 69	
	gatcagatct ggcgcgcat taaatcgtc tcgcgctt cggatgatgc gg	52
	<210> 70	
25	<211> 41	
	<212> ADN	
	<213> Artificial	
	<220>	
	<223> cebador para la amplificación de 0,37 kb de pd2EGFP	
30	<400> 70	
	aggcggatcc gaatgtatt agaaaaataa acaaataagg g	41
	<210>71	
	<211> 36	
	<212> ADN	
35	<213> Artificial	
	<220>	
	<223> cebador para amplificar el ORF del gen de resistencia a zeocina	

ES 2 384 391 T3

	<400> 71	
	gatcggatcc ttcgaaatgg ccaagttgac cagtg	36
	<210> 72	
	<211> 32	
5	<212> ADN	
	<213> Artificial	
	<220>	
	<223> cebador para amplificar el ORF del gen de resistencia a zeocina	
	<400> 72	
10	aggcgcggcc gcaattctca gtcctgctcc tc	32
	<210> 73	
	<211> 40	
	<212> ADN	
	<213> Artificial	
15	<220>	
	<223> cebador para amplificar el ORF de d2EGFP	
	<400> 73	
	gatcgaattc tcgcaatgg tgagcaagca gatcctgaag	40
	<210> 74	
20	<211> 52	
	<212> ADN	
	<213> Artificial	
	<220>	
	<223> cebador para amplificar el ORF de d2EGFP	
25	<400> 74	
	aggcgaattc accggtgtt aaacttacac ccaactcgtc aggctgccca gg	52
	<210> 75	
	<211> 41	
	<212> ADN	
30	<213> Artificial	
	<220>	
	<223> cebador para amplificar el gen de d2EGFP	
	<400> 75	
	ttggttgtc atgaatggtg agcaagggcg aggagctgtt c	41
35	<210> 76	
	<211> 36	
	<212> ADN	

ES 2 384 391 T3

	<213> Artificial	
	<220>	
	<223> cebador para amplificar el gen de d2EGFP	
	<400> 76	
5	attctctaga ctacacattg atcctagcag aagcac	36
	<210> 77	
	<211> 42	
	<212> ADN	
	<213> Artificial	
10	<220>	
	<223> oligo para preparar el ligador para crear MCS en pGL3-promoter-GFP	
	<400> 77	
	cgatatcttg gagatctact agtggcgcg cttgggctag ct	42
	<210> 78	
15	<211> 50	
	<212> ADN	
	<213> Artificial	
	<220>	
	<223> oligo para preparar el ligador para crear MCS en pGL3-promoter-GFP	
20	<400> 78	
	gatcagctag cccaaggcgc gccactagta gatctccaag atactgagct	50
	<210> 79	
	<211> 40	
	<212> ADN	
25	<213> Artificial	
	<220>	
	<223> cebador para la amplificación del promotor de EF-1alfa	
	<400> 79	
	gatcggcgcg ccatttaat ccgaaaagtg ccactgacg	40
30	<210> 80	
	<211> 34	
	<212> ADN	
	<213> Artificial	
	<220>	
35	<223> cebador para la amplificación del promotor de EF-1alfa	
	<400> 80	
	aggcgggacc ccctcacgac acctgaaatg gaag	34

<210> 81
 <211> 51
 <212> ADN
 <213> Artificial
 5 <220>
 <223> cebador para la amplificación del promotor de SV40
 <400> 81
 ttggttgggg cgcgccgag caccatggcc tgaataacc tctgaaagag g 51
 <210> 82
 10 <211> 55
 <212> ADN
 <213> Artificial
 <220>
 <223> cebador para la amplificación del promotor de SV40
 15 <400> 82
 ttggttggga gctcaagctt ttgcaaaag cctaggcctc caaaaaagcc tctc 55
 <210> 83
 <211> 11
 <212> ADN
 20 <213> Artificial
 <220>
 <223> secuencia alrededor del codón de inicio del gen de Resistencia a zeocina de tipo salvaje
 <400> 83
 aaaccatggc c 11
 25 <210> 84
 <211> 50
 <212> ADN
 <213> Artificial
 <220>
 30 <223> cebador ZEOforwardMUT
 <400> 84
 gatctcgga tacaggattt atgttgcca agttgaccag tgccgttccg 50
 <210> 85
 <211> 26
 35 <212> ADN
 <213> Artificial
 <220>

ES 2 384 391 T3

	<223> cebador ZEO-WTreverse	
	<400> 85	
	aggcgaattc agtcctgctc ctcggc	26
	<210> 86	
5	<211> 45	
	<212> ADN	
	<213> Artificial	
	<220>	
	<223> cebador ZEO-LEUreverse	
10	<400> 86	
	aggccccgcc cccacggctg ctcgccgatc tcggtcaagg ccggc	45
	<210> 87	
	<211> 45	
	<212> ADN	
15	<213> Artificial	
	<220>	
	<223> cebador ZEO-THRreverse	
	<400> 87	
	aggccccgcc cccacggctg ctcgccgatc tcggtggtgg ccggc	45
20	<210> 88	
	<211> 44	
	<212> ADN	
	<213> Artificial	
	<220>	
25	<223> cebador ZEO-VALreverse	
	<400> 88	
	aggccccgcc cccacggctg ctcgccgatc tcggtccacg ccgg	44
	<210> 89	
	<211> 11	
30	<212> ADN	
	<213> Artificial	
	<220>	
	<223> secuencia alrededor del codón de inicio de d2EGFP de tipo salvaje	
	<400> 89	
35	gaattcatgg g	11
	<210> 90	
	<211> 49	

ES 2 384 391 T3

<212> ADN
 <213> Artificial
 <220>
 <223> cebador d2EGFPforwardBamHI
 5 <400> 90
 gatcggatcc tatgaggaat tcgccacat ggtgagcaag ggcgaggag 49
 <210> 91
 <211> 41
 <212> ADN
 10 <213> Artificial
 <220>
 <223> cebador d2EGFPreverseNotI
 <400> 91
 aaggaaaaa gcggccgct acacattgat cctagcagaa g 41
 15 <210> 92
 <211> 57
 <212> ADN
 <213> Artificial
 <220>
 20 <223> secuencia espaciadora
 <400> 92
 tcgatccaaa gactgcaaaa tctagatccg agattttcag gagctaagga agctaaa 57
 <210> 93
 <211> 99
 25 <212> ADN
 <213> Artificial
 <220>
 <223> cebador ZEOforwardBamHI-ATGmut/space
 <400> 93
 gatcggatcc ttggtttatg tcgatccaaa gactgcaaaa tctagatccg agattttcag 60
 gagctaagga agctaaagcc aagttgacca gtgaagttc 99
 30
 <210> 94
 <211> 38
 <212> ADN
 <213> Artificial
 35 <220>

ES 2 384 391 T3

	<223> cebador ZEOforwardBamHI-GTG	
	<400> 94	
	gatcggatcc accgtggcca agttgaccag tgccgttc	38
	<210> 95	
5	<211> 38	
	<212> ADN	
	<213> Artificial	
	<220>	
	<223> cebador ZEOforwardBamHI-TTG	
10	<400> 95	
	gatcggatcc accttgcca agttgaccag tgccgttc	38
	<210> 96	
	<211> 35	
	<212> ADN	
15	<213> Artificial	
	<220>	
	<223> cebador BSDBamHIforward	
	<400> 96	
	gatcggatcc accatggcca agccttgc tcaag	35
20	<210> 97	
	<211> 25	
	<212> ADN	
	<213> Artificial	
	<220>	
25	<223> cebador BSD150reverse	
	<400> 97	
	gtaaaatgat atacgtgac accag	25
	<210> 98	
	<211> 25	
30	<212> ADN	
	<213> Artificial	
	<220>	
	<223> cebador BSD150forward	
	<400> 98	
35	ctgggtcaa cgtatatcat ttac	25
	<210> 99	
	<211> 24	

ES 2 384 391 T3

<212> ADN
 <213> Artificial
 <220>
 <223> cebador BSD250reverse
 5 <400> 99
 gccctgttct cgttccgat cgcg 24
 <210> 100
 <211> 24
 <212> ADN
 10 <213> Artificial
 <220>
 <223> cebador BSD250forward
 <400> 100
 cgcgatcgga aacgagaaca gggc 24
 15 <210> 101
 <211> 24
 <212> ADN
 <213> Artificial
 <220>
 20 <223> cebador BSD350reverse
 <400> 101
 gccgtcggct gtccgtcact gtcc 24
 <210> 102
 <211> 24
 25 <212> ADN
 <213> Artificial
 <220>
 <223> cebador BSD350forward
 <400> 102
 30 ggacagtgac ggacagccga cggc 24
 <210> 103
 <211> 38
 <212> ADN
 <213> Artificial
 35 <220>
 <223> cebador BSD399reverse
 <400> 103

ES 2 384 391 T3

	gatcgaattc ttagccctcc cacacgtaac cagagggc	38
	<210> 104	
	<211> 103	
	<212> ADN	
5	<213> Artificial	
	<220>	
	<223> cebador BSDforwardBamHIAvrII-ATGmut/space	
	<400> 104	
	gatcggatcc taggttggtt tatgtcgatc caaagactgc caaatctaga tccgagattt	60
	tcaggagcta aggaagctaa agccaagcct ttgtctcaag aag	103
10	<210> 105	
	<211> 44	
	<212> ADN	
	<213> Artificial	
	<220>	
15	<223> cebador BSD399reverseEcoRIAvrII	
	<400> 105	
	gatcgaattc cctaggttag cctcccaca cgtaaccaga gggc	44
	<210> 106	
	<211> 42	
20	<212> ADN	
	<213> Artificial	
	<220>	
	<223> cebador BSDforwardBamHIAvrII-GTG	
	<400> 106	
25	gatcggatcc taggaccgtg gccaagcctt tgtctcaaga ag	42
	<210> 107	
	<211> 42	
	<212> ADN	
	<213> Artificial	
30	<220>	
	<223> cebador BSDforwardBamHIAvrII-TTG	
	<400> 107	
	gatcggatcc taggacctg gccaagcctt tgtctcaaga ag	42
	<210> 108	
35	<211>375	

ES 2 384 391 T3

<212> ADN

<213> Artificial

<220>

<223> gen de resistencia a zeocina de tipo salvaje

5 <220>

<221> CDS

<222> (1)..(375)

<400> 108

```

atg gcc aag ttg acc agt gcc gtt ccg gtg ctc acc gcg cgc gac gtc      48
Met Ala Lys Leu Thr Ser Ala Val Pro Val Leu Thr Ala Arg Asp Val
1          5          10          15

gcc gga gcg gtc gag ttc tgg acc gac cgg ctc ggg ttc tcc cgg gac      96
Ala Gly Ala Val Glu Phe Trp Thr Asp Arg Leu Gly Phe Ser Arg Asp
          20          25          30

ttc gtg gag gac gac ttc gcc ggt gtg gtc cgg gac gac gtg acc ctg      144
Phe Val Glu Asp Asp Phe Ala Gly Val Val Arg Asp Asp Val Thr Leu
          35          40          45

ttc atc agc gcg gtc cag gac cag gtg gtg ccg gac aac acc ctg gcc      192
Phe Ile Ser Ala Val Gln Asp Gln Val Val Pro Asp Asn Thr Leu Ala
          50          55          60

tgg gtg tgg gtg cgc ggc ctg gac gag ctg tac gcc gag tgg tcg gag      240
Trp Val Trp Val Arg Gly Leu Asp Glu Leu Tyr Ala Glu Trp Ser Glu
65          70          75          80

gtc gtg tcc acg aac ttc cgg gac gcc tcc ggg ccg gcc atg acc gag      288
Val Val Ser Thr Asn Phe Arg Asp Ala Ser Gly Pro Ala Met Thr Glu
          85          90          95

atc ggc gag cag ccg tgg ggg cgg gag ttc gcc ctg cgc gac ccg gcc      336
Ile Gly Glu Gln Pro Trp Gly Arg Glu Phe Ala Leu Arg Asp Pro Ala
          100          105          110

ggc aac tgc gtg cac ttc gtg gcc gag gag cag gac tga      375
Gly Asn Cys Val His Phe Val Ala Glu Glu Gln Asp
          115          120

```

10 <210> 109

<211> 124

<212> PRT

<213> Artificial

<220>

15 <223> gen de resistencia a zeocina de tipo salvaje

<400> 109

ES 2 384 391 T3

Met Ala Lys Leu Thr Ser Ala Val Pro Val Leu Thr Ala Arg Asp Val
 1 5 10 15
 Ala Gly Ala Val Glu Phe Trp Thr Asp Arg Leu Gly Phe Ser Arg Asp
 20 25 30
 Phe Val Glu Asp Asp Phe Ala Gly Val Val Arg Asp Asp Val Thr Leu
 35 40 45
 Phe Ile Ser Ala Val Gln Asp Gln Val Val Pro Asp Asn Thr Leu Ala
 50 55 60
 Trp Val Trp Val Arg Gly Leu Asp Glu Leu Tyr Ala Glu Trp Ser Glu
 65 70 75 80
 Val Val Ser Thr Asn Phe Arg Asp Ala Ser Gly Pro Ala Met Thr Glu
 85 90 95
 Ile Gly Glu Gln Pro Trp Gly Arg Glu Phe Ala Leu Arg Asp Pro Ala
 100 105 110
 Gly Asn Cys Val His Phe Val Ala Glu Glu Gln Asp
 115 120

<210> 110

<211>399

<212> ADN

5 <213> Artificial

<220>

<223> gen de resistencia a blasticidina de tipo salvaje

<220>

<221> CDS

10 <222> (1)..(399)

<400> 110

ES 2 384 391 T3

atg gcc aag cct ttg tct caa gaa gaa tcc acc ctc att gaa aga gca 48
 Met Ala Lys Pro Leu Ser Gln Glu Glu Ser Thr Leu Ile Glu Arg Ala
 1 5 10 15

acg gct aca atc aac agc atc ccc atc tct gaa gac tac agc gtc gcc 96
 Thr Ala Thr Ile Asn Ser Ile Pro Ile Ser Glu Asp Tyr Ser Val Ala
 20 25 30

agc gca gct ctc tct agc gac ggc cgc atc ttc act ggt gtc aat gta 144
 Ser Ala Ala Leu Ser Ser Asp Gly Arg Ile Phe Thr Gly Val Asn Val
 35 40 45

tat cat ttt act ggg gga cct tgt gca gaa ctc gtg gtg ctg ggc act 192
 Tyr His Phe Thr Gly Gly Pro Cys Ala Glu Leu Val Val Leu Gly Thr
 50 55 60

gct gct gct gcg gca gct ggc aac ctg act tgt atc gtc gcg atc gga 240
 Ala Ala Ala Ala Ala Ala Gly Asn Leu Thr Cys Ile Val Ala Ile Gly
 65 70 75 80

aat gag aac agg ggc atc ttg agc ccc tgc gga cgg tgc cga cag gtg 288
 Asn Glu Asn Arg Gly Ile Leu Ser Pro Cys Gly Arg Cys Arg Gln Val
 85 90 95

ctt ctc gat ctg cat cct ggg atc aaa gcc ata gtg aag gac agt gat 336
 Leu Leu Asp Leu His Pro Gly Ile Lys Ala Ile Val Lys Asp Ser Asp
 100 105 110

gga cag ccg acg gca gtt ggg att cgt gaa ttg ctg ccc tct ggt tat 384
 Gly Gln Pro Thr Ala Val Gly Ile Arg Glu Leu Leu Pro Ser Gly Tyr
 115 120 125

gtg tgg gag ggc taa 399
 Val Trp Glu Gly
 130

<210> 111

<211> 132

<212> PRT

5 <213> Artificial

<220>

<223> gen de resistencia a blasticidina de tipo salvaje

<400> 111

Met Ala Lys Pro Leu Ser Gln Glu Glu Ser Thr Leu Ile Glu Arg Ala
 1 5 10 15

Thr Ala Thr Ile Asn Ser Ile Pro Ile Ser Glu Asp Tyr Ser Val Ala
 20 25 30

Ser Ala Ala Leu Ser Ser Asp Gly Arg Ile Phe Thr Gly Val Asn Val

ES 2 384 391 T3

	35		40		45														
	Tyr	His	Phe	Thr	Gly	Gly	Pro	Cys	Ala	Glu	Leu	Val	Val	Leu	Gly	Thr			
	50						55					60							
	Ala	Ala	Ala	Ala	Ala	Ala	Gly	Asn	Leu	Thr	Cys	Ile	Val	Ala	Ile	Gly			
	65					70					75				80				
	Asn	Glu	Asn	Arg	Gly	Ile	Leu	Ser	Pro	Cys	Gly	Arg	Cys	Arg	Gln	Val			
					85					90					95				
	Leu	Leu	Asp	Leu	His	Pro	Gly	Ile	Lys	Ala	Ile	Val	Lys	Asp	Ser	Asp			
			100						105					110					
	Gly	Gln	Pro	Thr	Ala	Val	Gly	Ile	Arg	Glu	Leu	Leu	Pro	Ser	Gly	Tyr			
			115					120					125						
	Val	Trp	Glu	Gly															
	130																		

<210> 112

<211>600

<212> ADN

5 <213> Artificial

<220>

<223> gen de resistencia a puromicina de tipo salvaje

<220>

<221> CDS

10 <222> (1)..(600)

<400> 112

ES 2 384 391 T3

```

atg acc gag tac aag ccc acg gtg cgc ctc gcc acc cgc gac gac gtc      48
Met Thr Glu Tyr Lys Pro Thr Val Arg Leu Ala Thr Arg Asp Asp Val
1                               5                               10                               15

ccc agg gcc gta cgc acc ctc gcc gcc gcg ttc gcc gac tac ccc gcc      96
Pro Arg Ala Val Arg Thr Leu Ala Ala Ala Phe Ala Asp Tyr Pro Ala
                20                               25                               30

acg cgc cac acc gtc gat ccg gac cgc cac atc gag cgg gtc acc gag      144
Thr Arg His Thr Val Asp Pro Asp Arg His Ile Glu Arg Val Thr Glu
                35                               40                               45

ctg caa gaa ctc ttc ctc acg cgc gtc ggg ctc gac atc gcc aag gtg      192
Leu Gln Glu Leu Phe Leu Thr Arg Val Gly Leu Asp Ile Gly Lys Val
                50                               55                               60

tgg gtc gcg gac gac ggc gcc gcg gtg gcg gtc tgg acc acg ccg gag      240
Trp Val Ala Asp Asp Gly Ala Ala Val Ala Val Trp Thr Thr Pro Glu
65                               70                               75                               80

agc gtc gaa gcg ggg gcg gtg ttc gcc gag atc gcc ccg cgc atg gcc      288
Ser Val Glu Ala Gly Ala Val Phe Ala Glu Ile Gly Pro Arg Met Ala
                85                               90                               95

gag ttg agc ggt tcc ccg ctg gcc gcg cag caa cag atg gaa gcc ctc      336
Glu Leu Ser Gly Ser Arg Leu Ala Ala Gln Gln Gln Met Glu Gly Leu
                100                               105                               110

ctg gcg ccg cac cgg ccc aag gag ccc gcg tgg ttc ctg gcc acc gtc      384
Leu Ala Pro His Arg Pro Lys Glu Pro Ala Trp Phe Leu Ala Thr Val
                115                               120                               125

ggc gtc tcg ccc gac cac cag ggc aag ggt ctg ggc agc gcc gtc gtg      432
Gly Val Ser Pro Asp His Gln Gly Lys Gly Leu Gly Ser Ala Val Val
                130                               135                               140

ctc ccc gga gtg gag gcg gcc gag cgc gcc ggg gtg ccc gcc ttc ctg      480
Leu Pro Gly Val Glu Ala Ala Glu Arg Ala Gly Val Pro Ala Phe Leu
145                               150                               155                               160

gag acc tcc gcg ccc cgc aac ctc ccc ttc tac gag cgg ctc gcc ttc      528
Glu Thr Ser Ala Pro Arg Asn Leu Pro Phe Tyr Glu Arg Leu Gly Phe
                165                               170                               175

acc gtc acc gcc gac gtc gag tgc ccg aag gac cgc gcg acc tgg tgc      576
Thr Val Thr Ala Asp Val Glu Cys Pro Lys Asp Arg Ala Thr Trp Cys
                180                               185                               190

atg acc cgc aag ccc ggt gcc tga
Met Thr Arg Lys Pro Gly Ala
                195

```

<210> 113

<211> 199

5 <212> PRT

<213> Artificial

<220>

<223> gen de resistencia a puromicina de tipo salvaje

<400> 113

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

<210> 114

<211> 564

5 <212> ADN

<213> Artificial

<220>

<223> gen de DHFR de tipo salvaje (procedente de ratón)

<220>

10 <221> CDS

<222> (1)..(564)

<400> 114

ES 2 384 391 T3

atg gtt cga cca ttg aac tgc atc gtc gcc gtg tcc caa aat atg ggg 48
 Met Val Arg Pro Leu Asn Cys Ile Val Ala Val Ser Gln Asn Met Gly
 1 5 10 15

att ggc aag aac gga gac cta ccc tgg cct ccg ctc agg aac gag ttc 96
 Ile Gly Lys Asn Gly Asp Leu Pro Trp Pro Pro Leu Arg Asn Glu Phe
 20 25 30

aag tac ttc caa aga atg acc aca acc tct tca gtg gaa ggt aaa cag 144
 Lys Tyr Phe Gln Arg Met Thr Thr Thr Ser Ser Val Glu Gly Lys Gln
 35 40 45

aat ctg gtg att atg ggt agg aaa acc tgg ttc tcc att cct gag aag 192
 Asn Leu Val Ile Met Gly Arg Lys Thr Trp Phe Ser Ile Pro Glu Lys
 50 55 60

aat cga cct tta aag gac aga att aat ata gtt ctc agt aga gaa ctc 240
 Asn Arg Pro Leu Lys Asp Arg Ile Asn Ile Val Leu Ser Arg Glu Leu
 65 70 75 80

aaa gaa cca cca cga gga gct cat ttt ctt gcc aaa agt ttg gat gat 288
 Lys Glu Pro Pro Arg Gly Ala His Phe Leu Ala Lys Ser Leu Asp Asp
 85 90 95

gcc tta aga ctt att gaa caa ccg gaa ttg gca agt aaa gta gac atg 336
 Ala Leu Arg Leu Ile Glu Gln Pro Glu Leu Ala Ser Lys Val Asp Met
 100 105 110

gtt tgg ata gtc gga ggc agt tct gtt tac cag gaa gcc atg aat caa 384
 Val Trp Ile Val Gly Gly Ser Ser Val Tyr Gln Glu Ala Met Asn Gln
 115 120 125

cca ggc cac ctc aga ctc ttt gtg aca agg atc atg cag gaa ttt gaa 432
 Pro Gly His Leu Arg Leu Phe Val Thr Arg Ile Met Gln Glu Phe Glu
 130 135 140

agt gac acg ttt ttc cca gaa att gat ttg ggg aaa tat aaa ctt ctc 480
 Ser Asp Thr Phe Phe Pro Glu Ile Asp Leu Gly Lys Tyr Lys Leu Leu
 145 150 155 160

cca gaa tac cca ggc gtc ctc tct gag gtc cag gag gaa aaa ggc atc 528
 Pro Glu Tyr Pro Gly Val Leu Ser Glu Val Gln Glu Glu Lys Gly Ile
 165 170 175

aag tat aag ttt gaa gtc tac gag aag aaa gac taa 564
 Lys Tyr Lys Phe Glu Val Tyr Glu Lys Lys Asp
 180 185

<210> 115

<211> 187

<212> PRT

5 <213> Artificial

<220>

<223> gen de DHFR de tipo salvaje (procedente de ratón)

<400> 115

ES 2 384 391 T3

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

<210> 116

<211> 1143

<212> ADN

5 <213> Artificial

<220>

<223> gen de resistencia a higromicina de tipo salvaje

<220>

<221> CDS

10 <222> (1)..(1143)

<400> 116

ES 2 384 391 T3

atg	aaa	aag	cct	gaa	ctc	acc	gcg	acg	tct	gtc	gag	aag	ttt	ctg	atc	48
Met	Lys	Lys	Pro	Glu	Leu	Thr	Ala	Thr	Ser	Val	Glu	Lys	Phe	Leu	Ile	
1				5					10					15		
gaa	aag	ttc	gac	agc	gtc	tcc	gac	ctg	atg	cag	ctc	tcg	gag	ggc	gaa	96
Glu	Lys	Phe	Asp	Ser	Val	Ser	Asp	Leu	Met	Gln	Leu	Ser	Glu	Gly	Glu	
			20					25					30			
gaa	tct	cgt	gct	ttc	agc	ttc	gat	gta	gga	ggg	cgt	gga	tat	gtc	ctg	144
Glu	Ser	Arg	Ala	Phe	Ser	Phe	Asp	Val	Gly	Gly	Arg	Gly	Tyr	Val	Leu	
		35					40					45				
cgg	gta	aat	agc	tgc	gcc	gat	ggt	ttc	tac	aaa	gat	cgt	tat	gtt	tat	192

ES 2 384 391 T3

Arg	Val	Asn	Ser	Cys	Ala	Asp	Gly	Phe	Tyr	Lys	Asp	Arg	Tyr	Val	Tyr		
50						55					60						
cgg	cac	ttt	gca	tcg	gcc	gcg	ctc	ccg	att	ccg	gaa	gtg	ctt	gac	att		240
Arg	His	Phe	Ala	Ser	Ala	Ala	Leu	Pro	Ile	Pro	Glu	Val	Leu	Asp	Ile		
65					70				75					80			
ggg	gaa	ttc	agc	gag	agc	ctg	acc	tat	tgc	atc	tcc	cgc	cgt	gca	cag		288
Gly	Glu	Phe	Ser	Glu	Ser	Leu	Thr	Tyr	Cys	Ile	Ser	Arg	Arg	Ala	Gln		
				85					90					95			
ggt	gtc	acg	ttg	caa	gac	ctg	cct	gaa	acc	gaa	ctg	ccc	gct	gtt	ctg		336
Gly	Val	Thr	Leu	Gln	Asp	Leu	Pro	Glu	Thr	Glu	Leu	Pro	Ala	Val	Leu		
			100					105					110				
cag	ccg	gtc	gcg	gag	gcc	atg	gat	gcg	atc	gct	gcg	gcc	gat	ctt	agc		384
Gln	Pro	Val	Ala	Glu	Ala	Met	Asp	Ala	Ile	Ala	Ala	Ala	Asp	Leu	Ser		
		115					120					125					
cag	acg	agc	ggg	ttc	ggc	cca	ttc	gga	ccg	caa	gga	atc	ggt	caa	tac		432
Gln	Thr	Ser	Gly	Phe	Gly	Pro	Phe	Gly	Pro	Gln	Gly	Ile	Gly	Gln	Tyr		
	130					135					140						
act	aca	tgg	cgt	gat	ttc	ata	tgc	gcg	att	gct	gat	ccc	cat	gtg	tat		480
Thr	Thr	Trp	Arg	Asp	Phe	Ile	Cys	Ala	Ile	Ala	Asp	Pro	His	Val	Tyr		
145					150					155					160		
cac	tgg	caa	act	gtg	atg	gac	gac	acc	gtc	agt	gcg	tcc	gtc	gcg	cag		528
His	Trp	Gln	Thr	Val	Met	Asp	Asp	Thr	Val	Ser	Ala	Ser	Val	Ala	Gln		
				165					170					175			
gct	ctc	gat	gag	ctg	atg	ctt	tgg	gcc	gag	gac	tgc	ccc	gaa	gtc	cgg		576
Ala	Leu	Asp	Glu	Leu	Met	Leu	Trp	Ala	Glu	Asp	Cys	Pro	Glu	Val	Arg		
			180					185					190				
cac	ctc	gtg	cac	gcg	gat	ttc	ggc	tcc	aac	aat	gtc	ctg	acg	gac	aat		624
His	Leu	Val	His	Ala	Asp	Phe	Gly	Ser	Asn	Asn	Val	Leu	Thr	Asp	Asn		
		195					200					205					
ggc	cgc	ata	aca	gcg	gtc	att	gac	tgg	agc	gag	gcg	atg	ttc	ggg	gat		672
Gly	Arg	Ile	Thr	Ala	Val	Ile	Asp	Trp	Ser	Glu	Ala	Met	Phe	Gly	Asp		
	210					215					220						
tcc	caa	tac	gag	gtc	gcc	aac	atc	ttc	ttc	tgg	agg	ccg	tgg	ttg	gct		720
Ser	Gln	Tyr	Glu	Val	Ala	Asn	Ile	Phe	Phe	Trp	Arg	Pro	Trp	Leu	Ala		
225					230					235				240			
tgt	atg	gag	cag	cag	acg	cgc	tac	ttc	gag	cgg	agg	cat	ccg	gag	ctt		768
Cys	Met	Glu	Gln	Gln	Thr	Arg	Tyr	Phe	Glu	Arg	Arg	His	Pro	Glu	Leu		
				245					250					255			
gca	gga	tcg	ccg	cgg	ctc	cgg	gcg	tat	atg	ctc	cgc	att	ggt	ctt	gac		816
Ala	Gly	Ser	Pro	Arg	Leu	Arg	Ala	Tyr	Met	Leu	Arg	Ile	Gly	Leu	Asp		
			260					265					270				
caa	ctc	tat	cag	agc	ttg	gtt	gac	ggc	aat	ttc	gat	gat	gca	gct	tgg		864
Gln	Leu	Tyr	Gln	Ser	Leu	Val	Asp	Gly	Asn	Phe	Asp	Asp	Ala	Ala	Trp		
		275					280					285					
gcg	cag	ggt	cga	tgc	gac	gca	atc	gtc	cga	tcc	gga	gcc	ggg	act	gtc		912
Ala	Gln	Gly	Arg	Cys	Asp	Ala	Ile	Val	Arg	Ser	Gly	Ala	Gly	Thr	Val		
	290					295					300						
ggg	cgt	aca	caa	atc	gcc	cgc	aga	agc	gcg	gcc	gtc	tgg	acc	gat	ggc		960
Gly	Arg	Thr	Gln	Ile	Ala	Arg	Arg	Ser	Ala	Ala	Val	Trp	Thr	Asp	Gly		
305					310					315					320		
tgt	gta	gaa	gta	ctc	gcc	gat	agt	gga	aac	cga	cgc	ccc	agc	act	cgt		1008

ES 2 384 391 T3

Cys	Val	Glu	Val	Leu	Ala	Asp	Ser	Gly	Asn	Arg	Arg	Pro	Ser	Thr	Arg	
				325					330					335		
ccg	gag	gca	aag	gaa	ttc	ggg	aga	tgg	ggg	agg	cta	act	gaa	aca	cgg	1056
Pro	Glu	Ala	Lys	Glu	Phe	Gly	Arg	Trp	Gly	Arg	Leu	Thr	Glu	Thr	Arg	
			340					345					350			
aag	gag	aca	ata	ccg	gaa	gga	acc	cgc	gct	atg	acg	gca	ata	aaa	aga	1104
Lys	Glu	Thr	Ile	Pro	Glu	Gly	Thr	Arg	Ala	Met	Thr	Ala	Ile	Lys	Arg	
		355					360					365				
cag	aat	aaa	acg	cac	ggg	tgt	tgg	gtc	gtt	tgt	tca	taa				1143
Gln	Asn	Lys	Thr	His	Gly	Cys	Trp	Val	Val	Cys	Ser					
	370					375					380					

<210> 117

<211> 380

<212> PRT

5 <213> Artificial

<220>

<223> gen de resistencia a higromicina de tipo salvaje

<400> 117

ES 2 384 391 T3

Met Lys Lys Pro Glu Leu Thr Ala Thr Ser Val Glu Lys Phe Leu Ile
 1 5 10 15

Glu Lys Phe Asp Ser Val Ser Asp Leu Met Gln Leu Ser Glu Gly Glu
 20 25 30

Glu Ser Arg Ala Phe Ser Phe Asp Val Gly Gly Arg Gly Tyr Val Leu
 35 40 45

Arg Val Asn Ser Cys Ala Asp Gly Phe Tyr Lys Asp Arg Tyr Val Tyr
 50 55 60

Arg His Phe Ala Ser Ala Ala Leu Pro Ile Pro Glu Val Leu Asp Ile
 65 70 75 80

Gly Glu Phe Ser Glu Ser Leu Thr Tyr Cys Ile Ser Arg Arg Ala Gln
 85 90 95

Gly Val Thr Leu Gln Asp Leu Pro Glu Thr Glu Leu Pro Ala Val Leu
 100 105 110

Gln Pro Val Ala Glu Ala Met Asp Ala Ile Ala Ala Ala Asp Leu Ser
 115 120 125

Gln Thr Ser Gly Phe Gly Pro Phe Gly Pro Gln Gly Ile Gly Gln Tyr
 130 135 140

Thr Thr Trp Arg Asp Phe Ile Cys Ala Ile Ala Asp Pro His Val Tyr
 145 150 155 160

His Trp Gln Thr Val Met Asp Asp Thr Val Ser Ala Ser Val Ala Gln
 165 170 175

Ala Leu Asp Glu Leu Met Leu Trp Ala Glu Asp Cys Pro Glu Val Arg
 180 185 190

His Leu Val His Ala Asp Phe Gly Ser Asn Asn Val Leu Thr Asp Asn
 195 200 205

Gly Arg Ile Thr Ala Val Ile Asp Trp Ser Glu Ala Met Phe Gly Asp
 210 215 220

ES 2 384 391 T3

Ser Gln Tyr Glu Val Ala Asn Ile Phe Phe Trp Arg Pro Trp Leu Ala
 225 230 235 240
 Cys Met Glu Gln Gln Thr Arg Tyr Phe Glu Arg Arg His Pro Glu Leu
 245 250 255
 Ala Gly Ser Pro Arg Leu Arg Ala Tyr Met Leu Arg Ile Gly Leu Asp
 260 265 270
 Gln Leu Tyr Gln Ser Leu Val Asp Gly Asn Phe Asp Asp Ala Ala Trp
 275 280 285
 Ala Gln Gly Arg Cys Asp Ala Ile Val Arg Ser Gly Ala Gly Thr Val
 290 295 300
 Gly Arg Thr Gln Ile Ala Arg Arg Ser Ala Ala Val Trp Thr Asp Gly
 305 310 315 320
 Cys Val Glu Val Leu Ala Asp Ser Gly Asn Arg Arg Pro Ser Thr Arg
 325 330 335
 Pro Glu Ala Lys Glu Phe Gly Arg Trp Gly Arg Leu Thr Glu Thr Arg
 340 345 350
 Lys Glu Thr Ile Pro Glu Gly Thr Arg Ala Met Thr Ala Ile Lys Arg
 355 360 365
 Gln Asn Lys Thr His Gly Cys Trp Val Val Cys Ser
 370 375 380

<210> 118

<211> 804

<212> ADN

5 <213> Artificial

<220>

<223> gen de resistencia a neomicina de tipo salvaje

<220>

<221> CDS

10 <222> (1)..(804)

<400> 118

ES 2 384 391 T3

atg gga tcg gcc att gaa caa gat gga ttg cac gca ggt tct ccg gcc	48
Met Gly Ser Ala Ile Glu Gln Asp Gly Leu His Ala Gly Ser Pro Ala	
1 5 10 15	
gct tgg gtg gag agg cta ttc ggc tat gac tgg gca caa cag aca atc	96
Ala Trp Val Glu Arg Leu Phe Gly Tyr Asp Trp Ala Gln Gln Thr Ile	
20 25 30	
ggc tgc tct gat gcc gcc gtg ttc cgg ctg tca gcg cag ggg cgc ccg	144
Gly Cys Ser Asp Ala Ala Val Phe Arg Leu Ser Ala Gln Gly Arg Pro	
35 40 45	
gtt ctt ttt gtc aag acc gac ctg tcc ggt gcc ctg aat gaa ctg cag	192
Val Leu Phe Val Lys Thr Asp Leu Ser Gly Ala Leu Asn Glu Leu Gln	
50 55 60	
gac gag gca gcg cgg cta tcg tgg ctg gcc acg acg ggc gtt cct tgc	240
Asp Glu Ala Ala Arg Leu Ser Trp Leu Ala Thr Thr Gly Val Pro Cys	
65 70 75 80	
gca gct gtg ctc gac gtt gtc act gaa gcg gga agg gac tgg ctg cta	288
Ala Ala Val Leu Asp Val Val Thr Glu Ala Gly Arg Asp Trp Leu Leu	
85 90 95	

ES 2 384 391 T3

ttg ggc gaa gtg ccg ggg cag gat ctc ctg tca tct cac ctt gct cct	336
Leu Gly Glu Val Pro Gly Gln Asp Leu Leu Ser Ser His Leu Ala Pro	
100 105 110	
gcc gag aaa gta tcc atc atg gct gat gca atg cgg cgg ctg cat acg	384
Ala Glu Lys Val Ser Ile Met Ala Asp Ala Met Arg Arg Leu His Thr	
115 120 125	
ctt gat ccg gct acc tgc cca ttc gac cac caa gcg aaa cat cgc atc	432
Leu Asp Pro Ala Thr Cys Pro Phe Asp His Gln Ala Lys His Arg Ile	
130 135 140	
gag cga gca cgt act cgg atg gaa gcc ggt ctt gtc gat cag gat gat	480
Glu Arg Ala Arg Thr Arg Met Glu Ala Gly Leu Val Asp Gln Asp Asp	
145 150 155 160	
ctg gac gaa gag cat cag ggg ctc gcg cca gcc gaa ctg ttc gcc agg	528
Leu Asp Glu Glu His Gln Gly Leu Ala Pro Ala Glu Leu Phe Ala Arg	
165 170 175	
ctc aag gcg cgc atg ccc gac ggc gat gat ctc gtc gtg acc cat ggc	576
Leu Lys Ala Arg Met Pro Asp Gly Asp Asp Leu Val Val Thr His Gly	
180 185 190	
gat gcc tgc ttg ccg aat atc atg gtg gaa aat ggc cgc ttt tct gga	624
Asp Ala Cys Leu Pro Asn Ile Met Val Glu Asn Gly Arg Phe Ser Gly	
195 200 205	
ttc atc gac tgt ggc cgg ctg ggt gtg gcg gac cgc tat cag gac ata	672
Phe Ile Asp Cys Gly Arg Leu Gly Val Ala Asp Arg Tyr Gln Asp Ile	
210 215 220	
gcg ttg gct acc cgt gat att gct gaa gag ctt ggc ggc gaa tgg gct	720
Ala Leu Ala Thr Arg Asp Ile Ala Glu Glu Leu Gly Gly Glu Trp Ala	
225 230 235 240	
gac cgc ttc ctc gtg ctt tac ggt atc gcc gct ccc gat tgc cag cgc	768
Asp Arg Phe Leu Val Leu Tyr Gly Ile Ala Ala Pro Asp Ser Gln Arg	
245 250 255	
atc gcc ttc tat cgc ctt ctt gac gag ttc ttc tga	804
Ile Ala Phe Tyr Arg Leu Leu Asp Glu Phe Phe	
260 265	

<210> 119

<211> 267

<212> PRT

5 <213> Artificial

<220>

<223> gen de resistencia a neomicina de tipo salvaje

<400> 119

ES 2 384 391 T3

Met Gly Ser Ala Ile Glu Gln Asp Gly Leu His Ala Gly Ser Pro Ala
 1 5 10 15
 Ala Trp Val Glu Arg Leu Phe Gly Tyr Asp Trp Ala Gln Gln Thr Ile
 20 25 30
 Gly Cys Ser Asp Ala Ala Val Phe Arg Leu Ser Ala Gln Gly Arg Pro
 35 40 45
 Val Leu Phe Val Lys Thr Asp Leu Ser Gly Ala Leu Asn Glu Leu Gln
 50 55 60
 Asp Glu Ala Ala Arg Leu Ser Trp Leu Ala Thr Thr Gly Val Pro Cys
 65 70 75 80
 Ala Ala Val Leu Asp Val Val Thr Glu Ala Gly Arg Asp Trp Leu Leu
 85 90 95
 Leu Gly Glu Val Pro Gly Gln Asp Leu Leu Ser Ser His Leu Ala Pro
 100 105 110
 Ala Glu Lys Val Ser Ile Met Ala Asp Ala Met Arg Arg Leu His Thr
 115 120 125
 Leu Asp Pro Ala Thr Cys Pro Phe Asp His Gln Ala Lys His Arg Ile
 130 135 140
 Glu Arg Ala Arg Thr Arg Met Glu Ala Gly Leu Val Asp Gln Asp Asp
 145 150 155 160
 Leu Asp Glu Glu His Gln Gly Leu Ala Pro Ala Glu Leu Phe Ala Arg
 165 170 175
 Leu Lys Ala Arg Met Pro Asp Gly Asp Asp Leu Val Val Thr His Gly
 180 185 190
 Asp Ala Cys Leu Pro Asn Ile Met Val Glu Asn Gly Arg Phe Ser Gly
 195 200 205
 Phe Ile Asp Cys Gly Arg Leu Gly Val Ala Asp Arg Tyr Gln Asp Ile
 210 215 220
 Ala Leu Ala Thr Arg Asp Ile Ala Glu Glu Leu Gly Gly Glu Trp Ala
 225 230 235 240
 Asp Arg Phe Leu Val Leu Tyr Gly Ile Ala Ala Pro Asp Ser Gln Arg
 245 250 255
 Ile Ala Phe Tyr Arg Leu Leu Asp Glu Phe Phe
 260 265

<210> 120

<211> 1121

5 <212> ADN

<213> Artificial

<220>

<223> gen de glutamina sintasa de tipo salvaje (humano)

<220>

ES 2 384 391 T3

<221> CDS

<222> (1)..(1119)

<400> 120

atg acc acc tca gca agt tcc cac tta aat aaa ggc atc aag cag gtg	48
Met Thr Thr Ser Ala Ser Ser His Leu Asn Lys Gly Ile Lys Gln Val	
1 5 10 15	
tac atg tcc ctg cct cag ggt gag aaa gtc cag gcc atg tat atc tgg	96
Tyr Met Ser Leu Pro Gln Gly Glu Lys Val Gln Ala Met Tyr Ile Trp	
20 25 30	
atc gat ggt act gga gaa gga ctg cgc tgc aag acc cgg acc ctg gac	144
Ile Asp Gly Thr Gly Glu Gly Leu Arg Cys Lys Thr Arg Thr Leu Asp	
35 40 45	
agt gag ccc aag tgt gtg gaa gag ttg cct gag tgg aat ttc gat ggc	192
Ser Glu Pro Lys Cys Val Glu Glu Leu Pro Glu Trp Asn Phe Asp Gly	
50 55 60	

ES 2 384 391 T3

tcc agt act tta cag tct gag ggt tcc aac agt gac atg tat ctc gtg	240
Ser Ser Thr Leu Gln Ser Glu Gly Ser Asn Ser Asp Met Tyr Leu Val	
65 70 75 80	
cct gct gcc atg ttt cgg gac ccc ttc cgt aag gac cct aac aag ctg	288
Pro Ala Ala Met Phe Arg Asp Pro Phe Arg Lys Asp Pro Asn Lys Leu	
85 90 95	
gtg tta tgt gaa gtt ttc aag tac aat cga agg cct gca gag acc aat	336
Val Leu Cys Glu Val Phe Lys Tyr Asn Arg Arg Pro Ala Glu Thr Asn	
100 105 110	
ttg agg cac acc tgt aaa cgg ata atg gac atg gtg agc aac cag cac	384
Leu Arg His Thr Cys Lys Arg Ile Met Asp Met Val Ser Asn Gln His	
115 120 125	
ccc tgg ttt ggc atg gag cag gag tat acc ctc atg ggg aca gat ggg	432
Pro Trp Phe Gly Met Glu Gln Glu Tyr Thr Leu Met Gly Thr Asp Gly	
130 135 140	
cac ccc ttt ggt tgg cct tcc aac ggc ttc cca ggg ccc cag ggt cca	480
His Pro Phe Gly Trp Pro Ser Asn Gly Phe Pro Gly Pro Gln Gly Pro	
145 150 155 160	
tat tac tgt ggt gtg gga gca gac aga gcc tat ggc agg gac atc gtg	528
Tyr Tyr Cys Gly Val Gly Ala Asp Arg Ala Tyr Gly Arg Asp Ile Val	
165 170 175	
gag gcc cat tac cgg gcc tgc ttg tat gct gga gtc aag att gcg ggg	576
Glu Ala His Tyr Arg Ala Cys Leu Tyr Ala Gly Val Lys Ile Ala Gly	
180 185 190	
act aat gcc gag gtc atg cct gcc cag tgg gaa ttt cag att gga cct	624
Thr Asn Ala Glu Val Met Pro Ala Gln Trp Glu Phe Gln Ile Gly Pro	
195 200 205	
tgt gaa gga atc agc atg gga gat cat ctc tgg gtg gcc cgt ttc atc	672
Cys Glu Gly Ile Ser Met Gly Asp His Leu Trp Val Ala Arg Phe Ile	
210 215 220	
ttg cat cgt gtg tgt gaa gac ttt gga gtg ata gca acc ttt gat cct	720
Leu His Arg Val Cys Glu Asp Phe Gly Val Ile Ala Thr Phe Asp Pro	
225 230 235 240	
aag ccc att cct ggg aac tgg aat ggt gca ggc tgc cat acc aac ttc	768
Lys Pro Ile Pro Gly Asn Trp Asn Gly Ala Gly Cys His Thr Asn Phe	
245 250 255	
agc acc aag gcc atg cgg gag gag aat ggt ctg aag tac atc gag gag	816
Ser Thr Lys Ala Met Arg Glu Glu Asn Gly Leu Lys Tyr Ile Glu Glu	
260 265 270	
gcc att gag aaa cta agc aag cgg cac cag tac cac atc cgt gcc tat	864
Ala Ile Glu Lys Leu Ser Lys Arg His Gln Tyr His Ile Arg Ala Tyr	
275 280 285	
gat ccc aag gga ggc ctg gac aat gcc cga cgt cta act gga ttc cat	912
Asp Pro Lys Gly Gly Leu Asp Asn Ala Arg Arg Leu Thr Gly Phe His	
290 295 300	
gaa acc tcc aac atc aac gac ttt tct ggt ggt gta gcc aat cgt agc	960
Glu Thr Ser Asn Ile Asn Asp Phe Ser Gly Gly Val Ala Asn Arg Ser	
305 310 315 320	
gcc agc ata cgc att ccc cgg act gtt ggc cag gag aag aag ggt tac	1008
Ala Ser Ile Arg Ile Pro Arg Thr Val Gly Gln Glu Lys Lys Gly Tyr	
325 330 335	

ES 2 384 391 T3

ttt gaa gat cgt cgc ccc tct gcc aac tgc gac ccc ttt tcg gtg aca 1056
Phe Glu Asp Arg Arg Pro Ser Ala Asn Cys Asp Pro Phe Ser Val Thr
340 345 350

gaa gcc ctc atc cgc acg tgt ctt ctc aat gaa acc ggc gat gag ccc 1104
Glu Ala Leu Ile Arg Thr Cys Leu Leu Asn Glu Thr Gly Asp Glu Pro
355 360 365

ttc cag tac aaa aat ta 1121
Phe Gln Tyr Lys Asn
370

<210> 121

<211> 373

<212> PRT

5 <213> Artificial

<220>

<223> gen de glutamina sintasa de tipo salvaje (humano)

<400> 121

ES 2 384 391 T3

Met Thr Thr Ser Ala Ser Ser His Leu Asn Lys Gly Ile Lys Gln Val
1 5 10 15
Tyr Met Ser Leu Pro Gln Gly Glu Lys Val Gln Ala Met Tyr Ile Trp
20 25 30
Ile Asp Gly Thr Gly Glu Gly Leu Arg Cys Lys Thr Arg Thr Leu Asp
35 40 45
Ser Glu Pro Lys Cys Val Glu Glu Leu Pro Glu Trp Asn Phe Asp Gly
50 55 60
Ser Ser Thr Leu Gln Ser Glu Gly Ser Asn Ser Asp Met Tyr Leu Val
65 70 75 80
Pro Ala Ala Met Phe Arg Asp Pro Phe Arg Lys Asp Pro Asn Lys Leu
85 90 95
Val Leu Cys Glu Val Phe Lys Tyr Asn Arg Arg Pro Ala Glu Thr Asn
100 105 110
Leu Arg His Thr Cys Lys Arg Ile Met Asp Met Val Ser Asn Gln His
115 120 125
Pro Trp Phe Gly Met Glu Gln Glu Tyr Thr Leu Met Gly Thr Asp Gly
130 135 140
His Pro Phe Gly Trp Pro Ser Asn Gly Phe Pro Gly Pro Gln Gly Pro
145 150 155 160
Tyr Tyr Cys Gly Val Gly Ala Asp Arg Ala Tyr Gly Arg Asp Ile Val
165 170 175
Glu Ala His Tyr Arg Ala Cys Leu Tyr Ala Gly Val Lys Ile Ala Gly
180 185 190
Thr Asn Ala Glu Val Met Pro Ala Gln Trp Glu Phe Gln Ile Gly Pro
195 200 205
Cys Glu Gly Ile Ser Met Gly Asp His Leu Trp Val Ala Arg Phe Ile
210 215 220
Leu His Arg Val Cys Glu Asp Phe Gly Val Ile Ala Thr Phe Asp Pro
225 230 235 240

ES 2 384 391 T3

<223> cebador ZEOForwardGTG-Thr9
 <400> 124
 aattggatcc accgtggcca agtgaccag tgccgttacc gtgctc 46
 <210> 125
 5 <211> 46
 <212> ADN
 <213> Artificial
 <220>
 <223> cebador ZEOForward GTG-Phe9
 10 <400> 125
 aattggatcc accgtggcca agtgaccag tgccgtttc gtgctc 46
 <210> 126
 <211> 43
 <212> ADN
 15 <213> Artificial
 <220>
 <223> cebador TTGspaceBamHIF
 <400> 126
 gaattcggat ccacctggc gatccaaaga ctgccaatc tag 43
 20 <210> 127
 <211> 46
 <212> ADN
 <213> Artificial
 <220>
 25 <223> cebador ZEOForwardTTG-Thr9
 <400> 127
 aattggatcc acctggcca agtgaccag tgccgttacc gtgctc 46
 <210> 128
 <211> 46
 30 <212> ADN
 <213> Artificial
 <220>
 <223> cebador ZEOForwardTTG-Phe9
 <400> 128
 35 aattggatcc acctggcca agtgaccag tgccgtttc gtgctc 46
 <210> 129
 <211> 37

ES 2 384 391 T3

<212> ADN
 <213> Artificial
 <220>
 <223> cebador PURO BamHI F
 5 <400> 129
 gatcggatcc atggttaccg agtacaagcc cacggtg 37
 <210> 130
 <211> 35
 <212> ADN
 10 <213> Artificial
 <220>
 <223> cebador PURO300 R LEU
 <400> 130
 cagccgggaa cgcctcaact cggccaggcg cgggc 35
 15 <210> 131
 <211> 49
 <212> ADN
 <213> Artificial
 <220>
 20 <223> cebador PUR0300FLEU
 <400> 131
 cgagttgagc ggttcccggc tggccgcgca gcaacagctg gaaggcctc 49
 <210> 132
 <211> 44
 25 <212> ADN
 <213> Artificial
 <220>
 <223> cebador PURO600RLEU
 <400> 132
 30 aagcttgaat tcaggcaccg ggcttgcggg tcaggcacca ggtc 44
 <210> 133
 <211> 42
 <212> ADN
 <213> Artificial
 35 <220>
 <223> cebador PUROBamHI TTG1F
 <400> 133

ES 2 384 391 T3

gaattcggat ccaccttggg taccgagtac aagcccacgg tg

42

<210> 134

<211> 804

<212> ADN

5 <213> Artificial

<220>

<223> gen de resistencia a neomicina modificado que carece de secuencias ATG internas

<220>

<221> CDS

10 <222> (1)..(804)

<400> 134

atg gga tcg gcc att gaa caa gac gga ttg cac gca ggt tct ccg gcc	48
Met Gly Ser Ala Ile Glu Gln Asp Gly Leu His Ala Gly Ser Pro Ala	
1 5 10 15	
gct tgg gtg gag agg cta ttc ggc tac gac tgg gca caa cag aca atc	96
Ala Trp Val Glu Arg Leu Phe Gly Tyr Asp Trp Ala Gln Gln Thr Ile	
20 25 30	
ggc tgc tct gac gcc gcc gtg ttc cgg ctg tca gcg cag ggg cgc ccg	144
Gly Cys Ser Asp Ala Ala Val Phe Arg Leu Ser Ala Gln Gly Arg Pro	
35 40 45	
gtt ctt ttt gtc aag acc gac ctg tcc ggt gcc ctg aac gaa ctg cag	192
Val Leu Phe Val Lys Thr Asp Leu Ser Gly Ala Leu Asn Glu Leu Gln	
50 55 60	
gac gag gca gcg cgg cta tcg tgg ctg gcc acg acg ggc gtt cct tgc	240
Asp Glu Ala Ala Arg Leu Ser Trp Leu Ala Thr Thr Gly Val Pro Cys	
65 70 75 80	

ES 2 384 391 T3

Met Gly Ser Ala Ile Glu Gln Asp Gly Leu His Ala Gly Ser Pro Ala
 1 5 10 15

Ala Trp Val Glu Arg Leu Phe Gly Tyr Asp Trp Ala Gln Gln Thr Ile
 20 25 30

Gly Cys Ser Asp Ala Ala Val Phe Arg Leu Ser Ala Gln Gly Arg Pro

35

40

45

Val Leu Phe Val Lys Thr Asp Leu Ser Gly Ala Leu Asn Glu Leu Gln
 50 55 60

Asp Glu Ala Ala Arg Leu Ser Trp Leu Ala Thr Thr Gly Val Pro Cys
 65 70 75 80

Ala Ala Val Leu Asp Val Val Thr Glu Ala Gly Arg Asp Trp Leu Leu
 85 90 95

Leu Gly Glu Val Pro Gly Gln Asp Leu Leu Ser Ser His Leu Ala Pro
 100 105 110

Ala Glu Lys Val Ser Ile Leu Ala Asp Ala Leu Arg Arg Leu His Thr
 115 120 125

Leu Asp Pro Ala Thr Cys Pro Phe Asp His Gln Ala Lys His Arg Ile
 130 135 140

Glu Arg Ala Arg Thr Arg Leu Glu Ala Gly Leu Val Asp Gln Asp Asp
 145 150 155 160

Leu Asp Glu Glu His Gln Gly Leu Ala Pro Ala Glu Leu Phe Ala Arg
 165 170 175

Leu Lys Ala Arg Leu Pro Asp Gly Asp Asp Leu Val Val Thr His Gly
 180 185 190

Asp Ala Cys Leu Pro Asn Ile Leu Val Glu Asn Gly Arg Phe Ser Gly
 195 200 205

Phe Ile Asp Cys Gly Arg Leu Gly Val Ala Asp Arg Tyr Gln Asp Ile
 210 215 220

Ala Leu Ala Thr Arg Asp Ile Ala Glu Glu Leu Gly Gly Glu Trp Ala
 225 230 235 240

Asp Arg Phe Leu Val Leu Tyr Gly Ile Ala Ala Pro Asp Ser Gln Arg
 245 250 255

Ile Ala Phe Tyr Arg Leu Leu Asp Glu Phe Phe
 260 265

<210> 136

<211> 40

5 <212> ADN

<213> Artificial

<220>

<223> cebador NEO-F-HindIII

ES 2 384 391 T3

<400> 136
gatcaagctt ttggatcggc cattgaaaca agacggattg 40
<210> 137
<211> 36
5 <212> ADN
<213> Artificial
<220>
<223> cebador NEO EcoRI 800R
<400> 137
10 aagcttgaat tctcagaaga actcgtaag aaggcg 36
<210> 138
<211> 564
<212> ADN
<213> Artificial
15 <220>
<223> gen de dhfr modificado que carece de secuencias ATG internas
<220>
<221> CDS
<222> (1)..(564)
20 <400> 138

ES 2 384 391 T3

atg gtt cga cca ttg aac tgc atc gtc gcc gtg tcc caa aat ctg ggg 48
 Met Val Arg Pro Leu Asn Cys Ile Val Ala Val Ser Gln Asn Leu Gly
 1 5 10 15

att ggc aag aac gga gac cta ccc tgg cct ccg ctc agg aac gag ttc 96
 Ile Gly Lys Asn Gly Asp Leu Pro Trp Pro Pro Leu Arg Asn Glu Phe
 20 25 30

aag tac ttc caa aga ctg acc aca acc tct tca gtg gaa ggt aaa cag 144
 Lys Tyr Phe Gln Arg Leu Thr Thr Thr Ser Ser Val Glu Gly Lys Gln
 35 40 45

aat ctg gtg att ctg ggt agg aaa acc tgg ttc tcc att cct gag aag 192
 Asn Leu Val Ile Leu Gly Arg Lys Thr Trp Phe Ser Ile Pro Glu Lys
 50 55 60

aat cga cct tta aag gac aga att aat ata gtt ctc agt aga gaa ctc 240
 Asn Arg Pro Leu Lys Asp Arg Ile Asn Ile Val Leu Ser Arg Glu Leu
 65 70 75 80

aaa gaa cca cca cga gga gct cat ttt ctt gcc aaa agt ttg gac gac 288
 Lys Glu Pro Pro Arg Gly Ala His Phe Leu Ala Lys Ser Leu Asp Asp
 85 90 95

gcc tta aga ctt att gaa caa ccg gaa ttg gca agt aaa gta gac ctg 336
 Ala Leu Arg Leu Ile Glu Gln Pro Glu Leu Ala Ser Lys Val Asp Leu
 100 105 110

gtt tgg ata gtc gga ggc agt tct gtt tac cag gaa gcc ctg aat caa 384
 Val Trp Ile Val Gly Gly Ser Ser Val Tyr Gln Glu Ala Leu Asn Gln
 115 120 125

cca ggc cac ctc aga ctc ttt gtg aca agg att ctg cag gaa ttt gaa 432
 Pro Gly His Leu Arg Leu Phe Val Thr Arg Ile Leu Gln Glu Phe Glu
 130 135 140

agt gac acg ttt ttc cca gaa att gat ttg ggg aaa tat aaa ctt ctc 480
 Ser Asp Thr Phe Phe Pro Glu Ile Asp Leu Gly Lys Tyr Lys Leu Leu
 145 150 155 160

cca gaa tac cca ggc gtc ctc tct gag gtc cag gag gaa aaa ggc atc 528
 Pro Glu Tyr Pro Gly Val Leu Ser Glu Val Gln Glu Glu Lys Gly Ile
 165 170 175

aag tat aag ttt gaa gtc tac gag aag aaa gac taa 564
 Lys Tyr Lys Phe Glu Val Tyr Glu Lys Lys Asp
 180 185

<210> 139

<211> 187

<212> PRT

5 <213> Artificial

<220>

<223> gen de dhfr modificado que carece de secuencias ATG internas

<400> 139

ES 2 384 391 T3

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

<210> 140

<211> 36

<212> ADN

5 <213> Artificial

<220>

<223> cebador DHFR-F-HindIII

<400> 140

gatcaagctt ttgtcgacc attgaactgc atcgtc

36

10 <210> 141

<211> 36

<212> ADN

<213> Artificial

<220>

15 <223> cebador DHFR-EcoRI-600-R

<400> 141

aagcttgaat tcttagtctt tcttctcgta gacttc

36

ES 2 384 391 T3

<210> 142

<211> 154

<212> ADN

<213> Artificial

5 <220>

<223> secuencia de poliadenilación sintética combinada y señal de pausa del gen de alfa2 globina humano

<220>

<221> secuencia de poliadenilación sintética

<222> (1)..(49)

10 <220>

<221> sitio de clonación

<222> (50)..(62)

<220>

<221> señal de pausa del gen de alfa2 globina humano

15 <222> (63)..(154)

<400> 142

```
aataaaatat ctttatcttc attacatctg tgtggtgggt ttttgtgtga atcgatagta      60
ctaacatacg ctctccatca aaacaaaacg aaacaaaaca aactagcaaa ataggctgtc      120
cccagtgcaa gtgcaggtgc cagaacattt ctct                                     154
```

REIVINDICACIONES

- 5 1. Una molécula de ADN que comprende una unidad de transcripción multicistrónica que codifica i) un polipéptido marcador seleccionable capaz de ser seleccionado en una célula hospedante eucariota, y ii) un polipéptido de interés, teniendo el polipéptido de interés una secuencia de iniciación de la traducción separada de la del polipéptido marcador seleccionable,
- en la que
- la secuencia codificante para el polipéptido de interés está en dirección 3' de la secuencia codificante para el marcador seleccionable en dicha unidad de transcripción multicistrónica, y
- 10 la secuencia que codifica el polipéptido marcador seleccionable no tiene ninguna secuencia ATG en la hebra codificante tras el codón de inicio del polipéptido marcador seleccionable hasta el codón de inicio del polipéptido de interés, y en la que la secuencia de inicio de la traducción en la hebra codificante para el polipéptido marcador seleccionable se elige del grupo que consiste en:
- 15 a) una secuencia de inicio de la traducción que comprende una mutación en los nucleótidos en las posiciones -3 a -1 y/o +4 de la secuencia de consenso RCCATGG (en la que R es A o G, y en la que el A del codón de inicio ATG es nt +1);
- b) un codón de inicio GTG;
- c) un codón de inicio TTG;
- d) un codón de inicio CTG;
- e) un codón de inicio ATT; y
- 20 f) un codón de inicio ACG.
2. Una molécula de ADN según la reivindicación 1, en la que la secuencia de inicio de la traducción resultante en la hebra codificante para el polipéptido marcador seleccionable comprende un codón de inicio GTG o un codón de inicio TTG.
- 25 3. Una molécula de ADN según la reivindicación 1 ó 2, en la que cualquier secuencia ATG cuando está presente en el marco y que codifica metionina en la secuencia que codifica el polipéptido marcador seleccionable de tipo salvaje está mutada para codificar valina, treonina, isoleucina o leucina.
4. Una molécula de ADN según cualquiera de las reivindicaciones anteriores, en la que el polipéptido marcador seleccionable proporciona resistencia frente a los efectos letales o inhibidores del crecimiento de un agente de selección, por ejemplo un antibiótico.
- 30 5. Una molécula de ADN según la reivindicación 4, en la que dicho agente de selección se elige del grupo que consiste en: zeocina, puromicina, blastidina, higromicina, neomicina, metotrexato, metionina sulfoximina y kanamicina.
6. Una molécula de ADN según la reivindicación 5, en la que el agente de selección es zeocina.
- 35 7. Una molécula de ADN según la reivindicación 5, en la que el polipéptido marcador seleccionable está codificado por el gen de resistencia a neomicina.
8. Una molécula de ADN según una cualquiera de las reivindicaciones anteriores, en la que el polipéptido marcador seleccionable comprende además una mutación que reduce la actividad del polipéptido marcador seleccionable comparada con su contraparte de tipo salvaje, por ejemplo en la que el polipéptido marcador seleccionable es un polipéptido de resistencia a zeocina en el que la prolina en la posición 9 está cambiada por un aminoácido diferente.
- 40 9. Una molécula de ADN según una cualquiera de las reivindicaciones anteriores, en la que la secuencia codificante del polipéptido de interés comprende una secuencia de inicio de la traducción óptima.
10. Un casete de expresión que comprende una molécula de ADN según una cualquiera de las reivindicaciones anteriores, comprendiendo además dicho casete de expresión un promotor en dirección 5' de dicha unidad de expresión multicistrónica, y secuencias de terminación de la transcripción en dirección 3' de la unidad de expresión multicistrónica.
- 45 11. Un casete de expresión según la reivindicación 10, que comprende además al menos un elemento de control de cromatina, escogido del grupo que consiste en una región de unión de matriz o armazón (MAR/SAR), una secuencia aislante, un elemento de apertura de cromatina universal (UCOE), y una secuencia antirrepresora (STAR).

12. Un casete de expresión según la reivindicación 11, en el que dicho elemento de control de cromatina es una secuencia antirrepresora elegida del grupo que consiste en:

- a) una cualquiera de SEQ. ID. NO. 1 a SEQ. ID. NO. 66;
- b) fragmentos de una cualquiera de SEQ. ID. NO. 1 a SEQ. ID. NO. 66, en el que dichos fragmentos tienen actividad antirrepresora;
- c) secuencias que son al menos 70% idénticas en secuencia nucleotídica a a) o b), en el que dichas secuencias tienen actividad antirrepresora; y
- d) el complemento a una cualquiera de a) a c).

13. Un casete de expresión según la reivindicación 12, en el que dicho casete de expresión comprende uno de:

- a) SEQ. ID. NO. 66,
- b) un fragmento de a) que tiene actividad antirrepresora;
- c) una secuencia que es al menos 70% idéntica en secuencia nucleotídica a a) o b), en el que dicha secuencia tiene actividad antirrepresora,

en el que a), b) o c) está situado en dirección 5' del promotor que dirige la transcripción del gen multicistrónico.

14. Un casete de expresión según una cualquiera de las reivindicaciones 10-13, en el que dicho gen multicistrónico está flanqueado en ambos lados por al menos una secuencia antirrepresora.

15. Un casete de expresión según la reivindicación 14, que comprende: 5' - secuencia antirrepresora A - secuencia antirrepresora B - promotor - gen multicistrónico que codifica la proteína marcadora seleccionable funcional y, en dirección 3' de ella, el polipéptido de interés - secuencia de terminación de la transcripción - secuencia antirrepresora C - 3',

en el que las secuencias antirrepresoras A y C pueden ser iguales o diferentes y se eligen del grupo que consiste en:

- a) una cualquiera de SEQ. ID. NO. 1 a SEQ. ED. NO. 65;
- b) fragmentos de una cualquiera de las secuencias de a), teniendo dichos fragmentos actividad antirrepresora;
- c) secuencias que son al menos 70% idénticas en secuencia nucleotídica a a) o b), en el que dichas secuencias tienen actividad antirrepresora,

y en el que la secuencia antirrepresora B se elige del grupo que consiste en

- i) SEQ. ID. NO. 66,
- ii) un fragmento de i) que tiene actividad antirrepresora; y
- iii) una secuencia que es al menos 70% idéntica en secuencia nucleotídica a i) o ii), en el que dicha secuencia tiene actividad antirrepresora.

16. Un casete de expresión según la reivindicación 15, en el que las secuencias antirrepresoras A y C se eligen de:

- a) un cualquiera de SEQ. ID. NO. 7, SEQ. ID. NO. 9, SEQ. ID. NO. 17, SEQ. ID. NO. 27, SEQ. ID. NO. 29, SEQ. ID. NO. 43, SEQ. ID. NO. 44, SEQ. ID. NO. 45, SEQ. ID. NO. 47, o SEQ. ID. NO. 61;
- b) fragmentos de una de las secuencias de a), teniendo dichos fragmentos actividad antirrepresora; y
- c) secuencias que son al menos 70% idénticas en secuencia nucleotídica a a) o b), en el que dichas secuencias tienen actividad antirrepresora.

17. Un casete de expresión según una cualquiera de las reivindicaciones 10-16, en el que el polipéptido de interés es una parte de una proteína multimérica.

18. Un casete de expresión según la reivindicación 17, en el que el polipéptido de interés es una cadena ligera de inmunoglobulina o una cadena pesada de inmunoglobulina.

19. Un casete de expresión según una cualquiera de las reivindicaciones 10-18, que comprende además:

a) una secuencia de pausa de la transcripción (TRAP) en dirección 5' de dicho promotor y que está en una dirección 5' a 3'; o

b) una secuencia TRAP en dirección 3' de dicha secuencia de terminación de la transcripción y que está en una orientación 3' a 5'; o

5 c) tanto a) como b);

en el que una secuencia TRAP es una secuencia que, cuando se coloca en una unidad de transcripción, da como resultado un nivel reducido de transcripción en el ácido nucleico presente en el lado 3' de la TRAP, cuando se compara con el nivel de transcripción observado en el ácido nucleico en el lado 5' de la TRAP,

por ejemplo, en el que dicha TRAP es SEQ. ID. NO. 142.

10 20. Una molécula de ADN que comprende una secuencia que codifica un polipéptido marcador seleccionable funcional capaz de ser seleccionado en una célula hospedante eucariota, caracterizada porque dicha molécula de ADN:

i) tiene una secuencia de inicio de la traducción elegida del grupo que consiste en:

15 a) una secuencia de inicio de la traducción que comprende una mutación en los nucleótidos en las posiciones -3 a -1 y/o +4 de la secuencia de consenso RCCATGG (en la que R es A o G, y en la que el A del codón de inicio ATG es nt +1);

b) un codón de inicio GTG;

c) un codón de inicio TTG;

d) un codón de inicio CTG;

20 e) un codón de inicio ATT; y

f) un codón de inicio ACG, seguido de una secuencia codificante marcadora seleccionable de otro modo funcional, y

ii) porque la hebra codificante de la secuencia que define el polipéptido marcador seleccionable en dirección 3' del codón de inicio no óptimo y hasta el codón de parada está desprovista de secuencias ATG.

25 21. Una célula hospedante que comprende una molécula de ADN o un casete de expresión según una cualquiera de las reivindicaciones anteriores.

22. Una célula hospedante según la reivindicación 21, que es una célula hospedante eucariota, preferiblemente una célula hospedante de mamífero.

23. Una célula hospedante según la reivindicación 22, que es una célula de ovario de hámster chino (CHO).

30 24. Un método para generar una célula hospedante que expresa un polipéptido de interés, comprendiendo el método las etapas de:

a) introducir en una pluralidad de células precursoras una molécula de ADN según una cualquiera de las reivindicaciones 1-9 o un casete de expresión según una cualquiera de las reivindicaciones 10-19, y

35 b) cultivar las células generadas en condiciones que seleccionan la expresión del polipéptido marcador seleccionable, y

c) seleccionar al menos una célula hospedante que produce el polipéptido de interés.

25. Un método para producir un polipéptido de interés, que comprende cultivar una célula hospedante, comprendiendo dicha célula hospedante un casete de expresión según una cualquiera de las reivindicaciones 10-19, y expresar el polipéptido de interés a partir del casete de expresión.

40 26. Un método según la reivindicación 25, que comprende además aislar el polipéptido de interés.

Fig 1

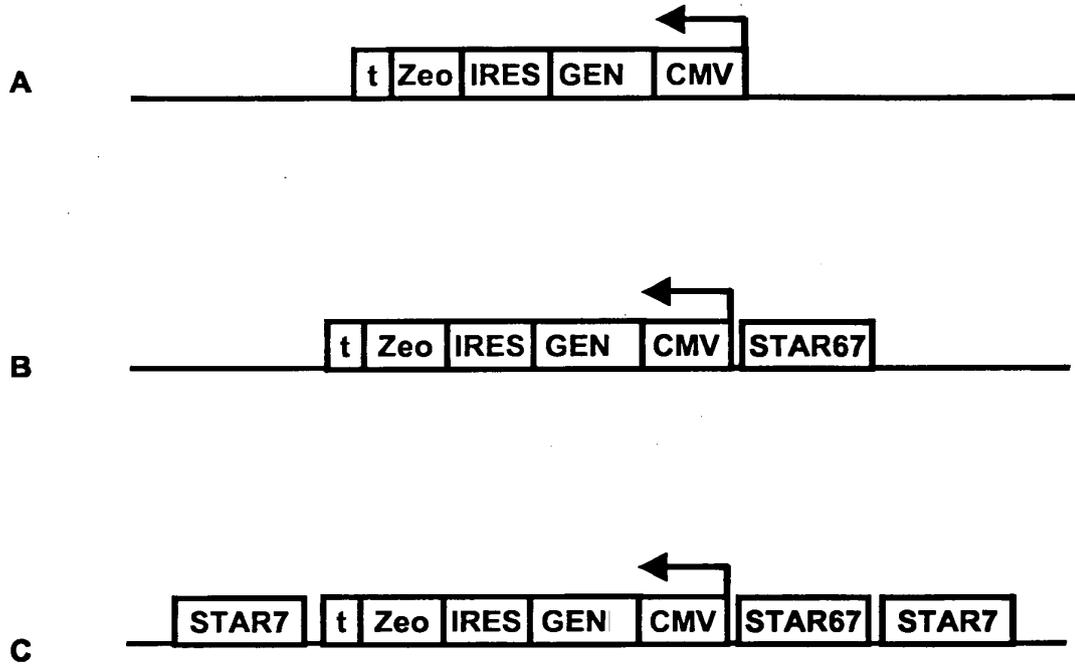


Fig 2

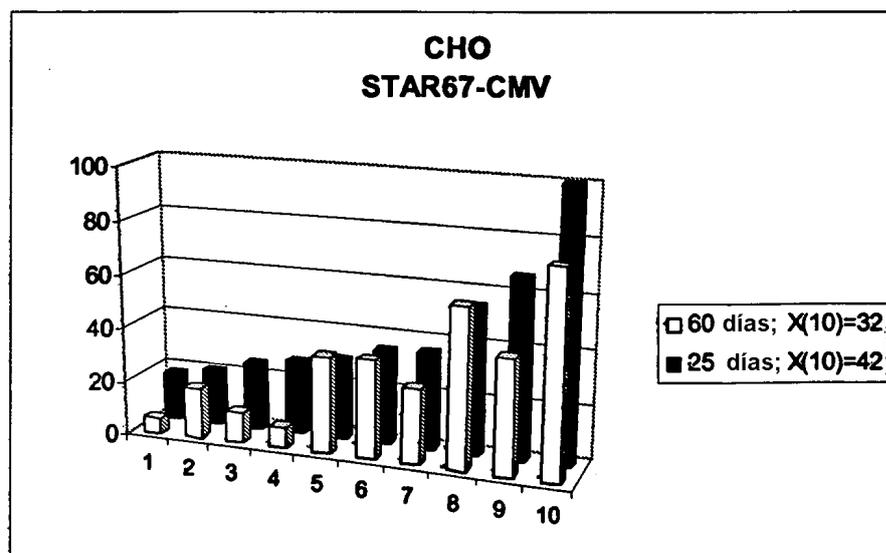
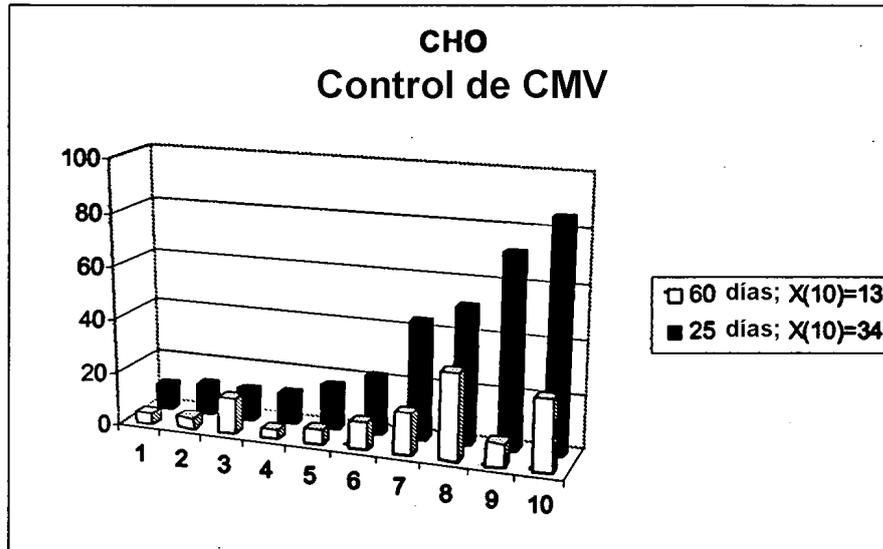
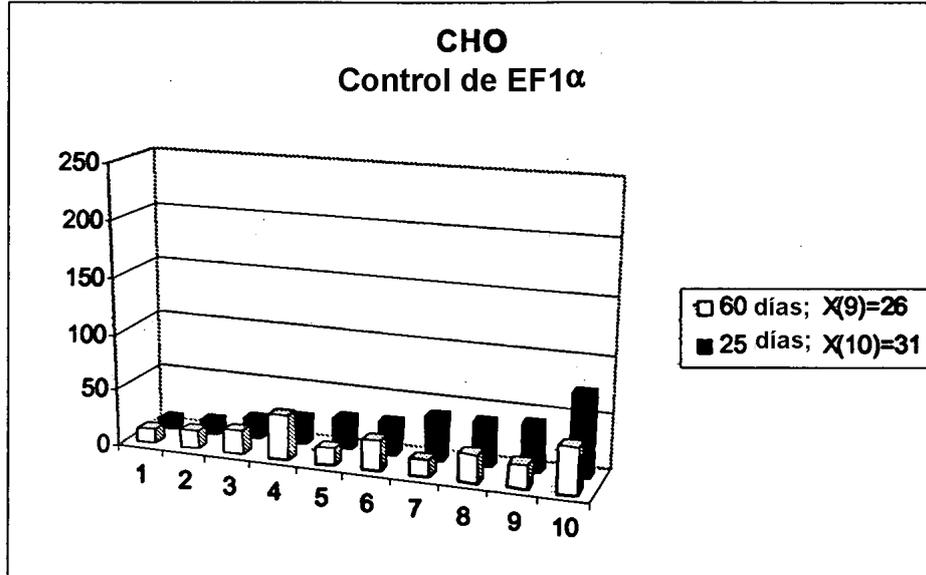


Fig 3

A



B

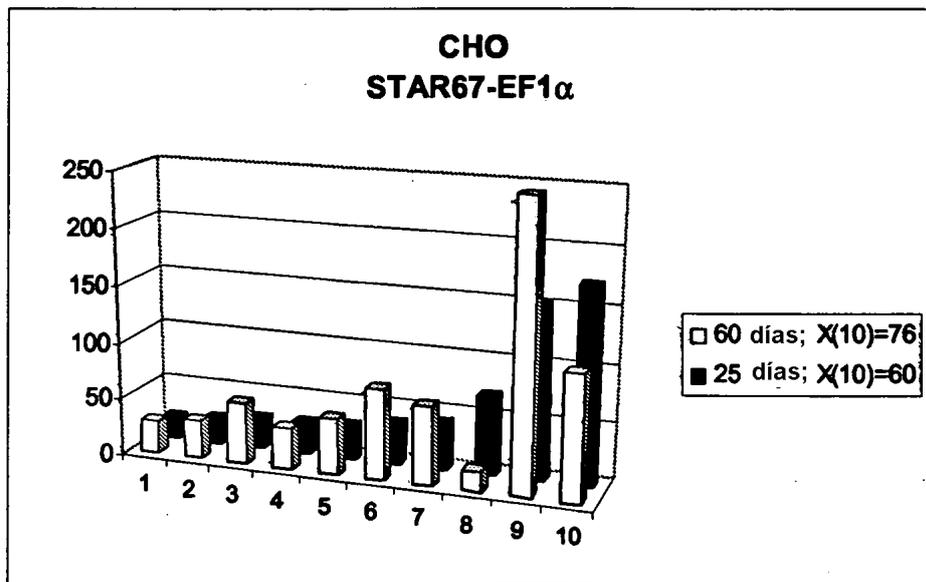


Fig 4

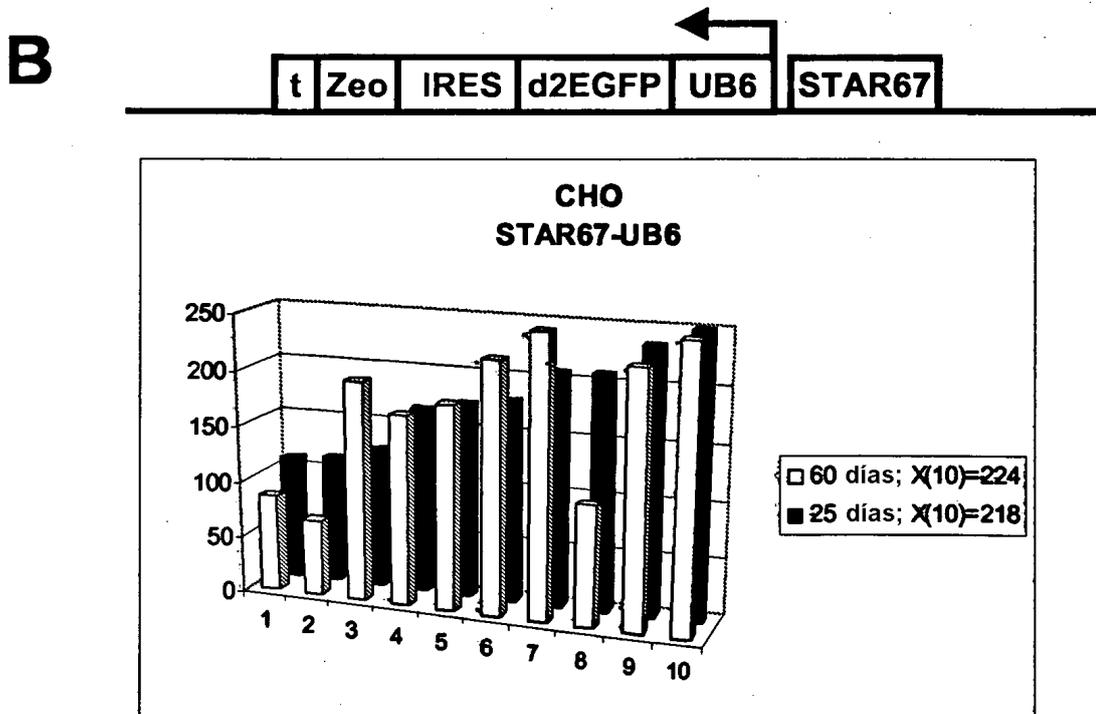
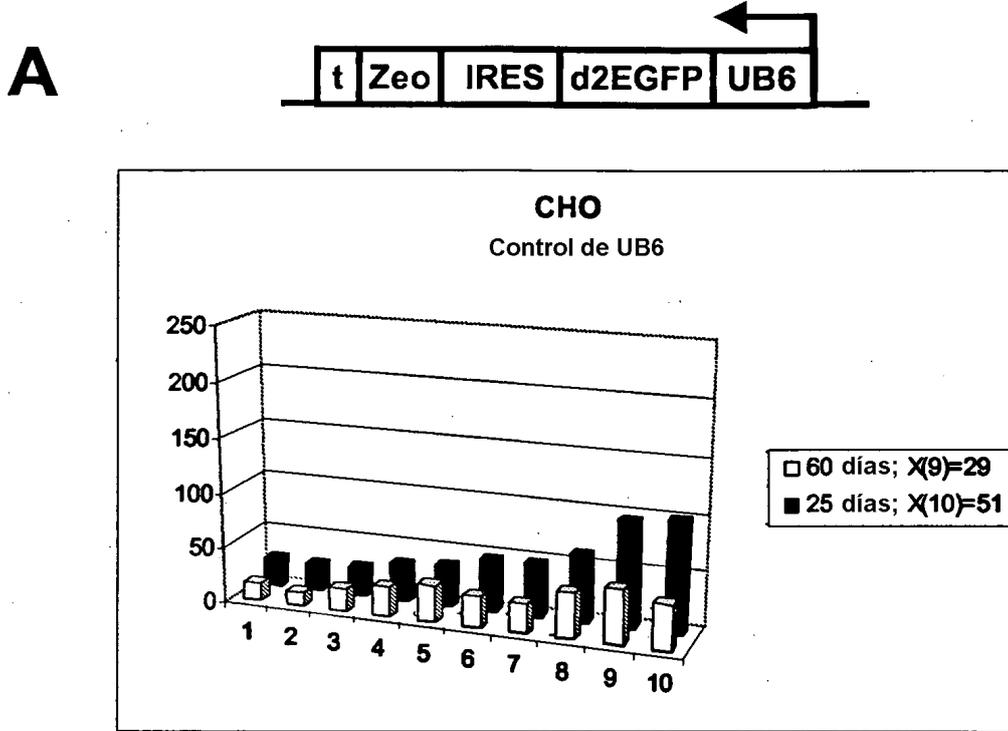
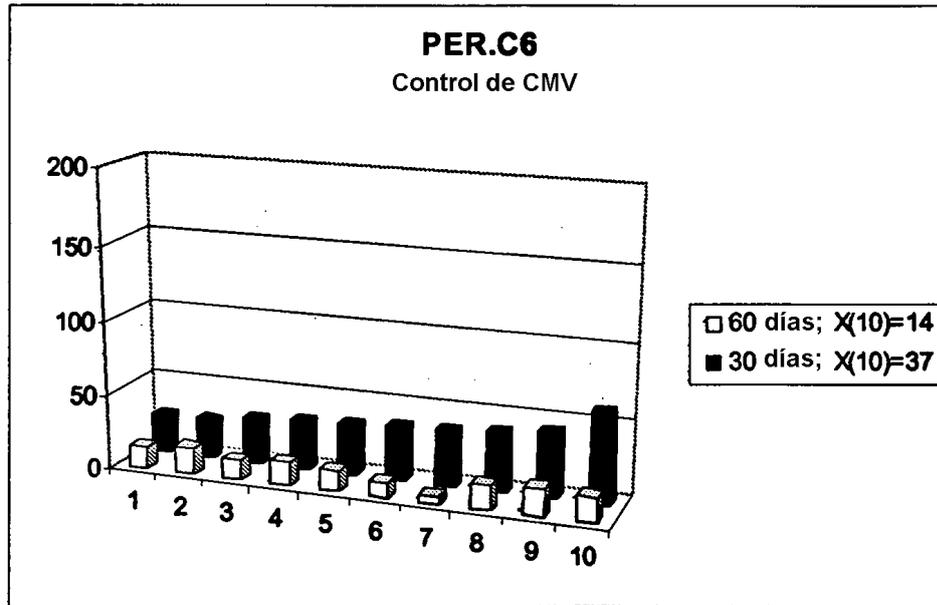


Fig 5

A



B

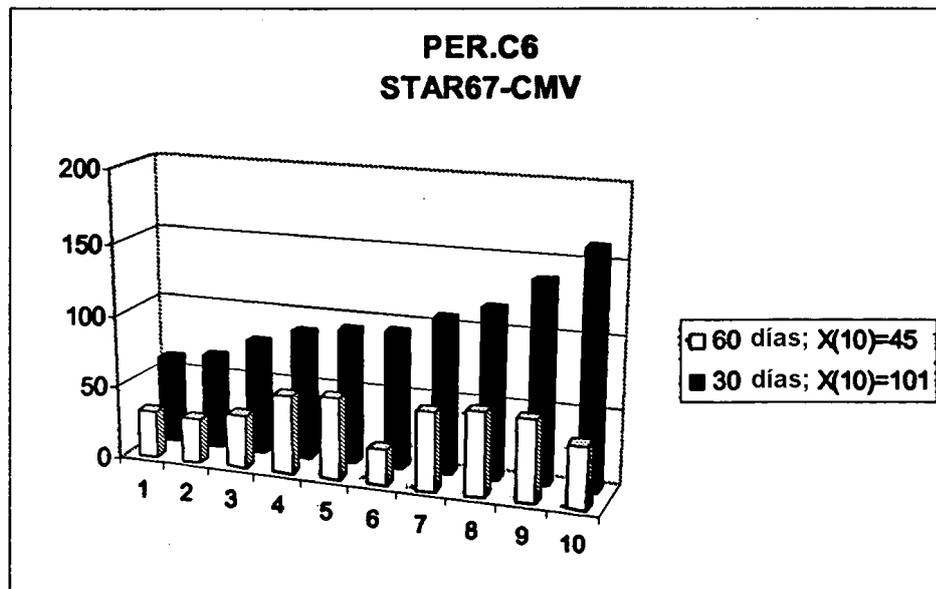
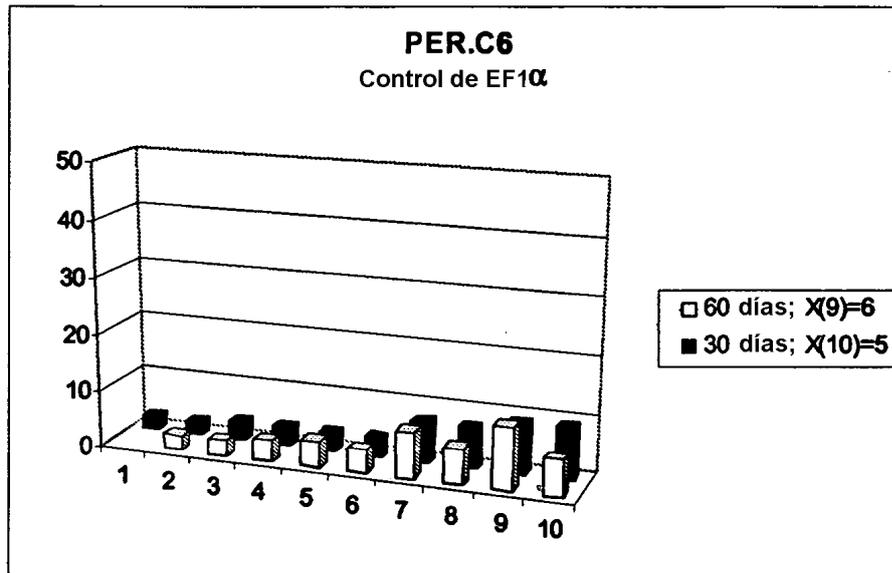


Fig 6

A



B

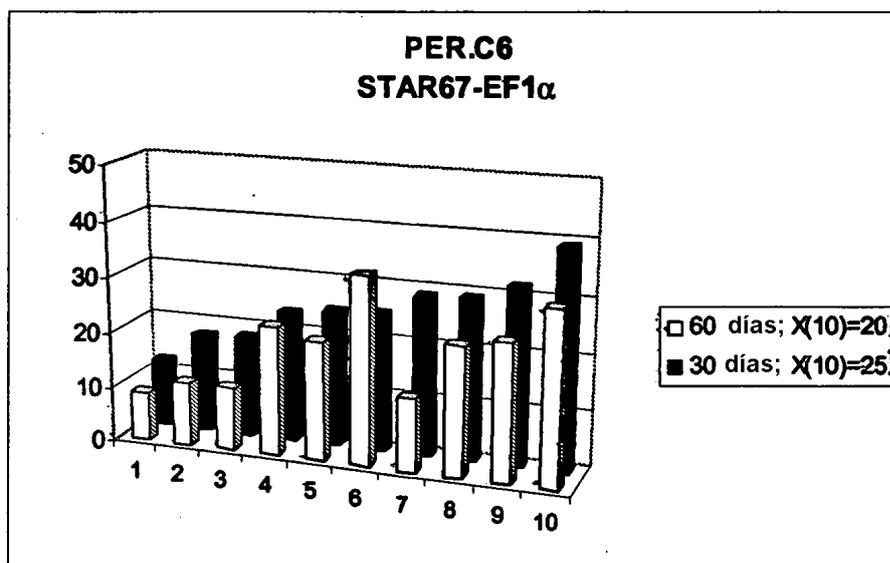
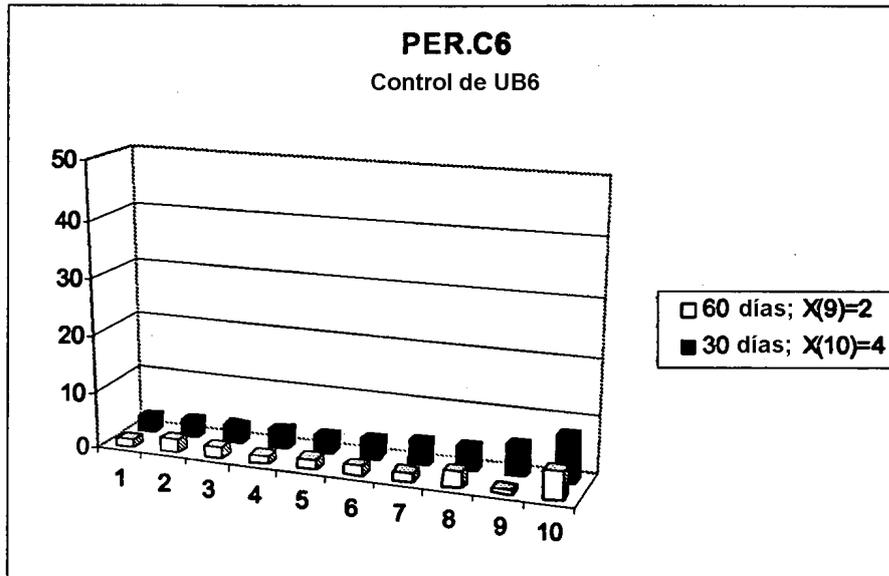


Fig 7

A



B

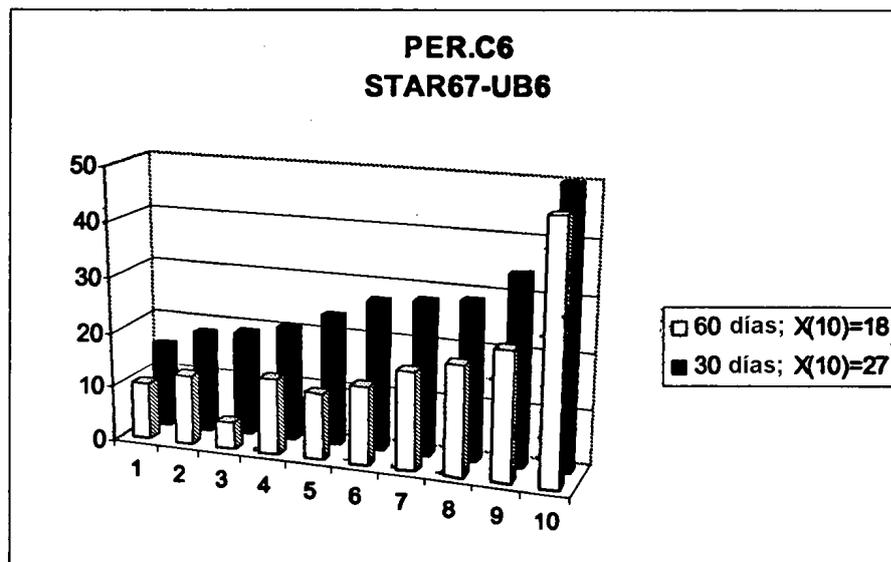


Fig 8

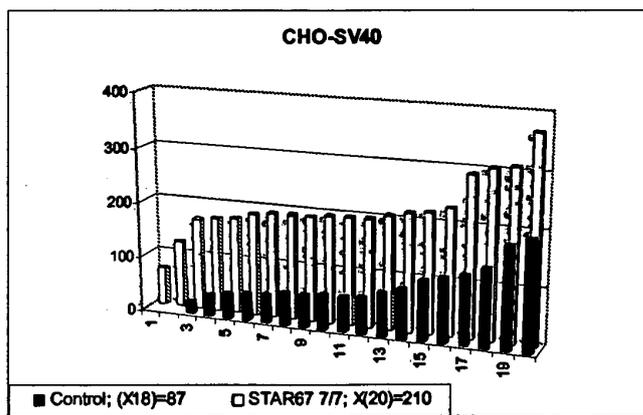
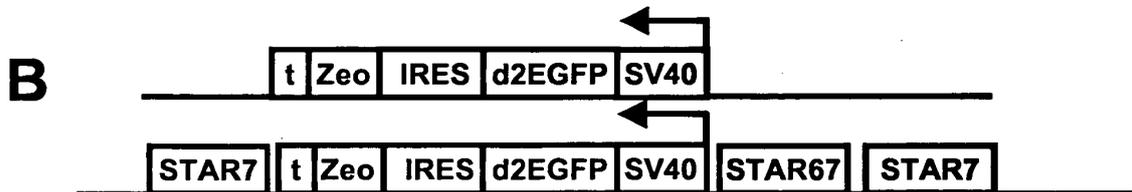
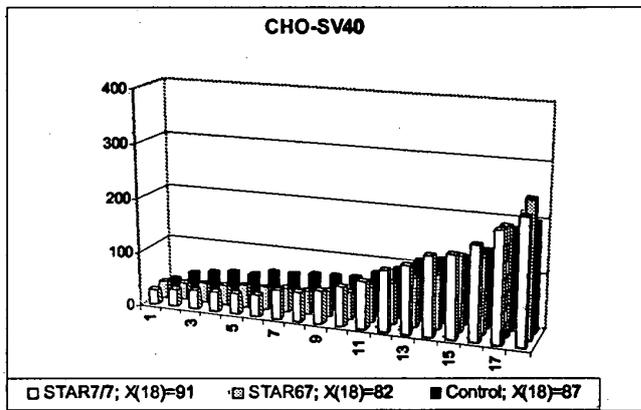
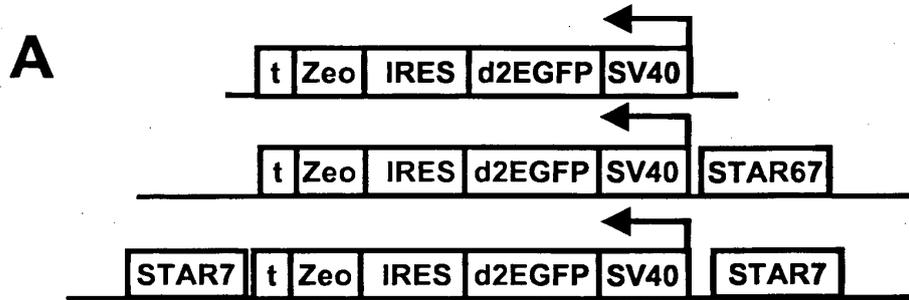


Fig 9

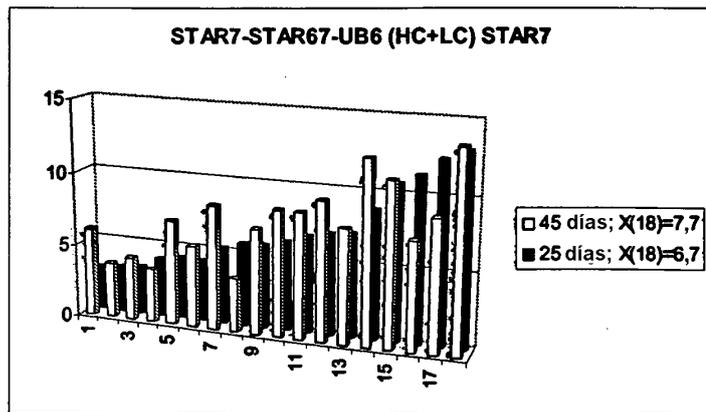
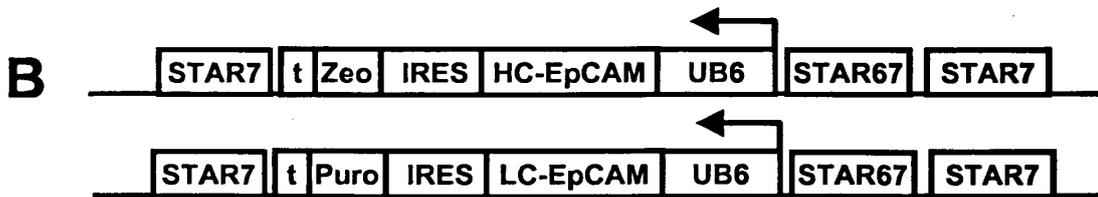
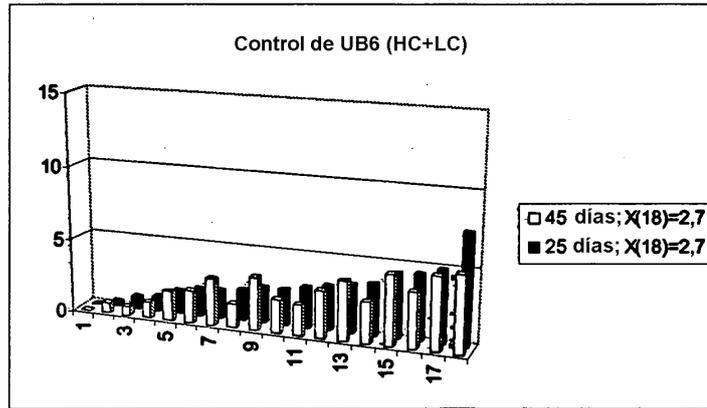
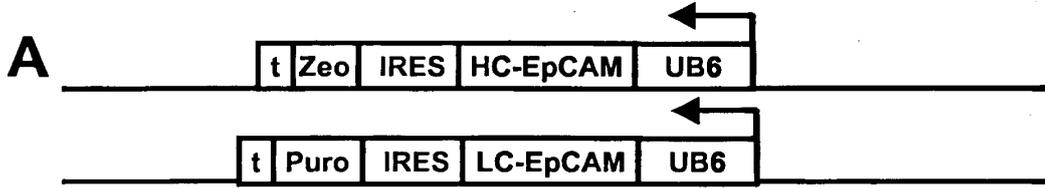


Fig 10

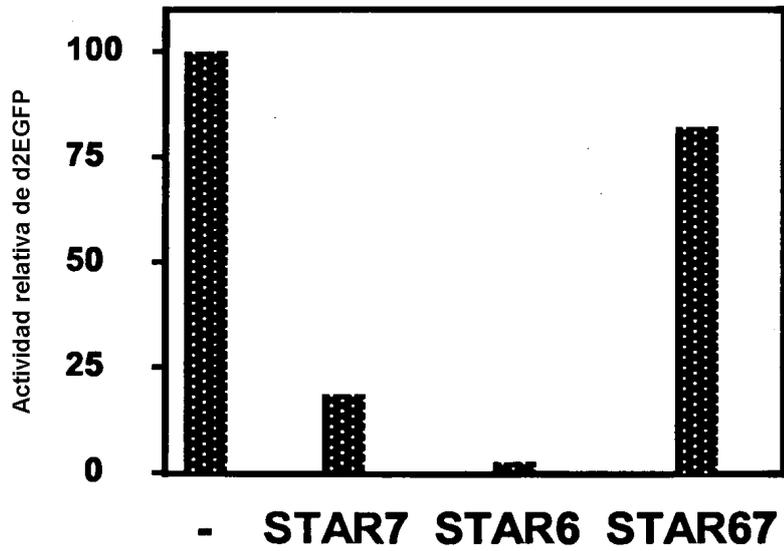
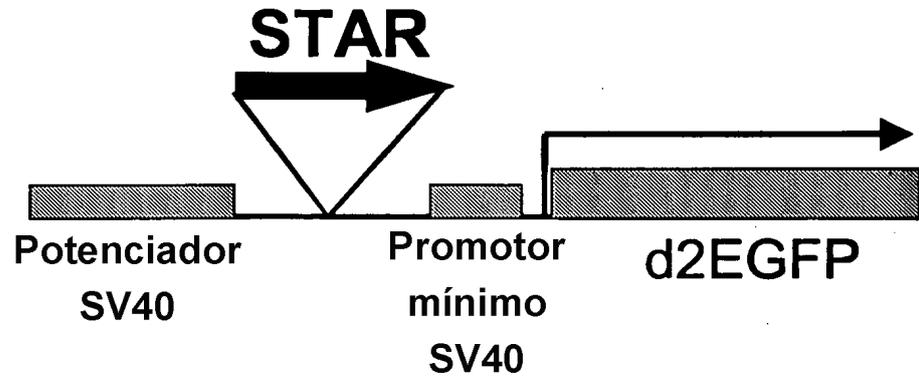
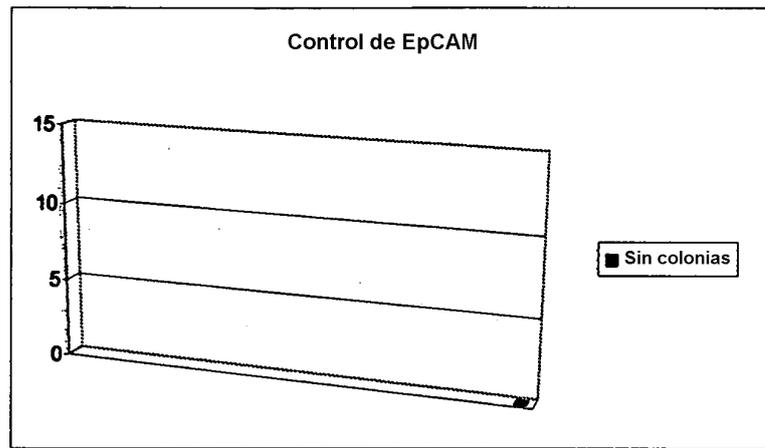
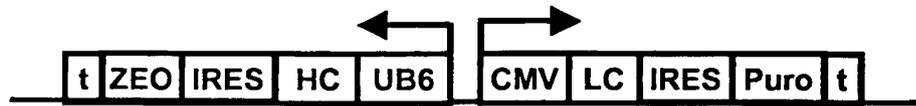


Fig 11

A



B

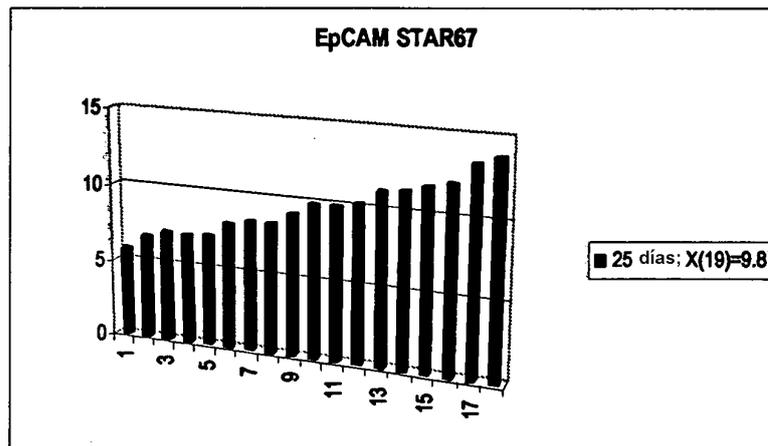


FIG 12

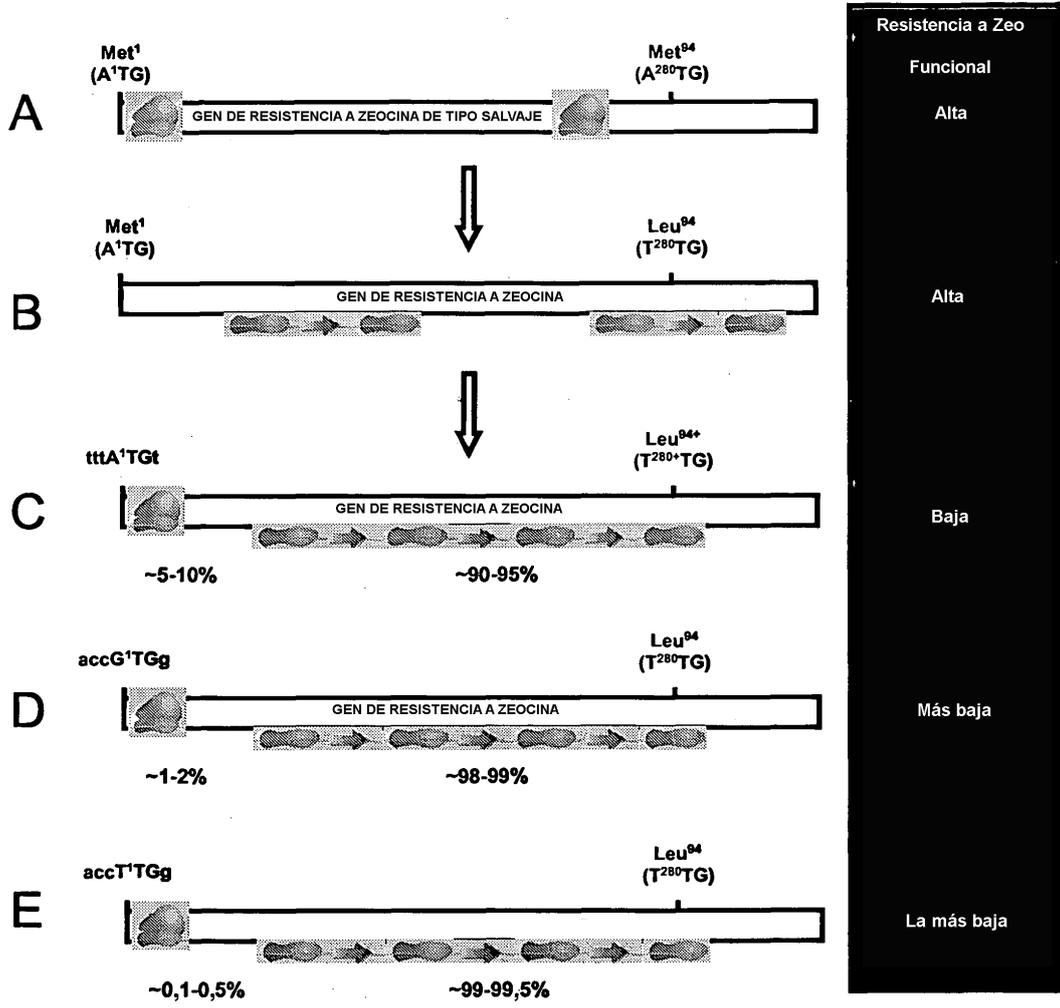


FIG 13

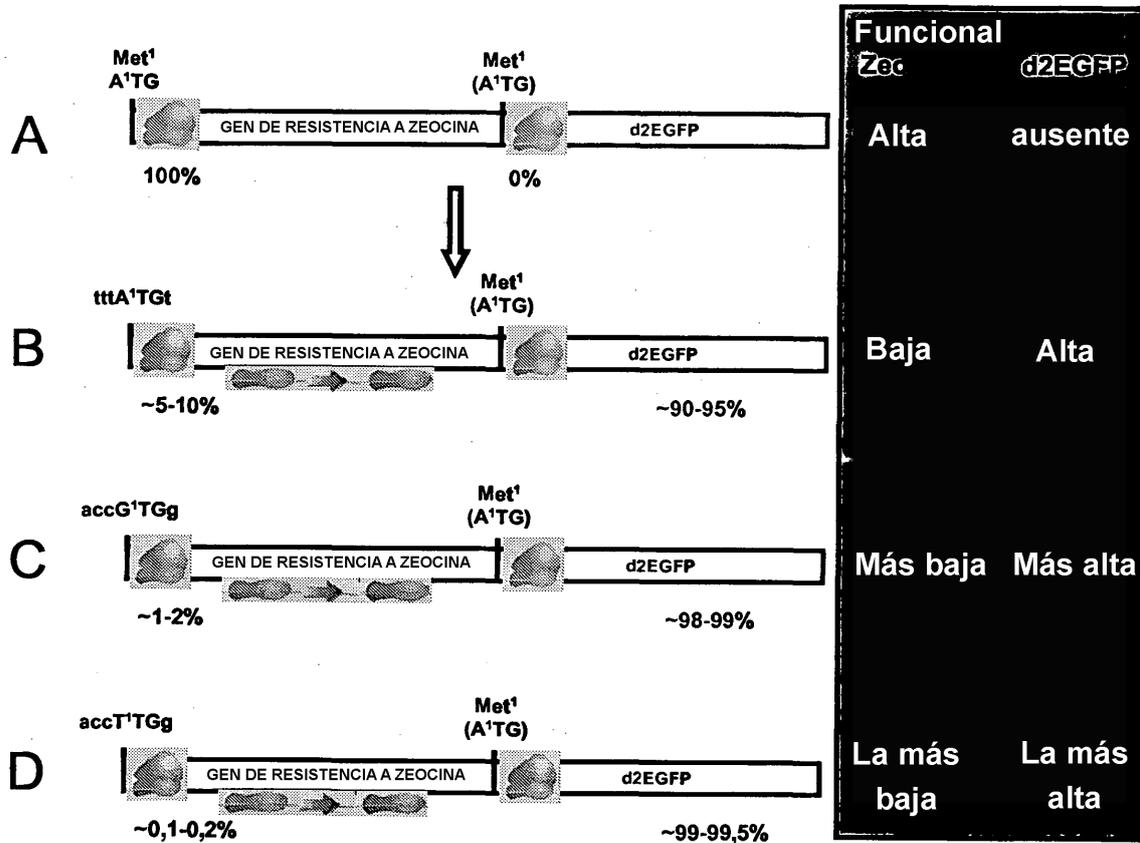
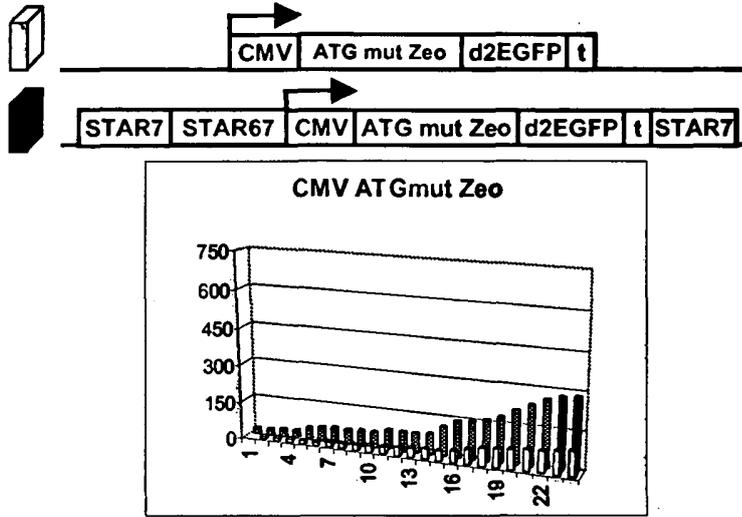
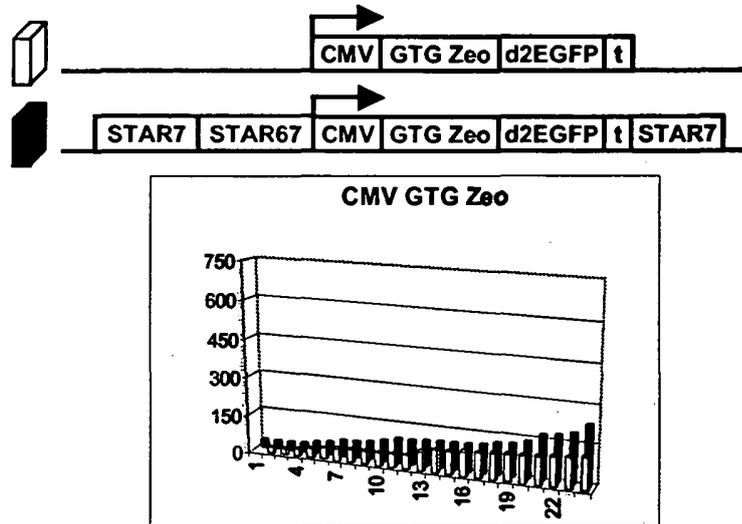


FIG 14

A



B



C

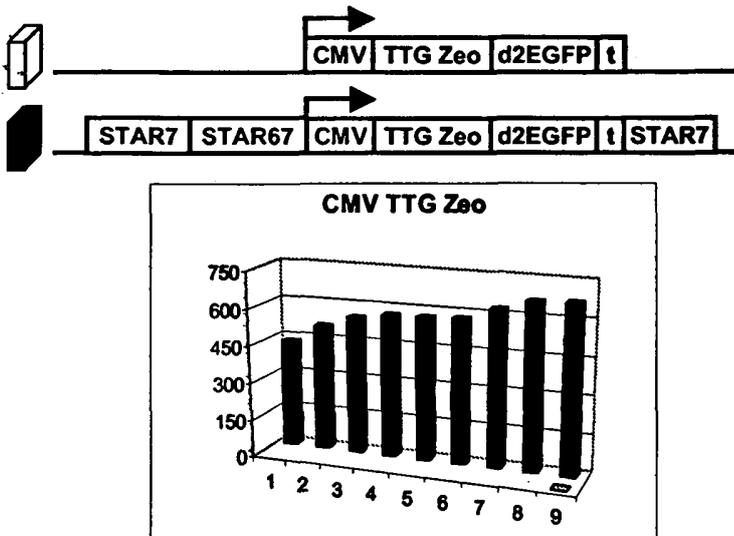


FIG 15

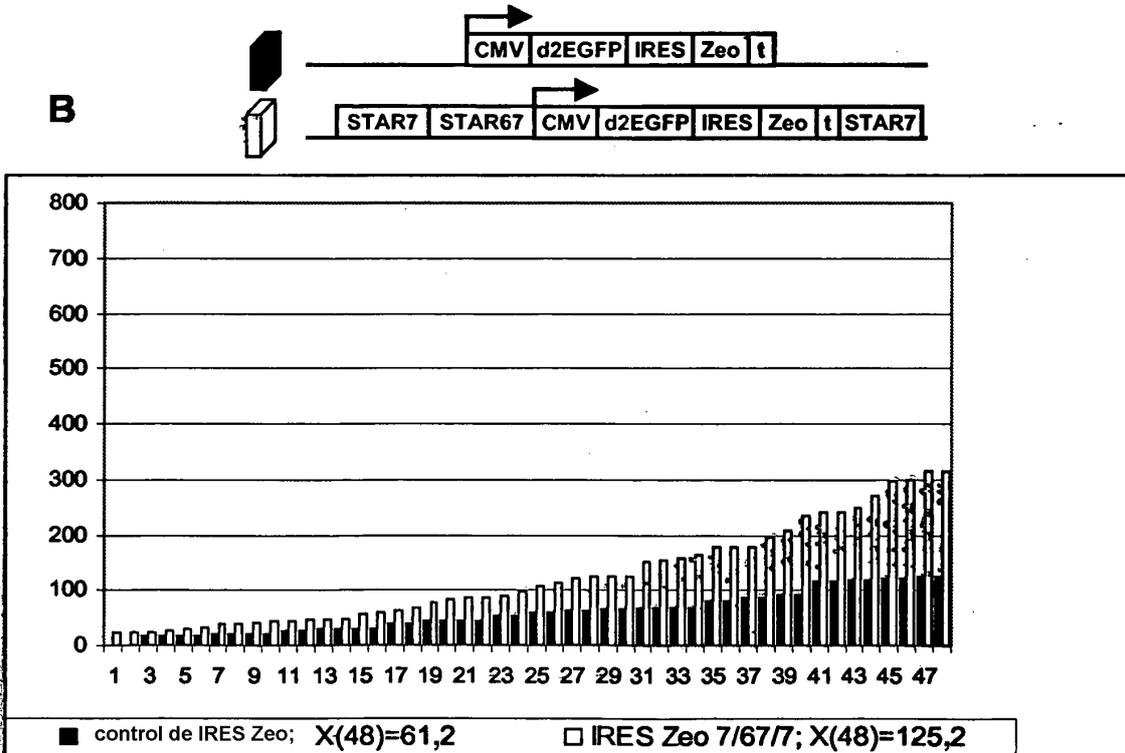
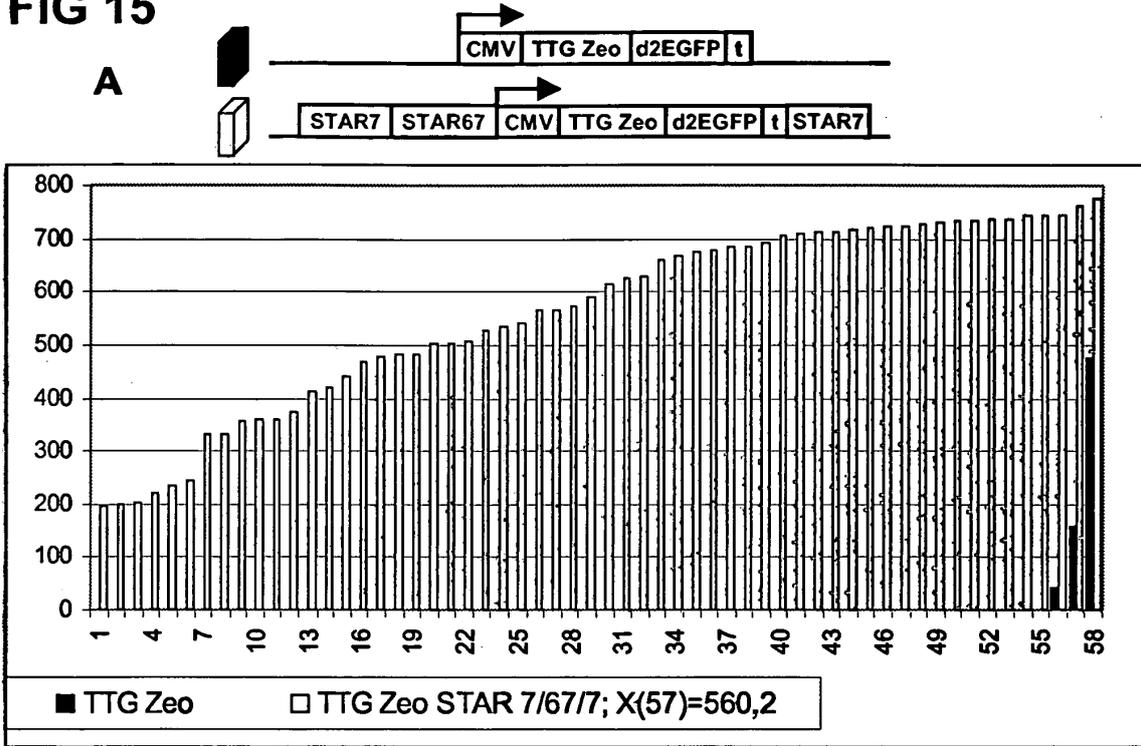


FIG 16

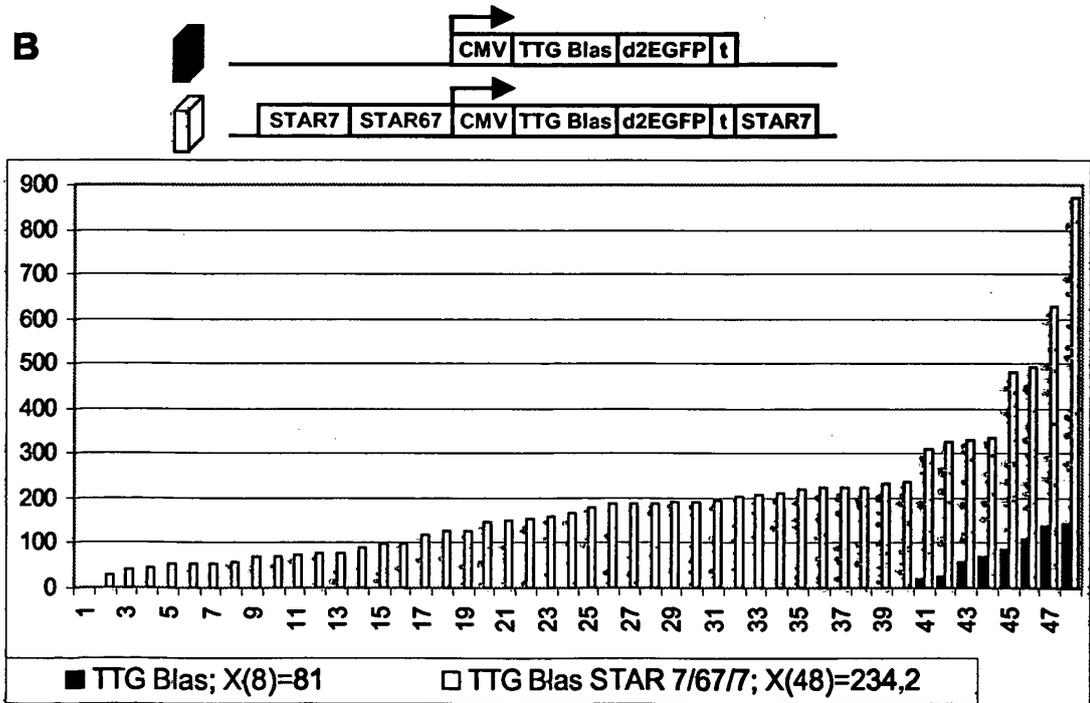
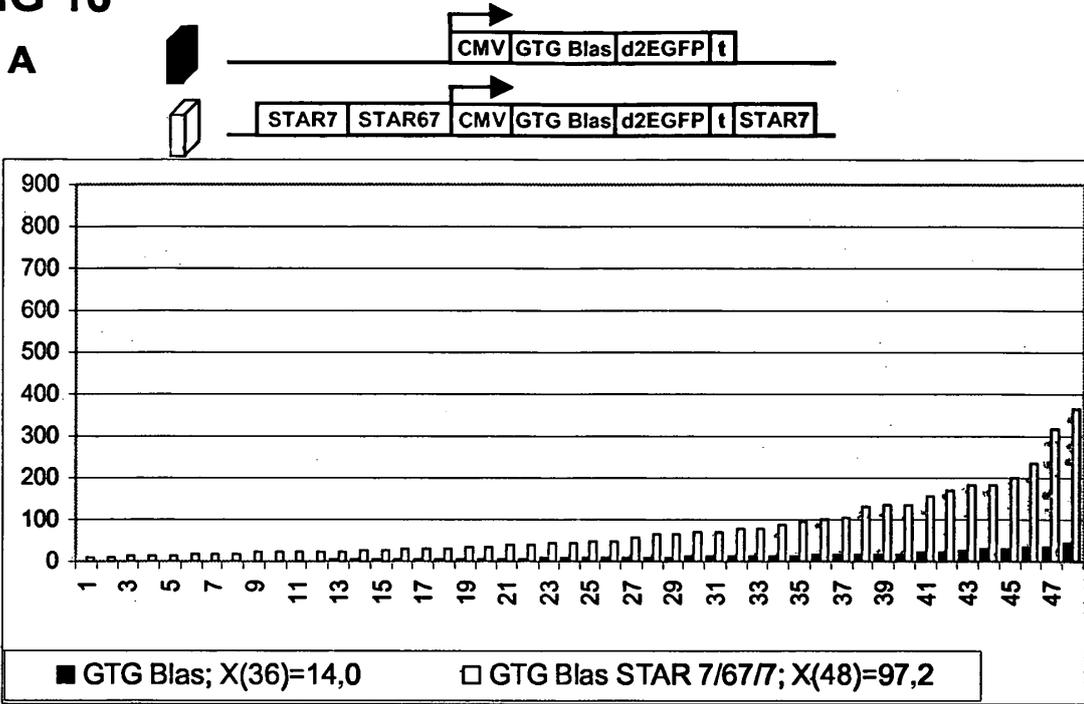


FIG 17

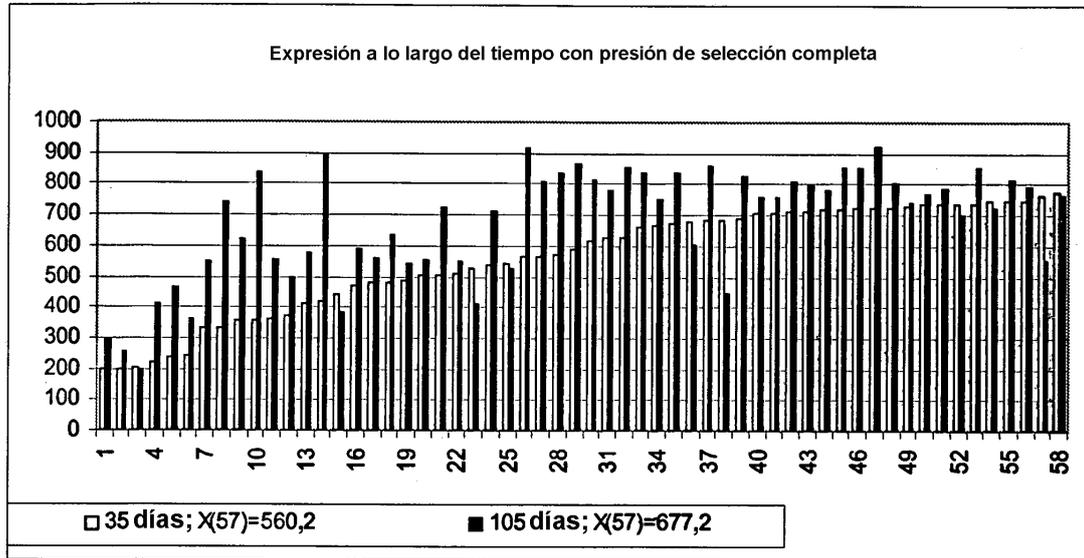


FIG 18

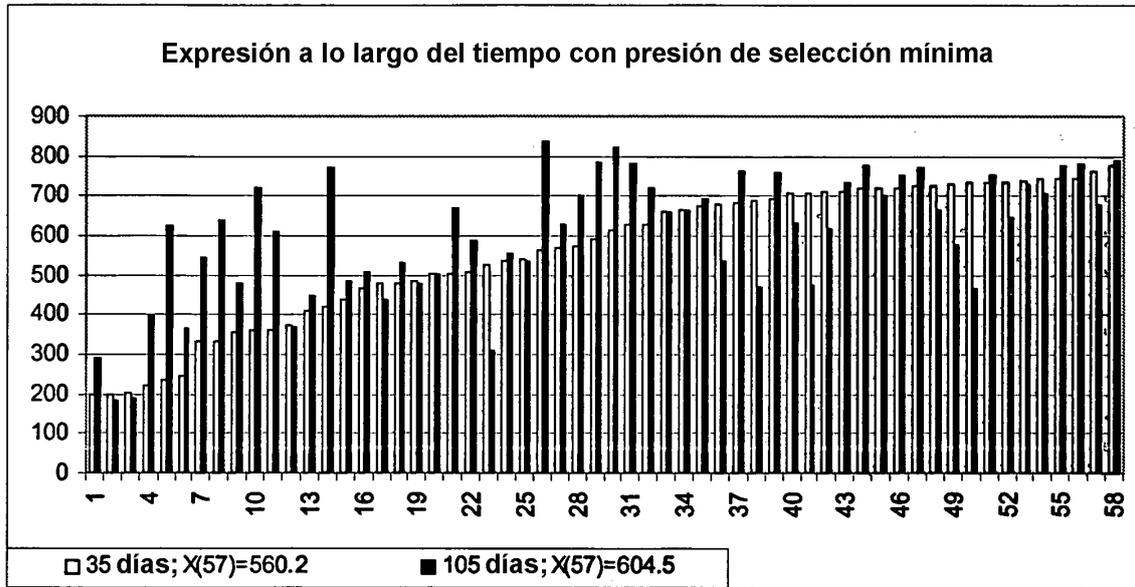


FIG 19

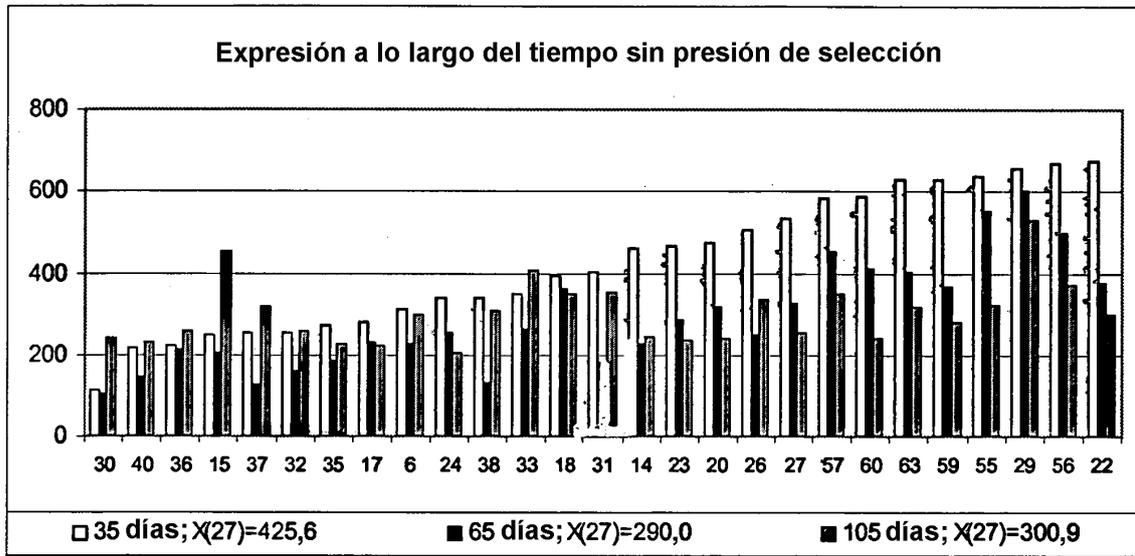


FIG 20

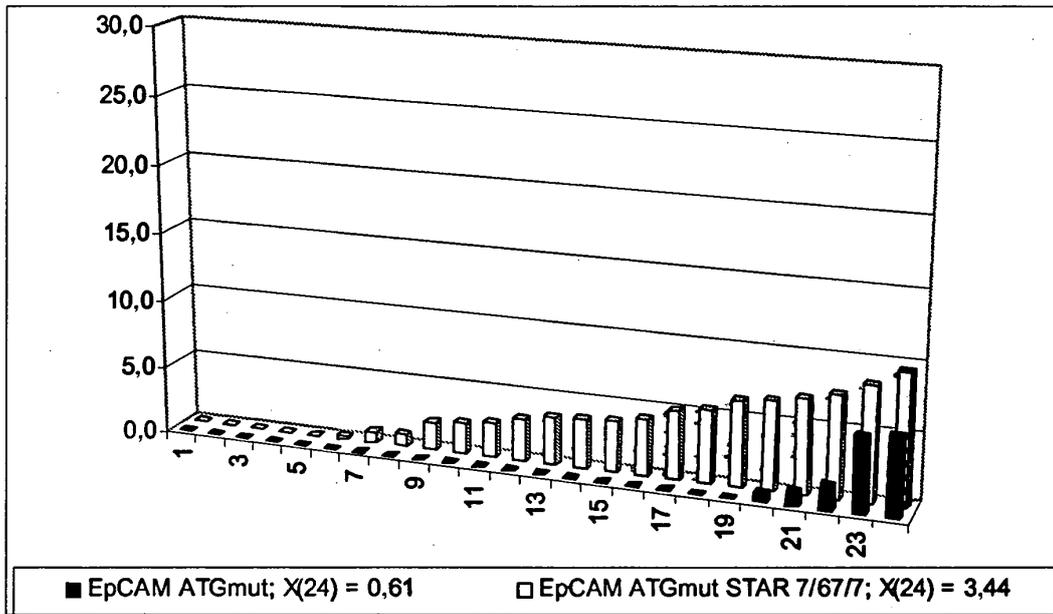
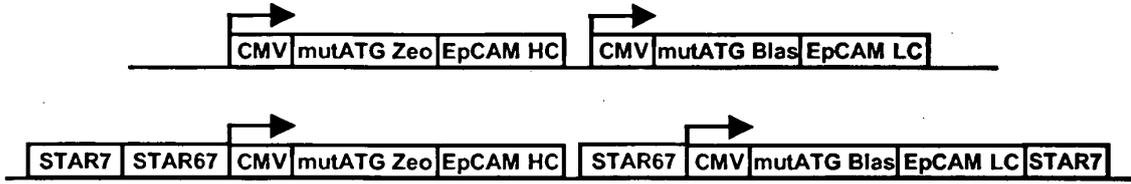


FIG 21

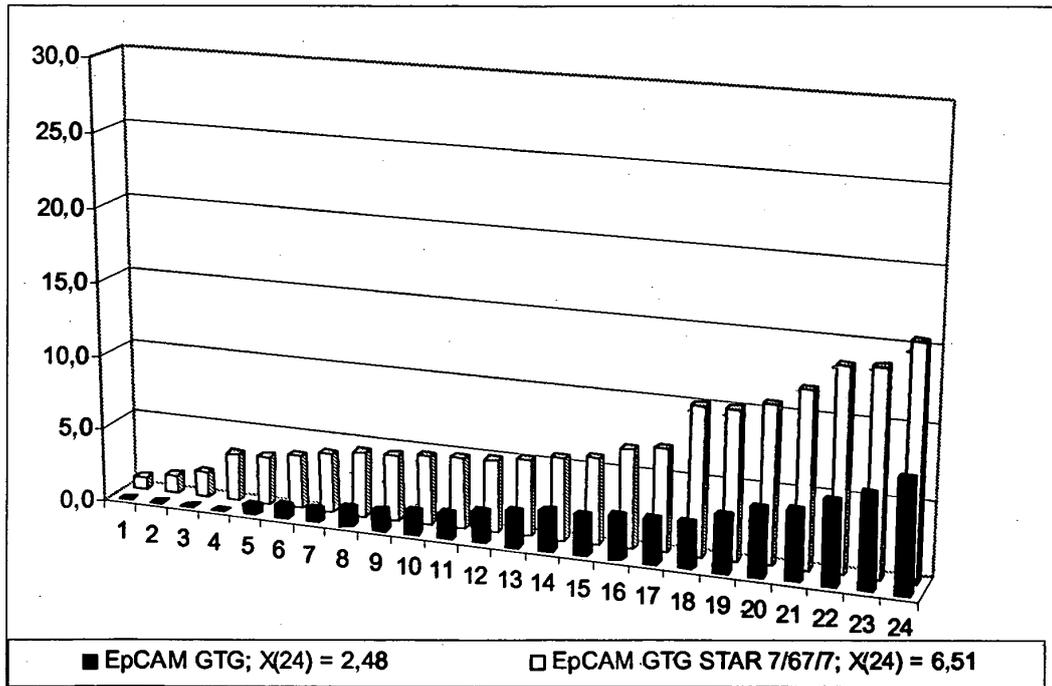
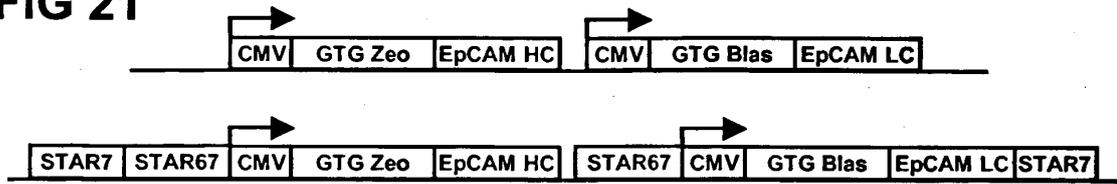


FIG 22

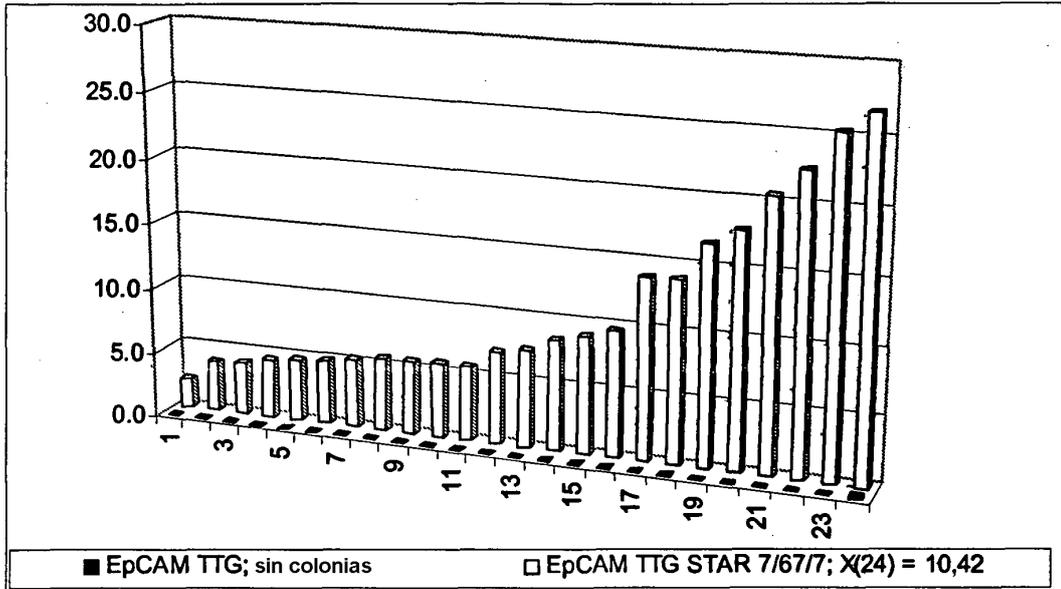
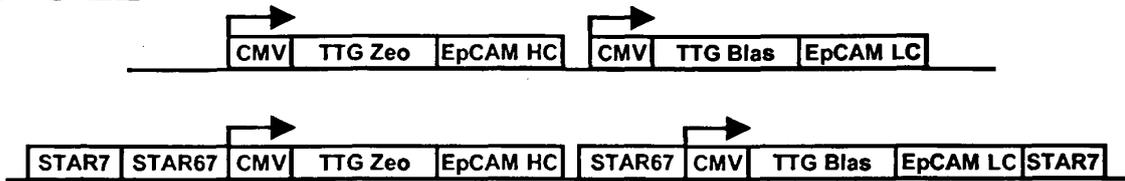


FIG 23

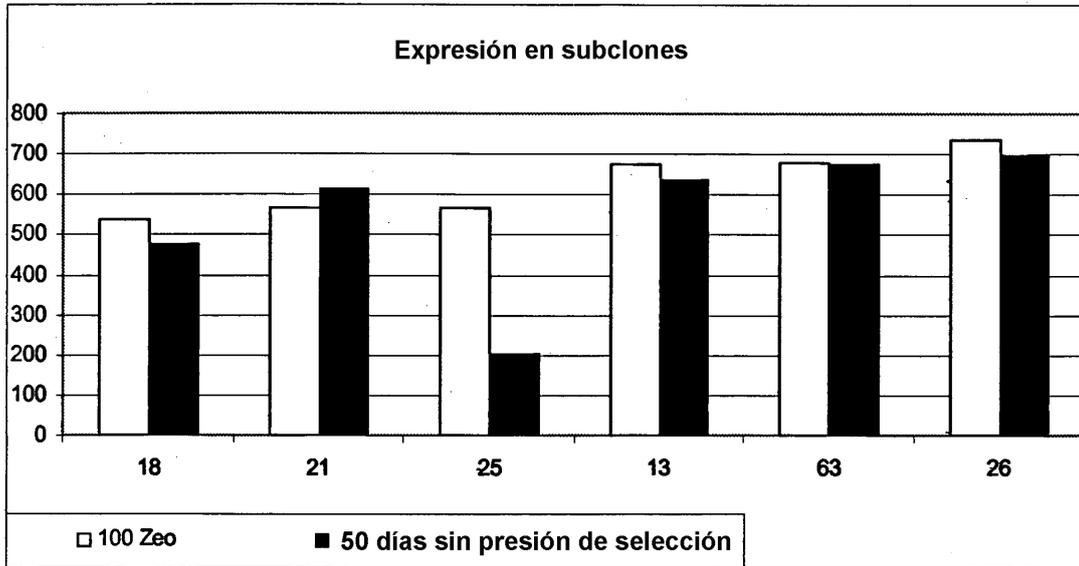


FIG 24

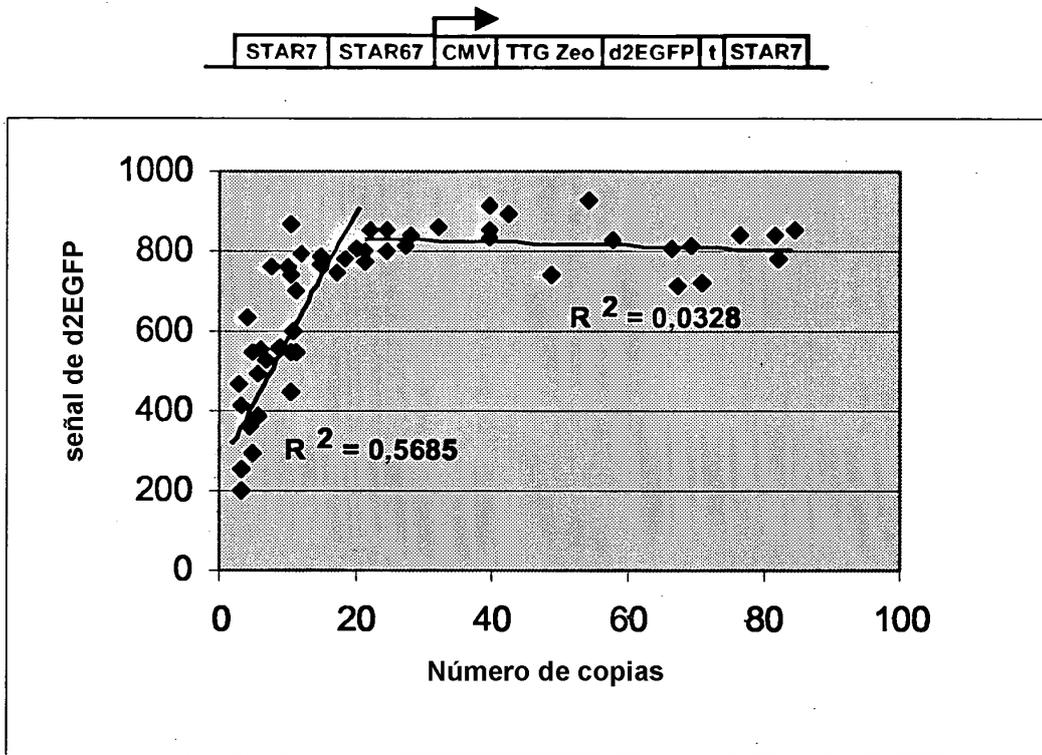


FIG 25

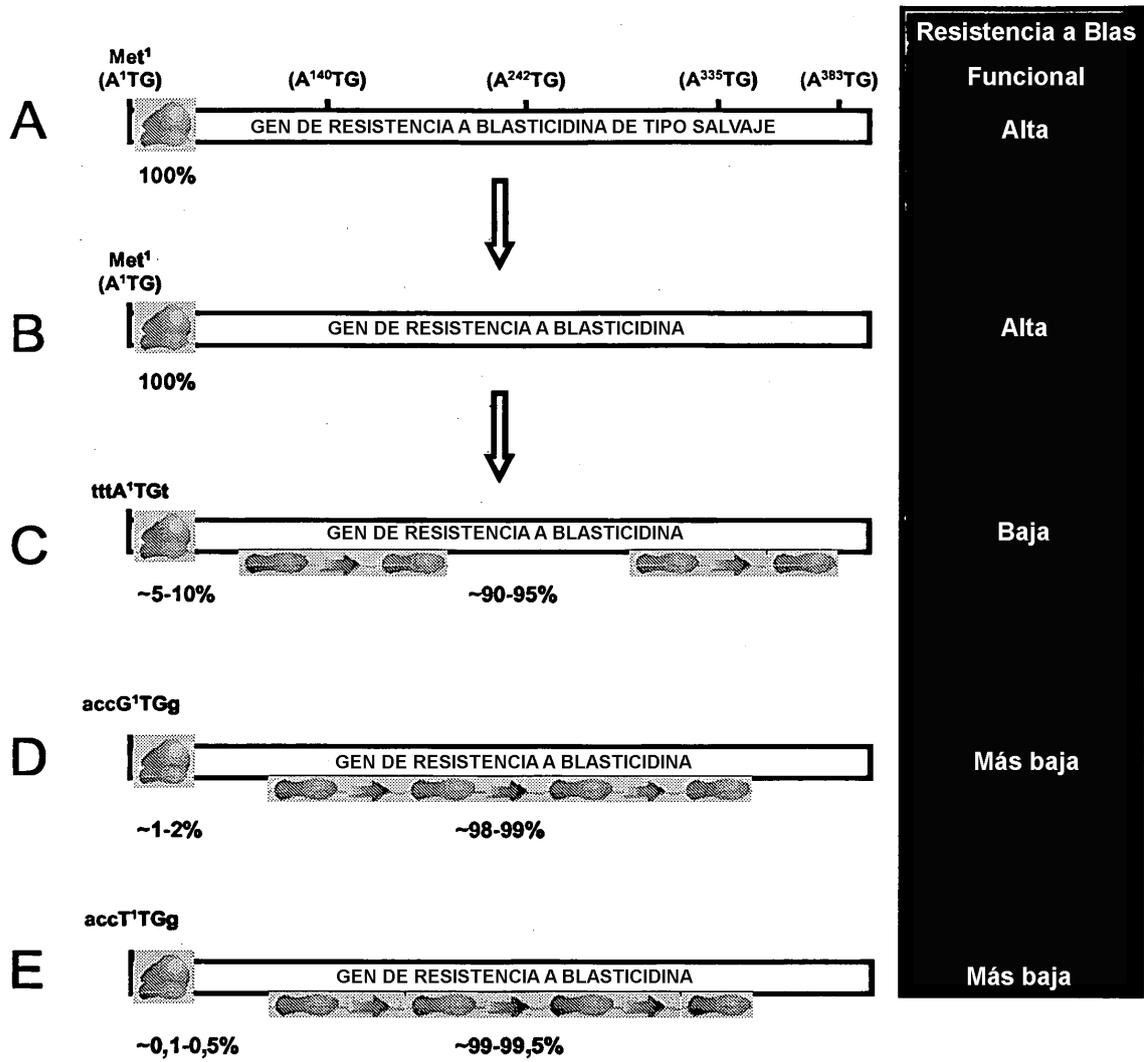


Fig. 26

ATGGCCAAGTTGACCAGTGCCGTTCCGGTGCTCACCGCGCGCGA
CGTCGCCGGAGCGGTTCGAGTTCTGGACCGACCGGCTCGGGTTCT
CCCGGGACTTCGTGGAGGACGACTTCGCCGGTGTGGTCCGGGAC
GACGTGACCCTGTTCATCAGCGCGGTCCAGGACCAGGTGGTGCC
GGACAACACCCTGGCCTGGGTGTGGGTGCGCGGCCTGGACGAGC
TGTACGCCGAGTGGTCGGAGGTCGTGTCCACGAACTTCCGGGAC
GCCTCCGGGCCGGCC**ATG**ACCGAGATCGGCGAGCAGCCGTGGGG
GCGGGAGTTCGCCCTGCGCGACCCGGCCGGCAACTGCGTGCACT
TCGTGGCCGAGGAGCAGGACTGA

Fig. 27

**ATGGCCAAGCCTTTGTCTCAAGAAGAATCCACCCTCATTGAAAGAGCAAC
GGCTACAATCAACAGCATCCCCATCTCTGAAGACTACAGCGTCGCCAGCG
CAGCTCTCTCTAGCGACGGCCGCATCTTCACTGGTGTCAATGTATATCAT
TTTACTGGGGGACCTTGTGCAGAACTCGTGGTGCTGGGCACTGCTGCTGC
TGCGGCAGCTGGCAACCTGACTTGTATCGTCGCGATCGGAAATGAGAACA
GGGGCATCTTGAGCCCCTGCGGACGGTGCCGACAGGTGCTTCTCGATCTG
CATCCTGGGATCAAAGCCATAGTGAAGGACAGTGATGGACAGCCGACGGC
AGTTGGGATTCGTGAATTGCTGCCCTCTGGTTATGTGTGGGAGGGCTAA**

Fig. 28

ATGACCGAGTACAAGCCCACGGTGCGCCTCGCCACCCGCGACGACGTCCCC
AGGGCCGTACGCACCCTCGCCGCCGCGTTCGCCGACTACCCCGCCACGCGC
CACACCGTCGATCCGGACCGCCACATCGAGCGGGTCACCGAGCTGCAAGAA
CTCTTCCTCACGCGCGTCGGGCTCGACATCGGCAAGGTGTGGGTGCGGGAC
GACGGCGCCGCGGTGGCGGTCTGGACCAGCCGGAGAGCGTCGAAGCGGGG
GCGGTGTTGCCCGAGATCGGCCCGCGC**ATGGCCGAGTTGAGCGGTTC**CCGG
CTGGCCGCGCAGCAACAG**ATGGA**AGGCCTCCTGGCGCCGCACCGGCCCAAG
GAGCCCGCGTGGTTCCTGGCCACCGTCGGCGTCTCGCCCGACCACCAGGGC
AAGGGTCTGGGCAGCGCCGTCGTGCTCCCCGGAGTGGAGGGCGGCCGAGCGC
GCCGGGGTGCCCGCCTTCCTGGAGACCTCCGCGCCCCGCAACCTCCCCTTC
TACGAGCGGCTCGGCTTCACCGTCACCGCCGACGTCGAGTGCCCGAAGGAC
CGCGCGACCTGGTGC**ATG**ACCCGCAAGCCCGGTGCCTGA

Fig. 29

ATGGTTCGACCATTGAACTGCATCGTCCGCGTGTCCCAA**ATATGGGGATT**
GGCAAGAACGGAGACCTACCCTGGCCTCCGCTCAGGAACGAGTTCAAGTAC
TTCCAAAGA**ATG**ACCACAACCTCTTCAGTGGAAGGTAAACAGAATCTGGTG
ATT**ATGGGT**AGGAAAACCTGGTTCTCCATTCTGAGAAGAATCGACCTTTA
AAGGACAGAATTAATATAGTTCTCAGTAGAGAACTCAAAGAACCACCACGA
GGAGCTCATTTTCTTGCCAAAAGTTTGGATGATGCCTTAAGACTTATTGAA
CAACCGGAATTGGCAAGTAAAGTAGAC**ATGGTT**TGGATAGTCGGAGGCAGT
TCTGTTTACCAGGAAGCC**ATGA**ATCAACCAGGCCACCTCAGACTCTTTGTG
ACAAGGATC**ATGC**AGGAATTTGAAAGTGACACGTTTTTCCAGAAATTGAT
TTGGGGAAATATAAACTTCTCCAGAATACCCAGGCGTCCTCTCTGAGGTC
CAGGAGGAAAAAGGCATCAAGTATAAGTTTGAAGTCTACGAGAAGAAAGAC
TAA

Fig. 30

ATGAAAAAGCCTGAACTCACCGCGACGTCTGTCGAGAAGTTTCTGATCGAA
AAGTTCGACAGCGTCTCCGACCTGATGCAGCTCTCGGAGGGCGAAGAATCT
CGTGCTTTCAGCTTCGATGTAGGAGGGCGTGGATATGTCCTGCGGGTAAAT
AGCTGCGCCGATGGTTTTCTACAAAGATCGTTATGTTTTATCGGCACITTTGCA
TCGGCCGCGCTCCCGATTCCGGAAGTGCTTGACATTGGGGAATTCAGCGAG
AGCCTGACCTATTGCATCTCCCGCCGTGCACAGGGTGTACGTTGCAAGAC
CTGCCTGAAACCGAACTGCCCGCTGTTCTGCAGCCGGTCGCGGAGGCCATG
GATGCGATCGCTGCGGCCGATCTTAGCCAGACGAGCGGGTTCGGCCCATTC
GGACCGCAAGGAATCGGTCAATACTACTACATGGCGTGATTTTCATATGCGCG
ATTGCTGATCCCCATGTGTATCACTGGCAAACGTGATGGACGACACCGTC
AGTGCGTCCGTGCGGCAGGCTCTCGATGAGCTGATGCTTTGGGCCGAGGAC
TGCCCCGAAGTCCGGCACCTCGTGCACGCGGATTTCCGGCTCCAACAATGTC
CTGACGGACAATGGCCGCATAACAGCGGTCATTGACTGGAGCGAGGCGATG
TTCCGGGGATTCCCAATACGAGGTCGCCAACATCTTCTTCTGGAGGCCGTGG
TTGGCTTGTATGGAGCAGCAGACGCGCTACTTCGAGCGGAGGCATCCGGAG
CTTGACAGGATCGCCGCGGCTCCGGGCGTATATGCTCCGCATTGGTCTTGAC
CAACTCTATCAGAGCTTGGTTGACGGCAATTTTCGATGATGCAGCTTGGGCG
CAGGGTCGATGCGACGCAATCGTCCGATCCGGAGCCGGGACTGTGCGGGCGT
ACACAAATCGCCCGCAGAAGCGCGGCCGTCTGGACCGATGGCTGTGTAGAA
GTACTCGCCGATAGTGGAAACCGACGCCCCAGCACTCGTCCGGAGGCAAAG
GAATTCGGGAGATGGGGGAGGCTAACTGAAACACGGAAGGAGACAATACCG
GAAGGAACCCGCGCTATGACGGCAATAAAAAGACAGAATAAAACGCACGGG
TGTTGGGTTCGTTTGTTTCATAA

Fig. 31

ATGGGATCGGCCATTGAACAAGATGGATTGCACGCAGGTTCTCCGGCCGCT
TGGGTGGAGAGGCTATTCGGCTATGACTGGGCACAACAGACAATCGGCTGC
TCTGATGCCGCCGTGTTCCGGCTGTCAGCGCAGGGGCGCCCGGTTCTTTTT
GTCAAGACCGACCTGTCCGGTGCCCTGAATGAACTGCAGGACGAGGCAGCG
CGGCTATCGTGGCTGGCCACGACGGGCGTTCCTTGCGCAGCTGTGCTCGAC
GTTGTCACTGAAGCGGGAAGGGACTGGCTGCTATTGGGCGAAGTGCCGGGG
CAGGATCTCCTGTCATCTCACCTTGCTCCTGCCGAGAAAGTATCCATCATG
GCTGATGCAATGCGGCGGCTGCATACGCTTGATCCGGCTACCTGCCCATT
GACCACCAAGCGAAACATCGCATCGAGCGAGCACGTACTCGGATGGAAGCC
GGTCTTGTCGATCAGGATGATCTGGACGAAGAGCATCAGGGGCTCGCGCCA
GCCGAACTGTTCGCCAGGCTCAAGGCGCGCATGCCCGACGGCGATGATCTC
GTCGTGACCCATGGCGATGCCCTGCTTGCCGAATATCATGGTGGAATAATGGC
CGCTTTTCTGGATTCATCGACTGTGGCCGGCTGGGTGTGGCGGACCGCTAT
CAGGACATAGCGTTGGCTACCCGTGATATTGCTGAAGAGCTTGGCGGCGAA
TGGGCTGACCGCTTCCTCGTGCTTTACGGTATCGCCGCTCCCGATTCCGAG
CGCATCGCCTTCTATCGCCTTCTTGACGAGTTCTTCTGA

Fig. 32

ATGACCACCTCAGCAAGTTCCC**ACTTAAATAAAGGCATCAAGCAGGTGTAC**
ATGTCCCTGCCTCAGGGTGAGAAAGTCCAGGCC**ATGTATATCTGGATCGAT**
GGTACTGGAGAAGGACTGCGCTGCAAGACCCGGACCCTGGACAGTGAGCCC
AAGTGTGTGGAAGAGTTGCCTGAGTGGAATTT**CGATGGCTCCAGTACTTTA**
CAGTCTGAGGGTTCCAACAGTGAC**ATGTATCTCGTGCCTGCTGCCATGTTT**
CGGGACCCCTTCCGTAAGGACCCTAACAAGCTGGTGT**TATGTGAAGTTTTC**
AAGTACAATCGAAGGCCTGCAGAGACCAATTTGAGGCACACCTGTAAACGG
ATA**ATGGACATGGT**GAGCAACCAGCACCCCTGGTTTGGC**ATGGAGCAGGAG**
TATACCCT**ATGGGGACAGATGGGCACCCCTTTGGTTGGCCTTCCAACGGC**
TCCCAGGGCCCCAGGGTCCATAT**ACTGTGGTGTGGGAGCAGACAGAGCC**
TATGGCAGGGACATCGTGGAGGCCATTACCGGGCCTGCTTGTATGCTGGA
GTCAAGATTGCGGGGACTAAT**GCCGAGGTCATGCCTGCCAGTGGGAATTT**
CAGATTGGACCTTGTGAAGGAATCAGC**ATGGGAGATCATCTCTGGGTGGCC**
CGTTTCATCTTGCATCGTGTGTGAAGACTTTGGAGTGATAGCAACCTTT
GATCCTAAGCCCATTCCTGGGA**ACTGGAATGGTGCAGGCTGCCATACCAAC**
TTCAGCACCAAGGCC**ATGCGGGAGGAGAATGGTCTGAAGTACATCGAGGAG**
GCCATTGAGAACTAAGCAAGCGGCACCAGTACCACATCCGTGCCT**ATGAT**
CCCAAGGGAGGCTGGACAAT**GCCCGACGTCTAACTGGATTCCATGAAACC**
TCCAACATCAACGACTTTTCTGGTGGTGTAGCCAA**TCGTAGCGCCAGCATA**
CGCATTCCCCGGACTGTTGGCCAGGAGAAGAAGGGT**ACTTTGAAGATCGT**
CGCCCCTCTGCCAACTGCGACCCCTTTT**CGGTGACAGAAGCCCTCATCCGC**
ACGTGTCTTCTCAAT**GAAACCGGCGATGAGCCCTTCCAGTACAAAAATTA**

FIG 33

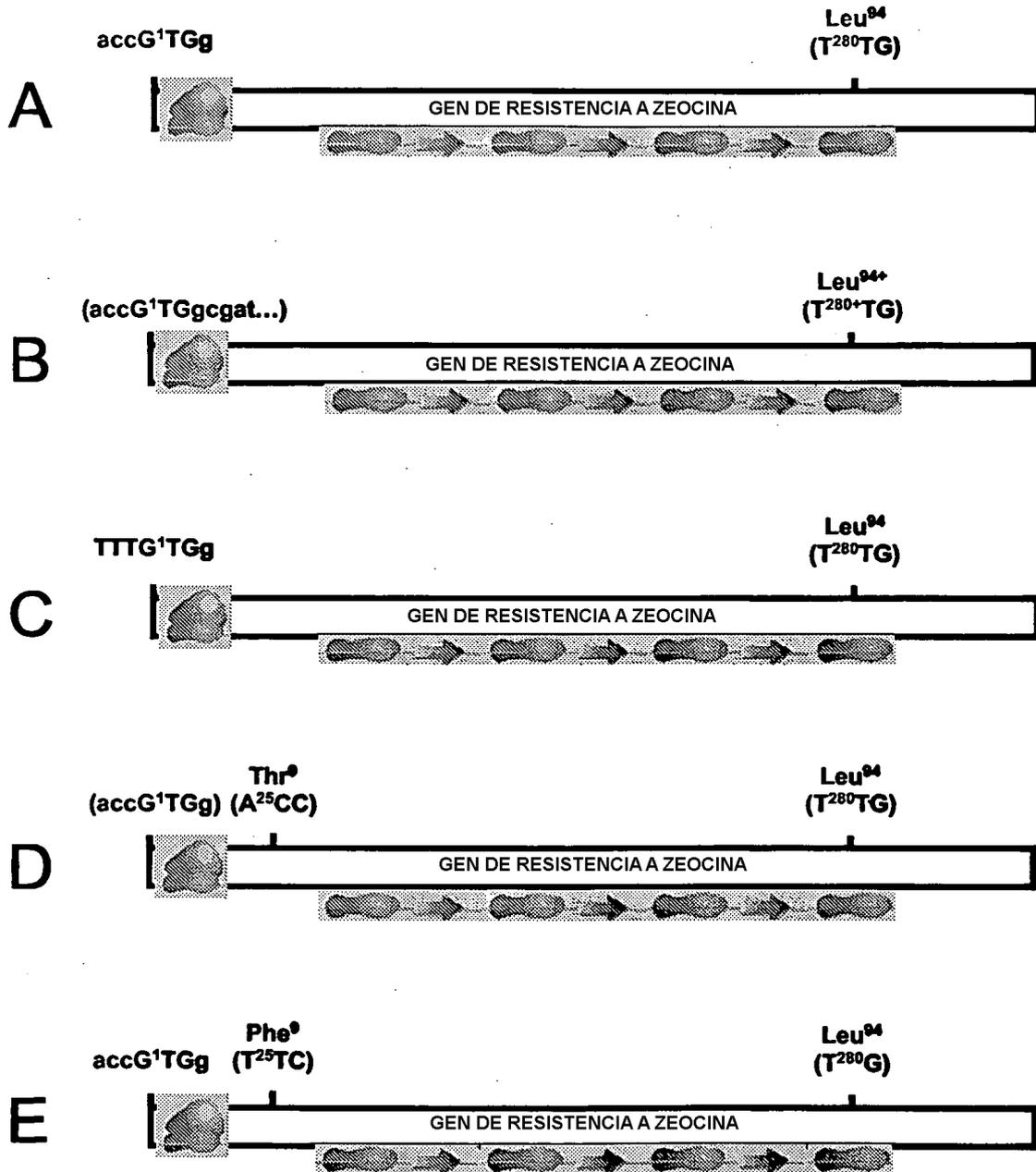


FIG 34

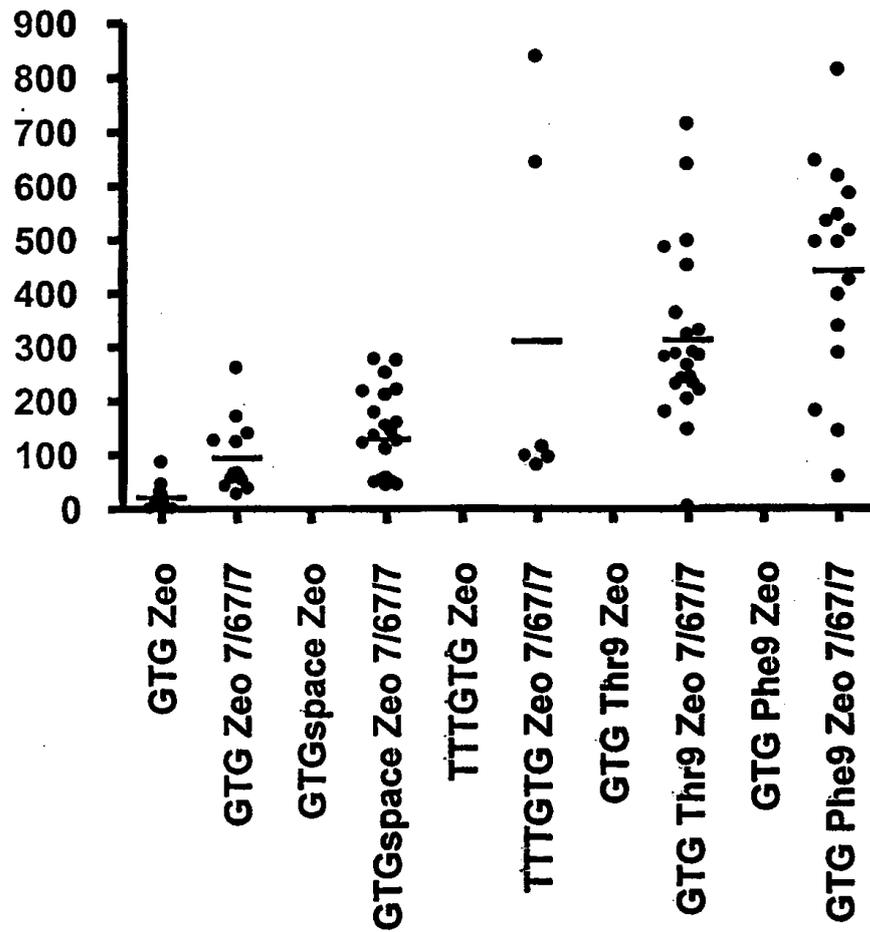
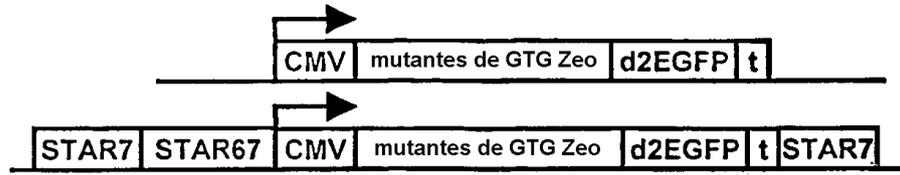


FIG 35

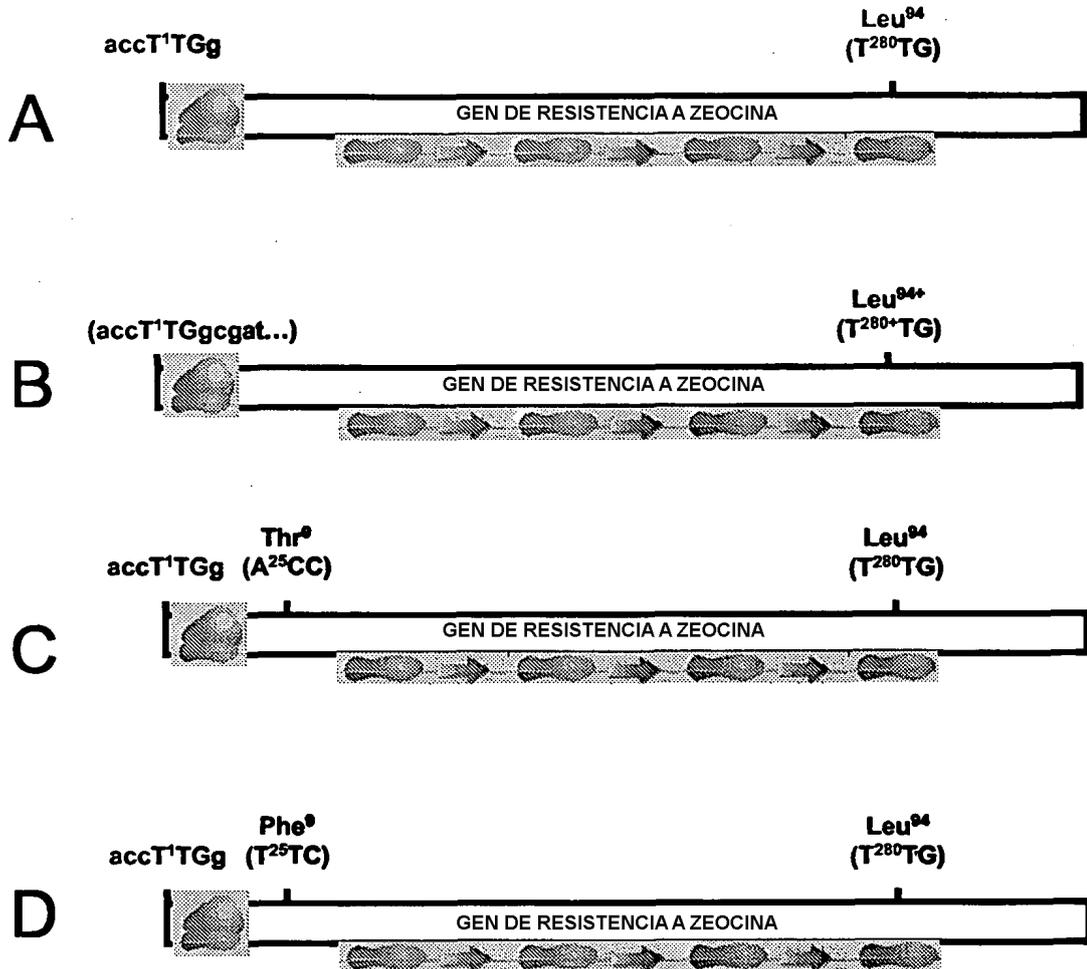


FIG 36

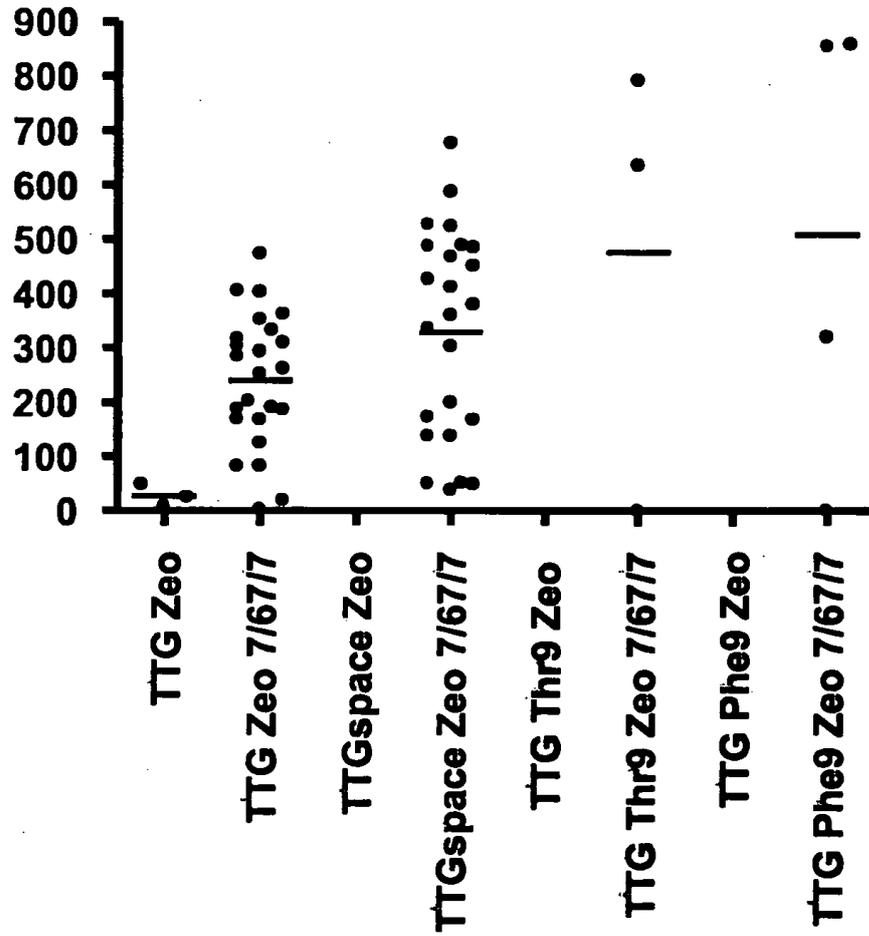
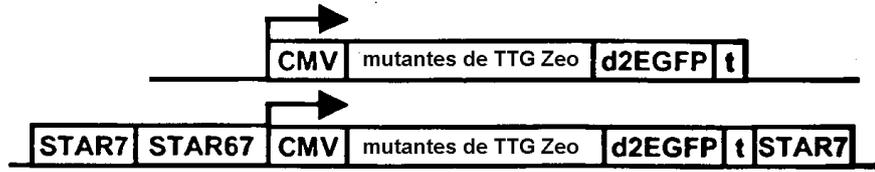


FIG 37

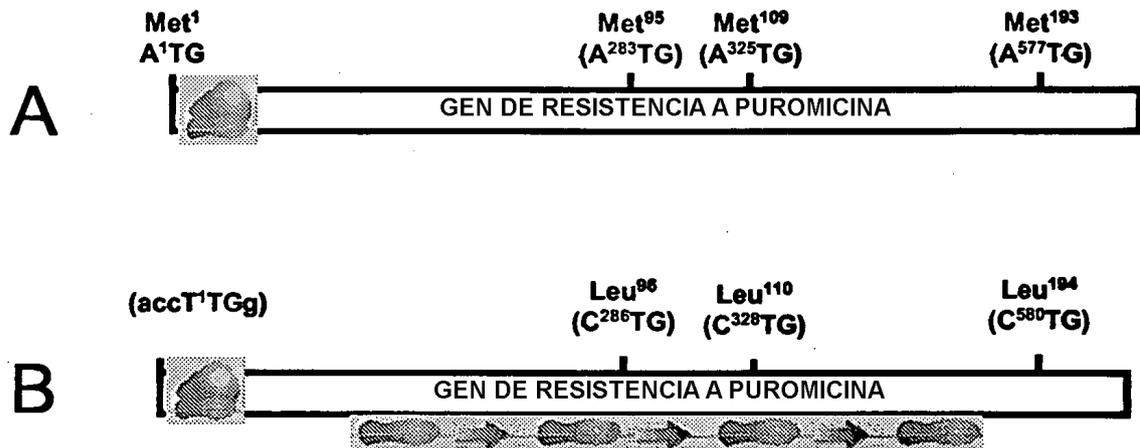


FIG 38

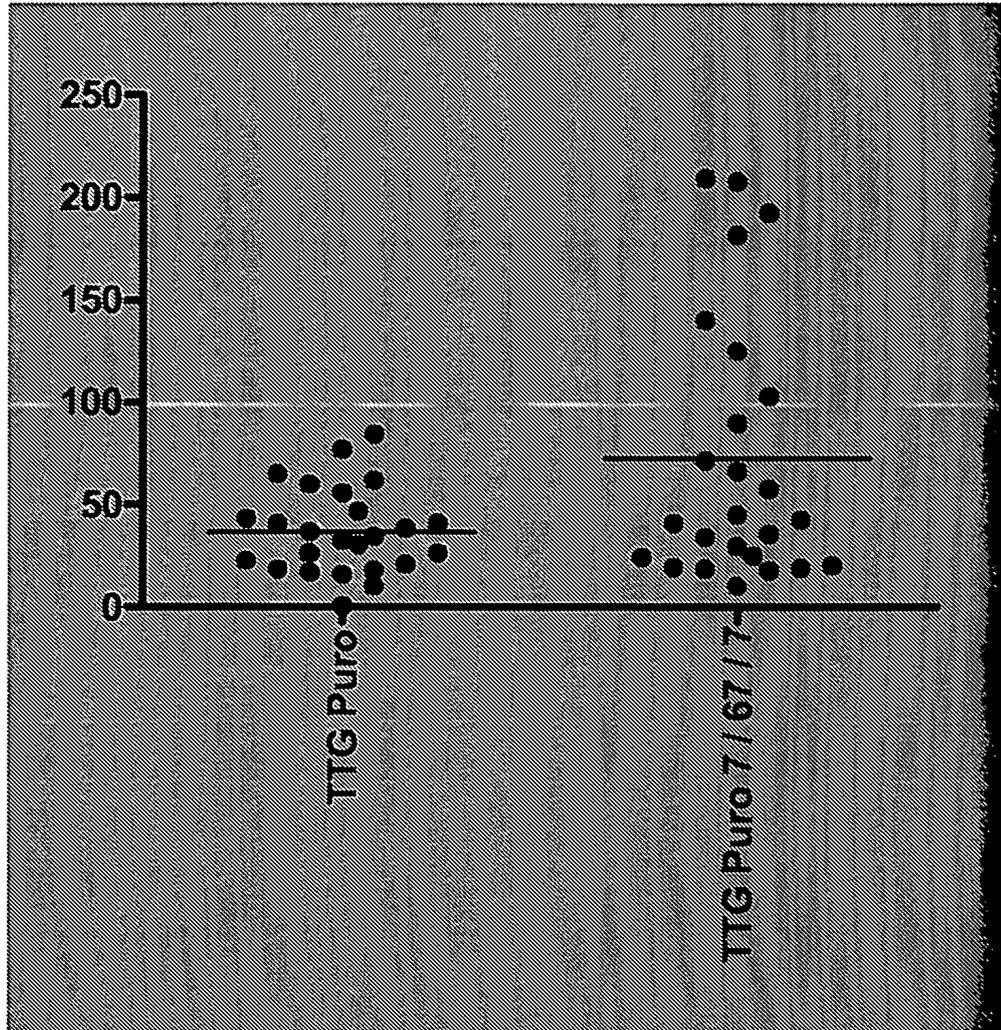
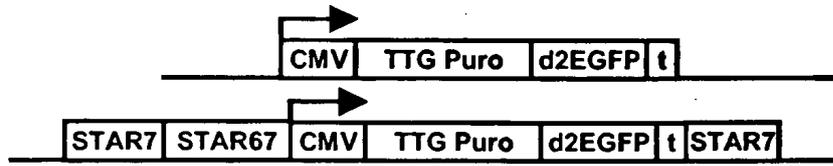


FIG 39

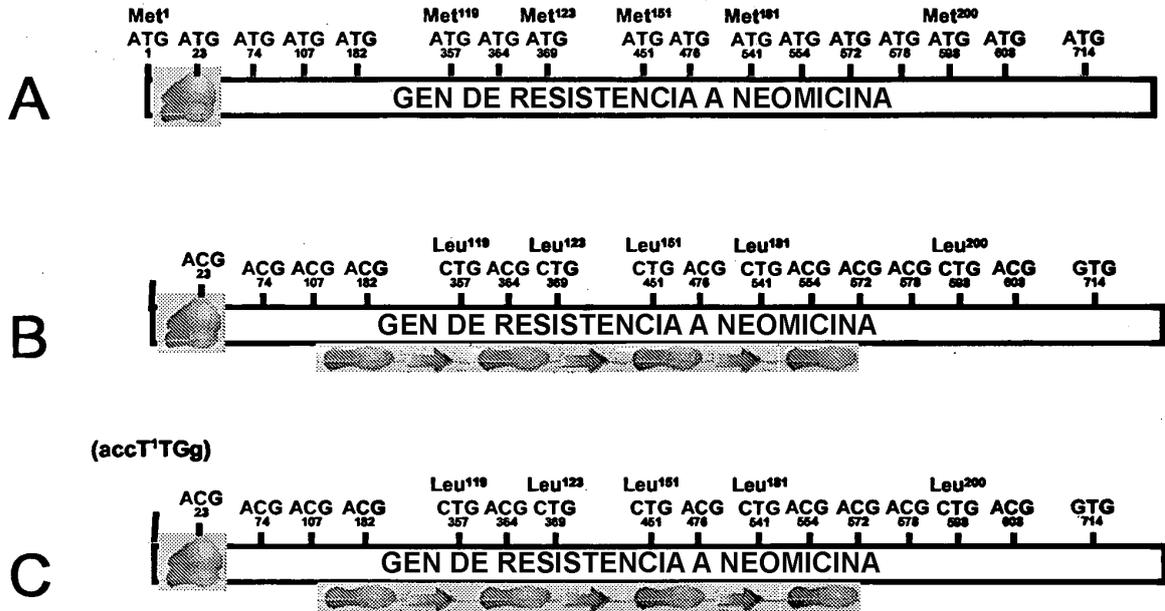


FIG 40

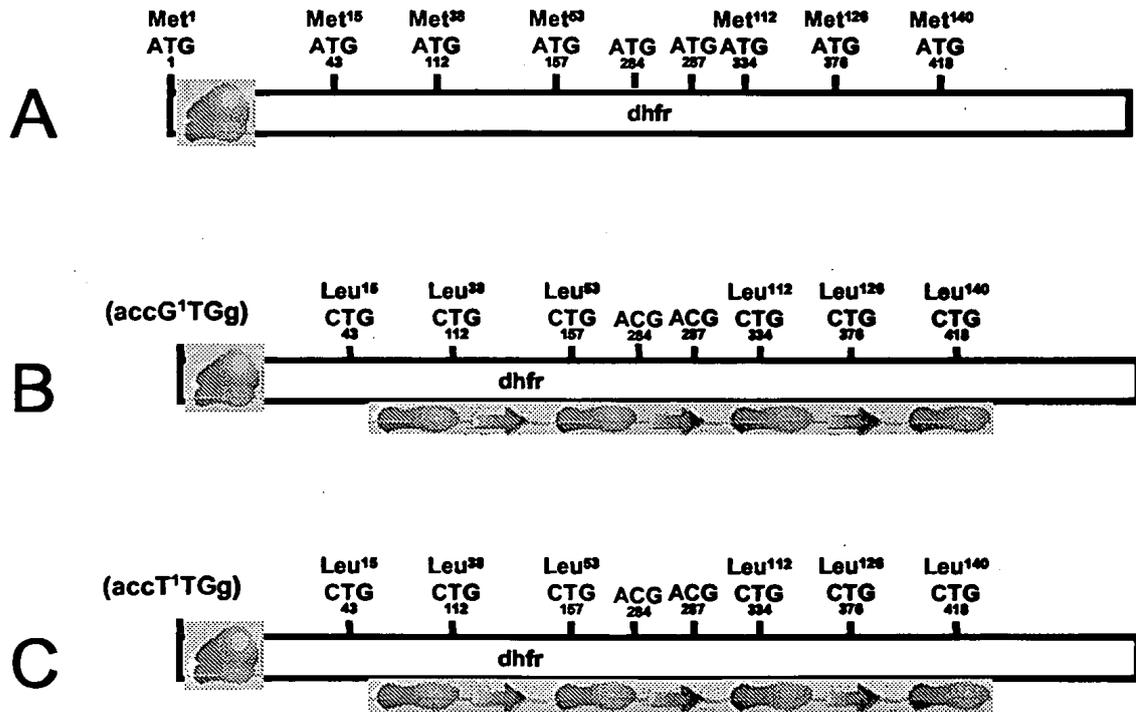


FIG 42

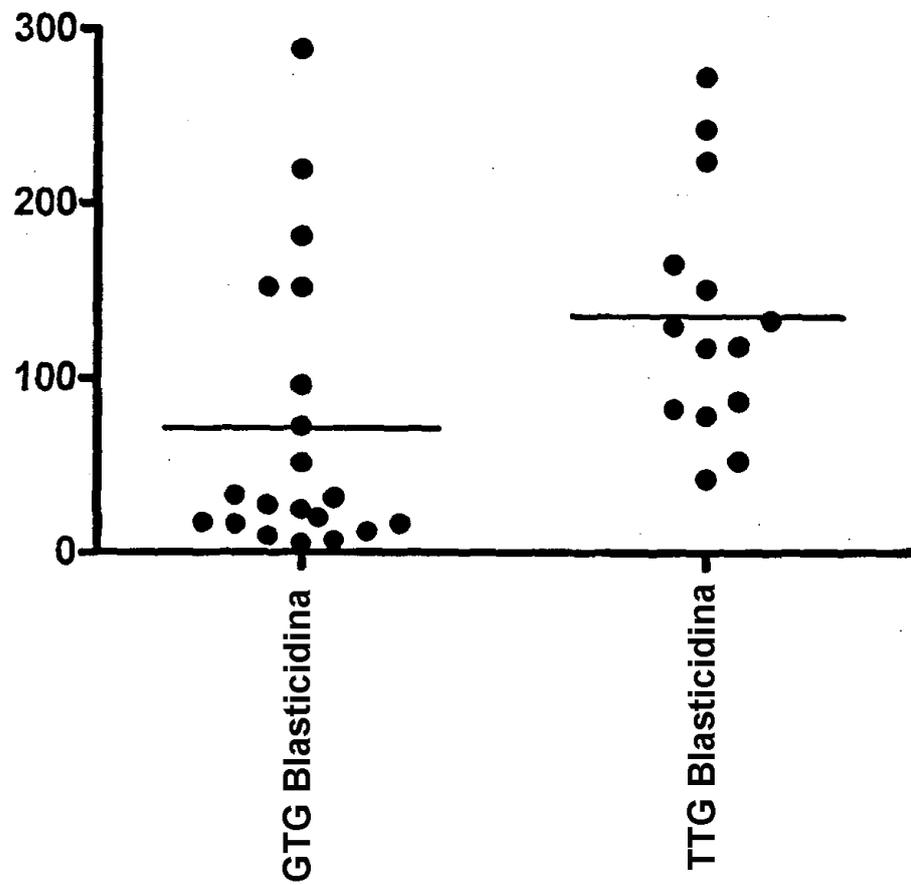
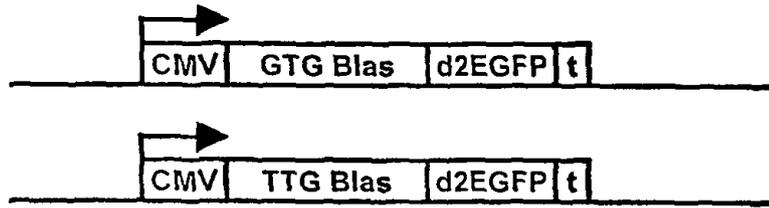


FIG 43

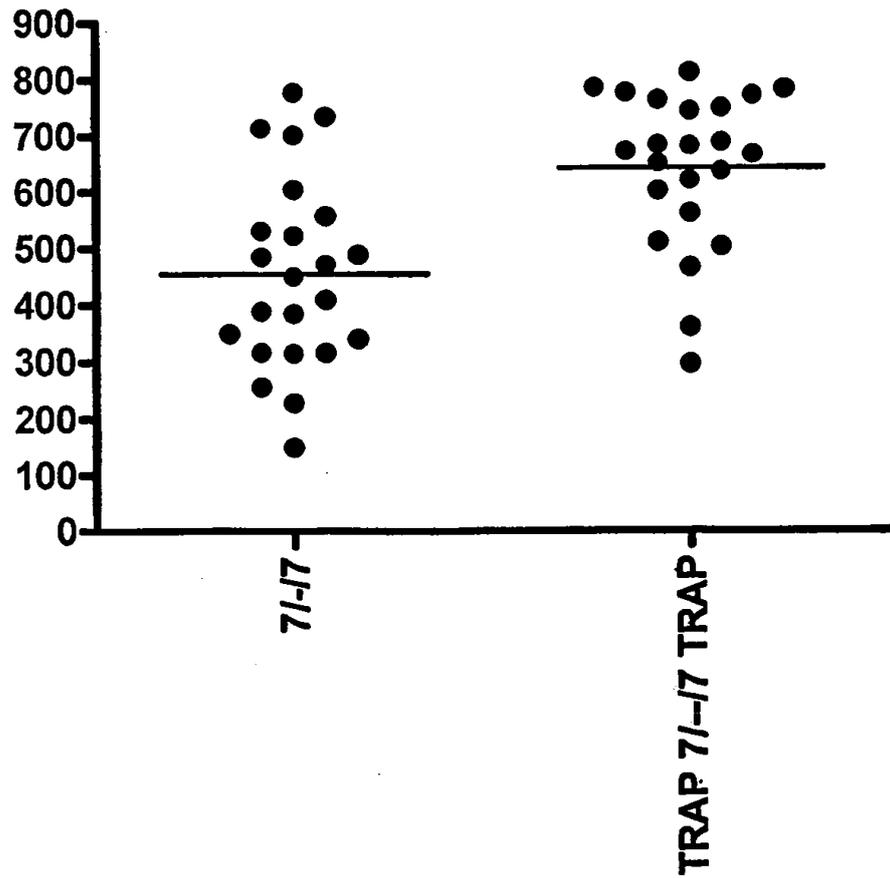
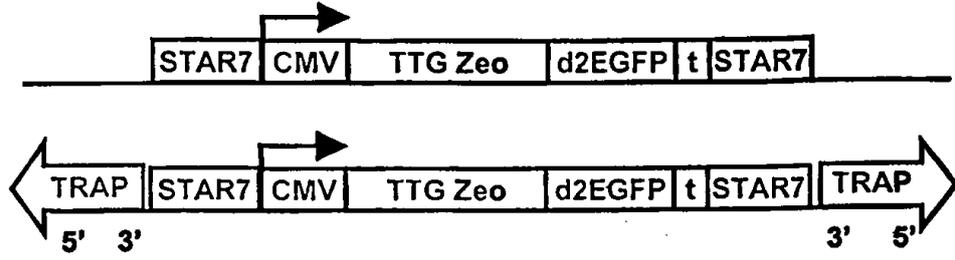


FIG 44

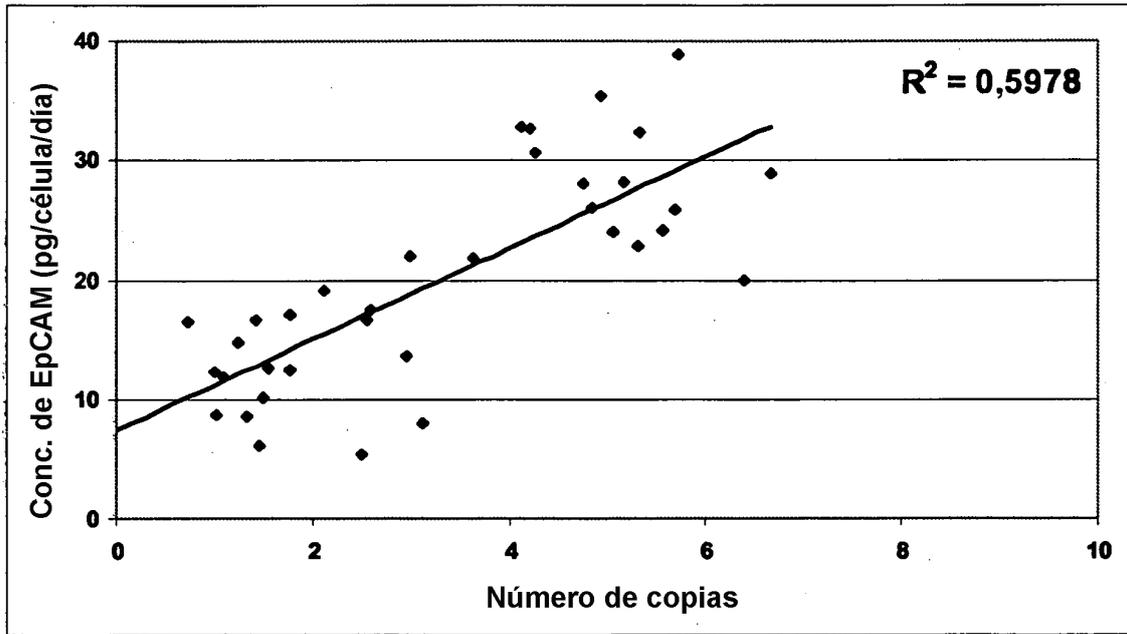


FIG 45

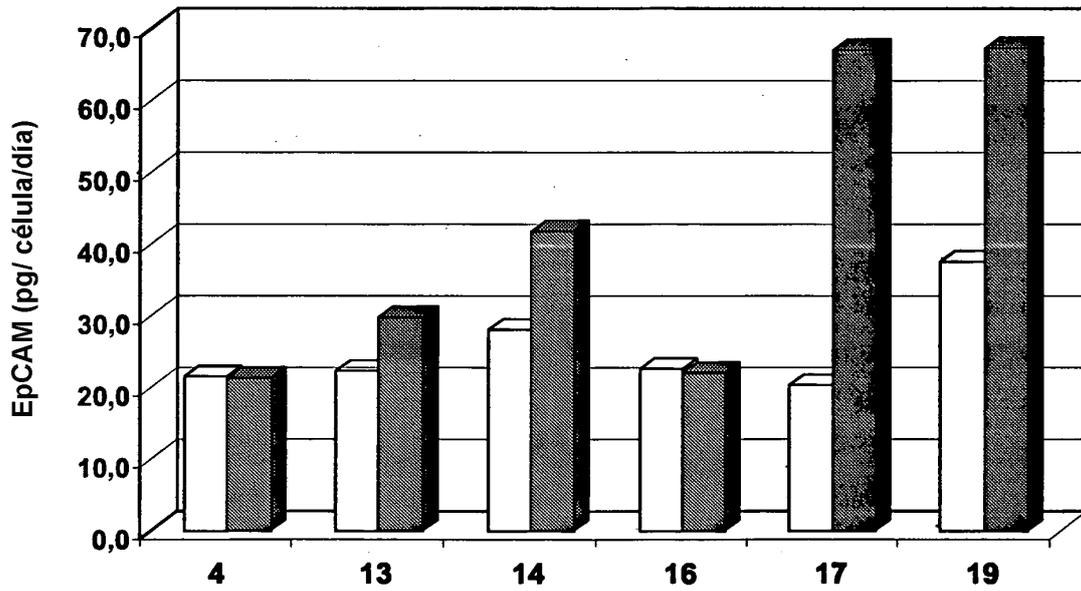
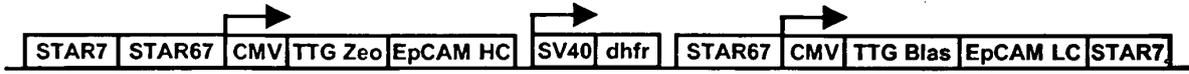


FIG 46

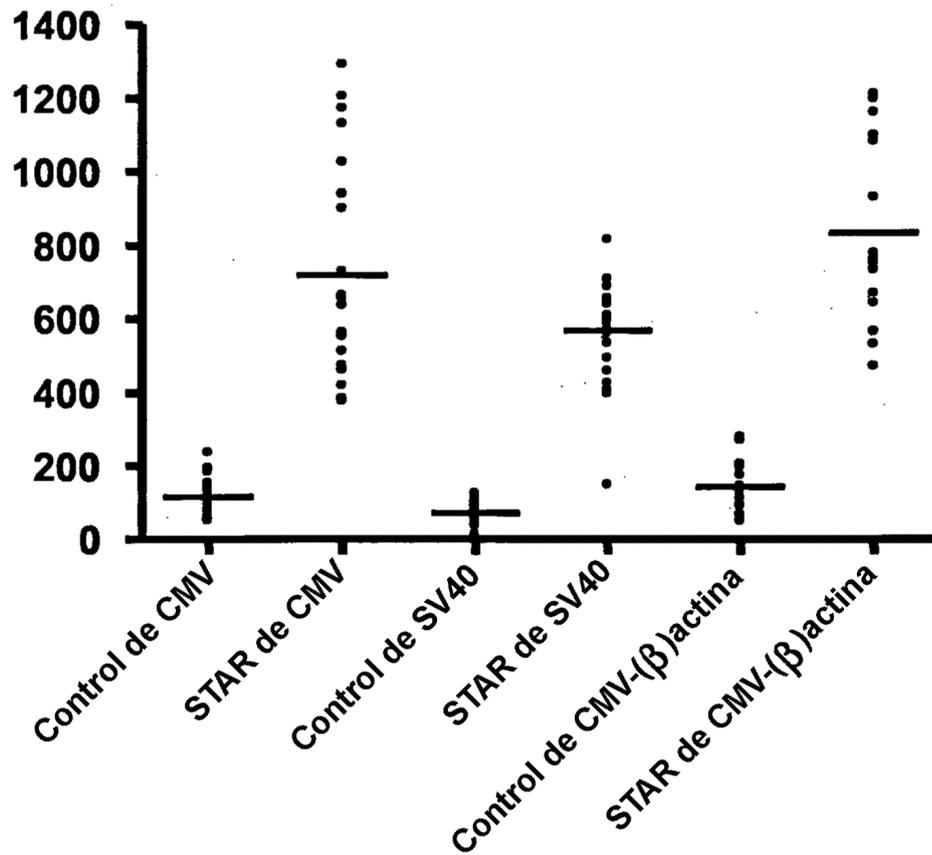
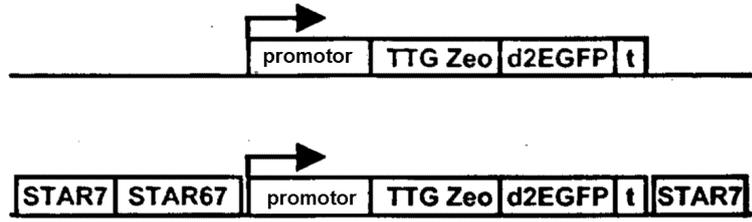


FIG 47

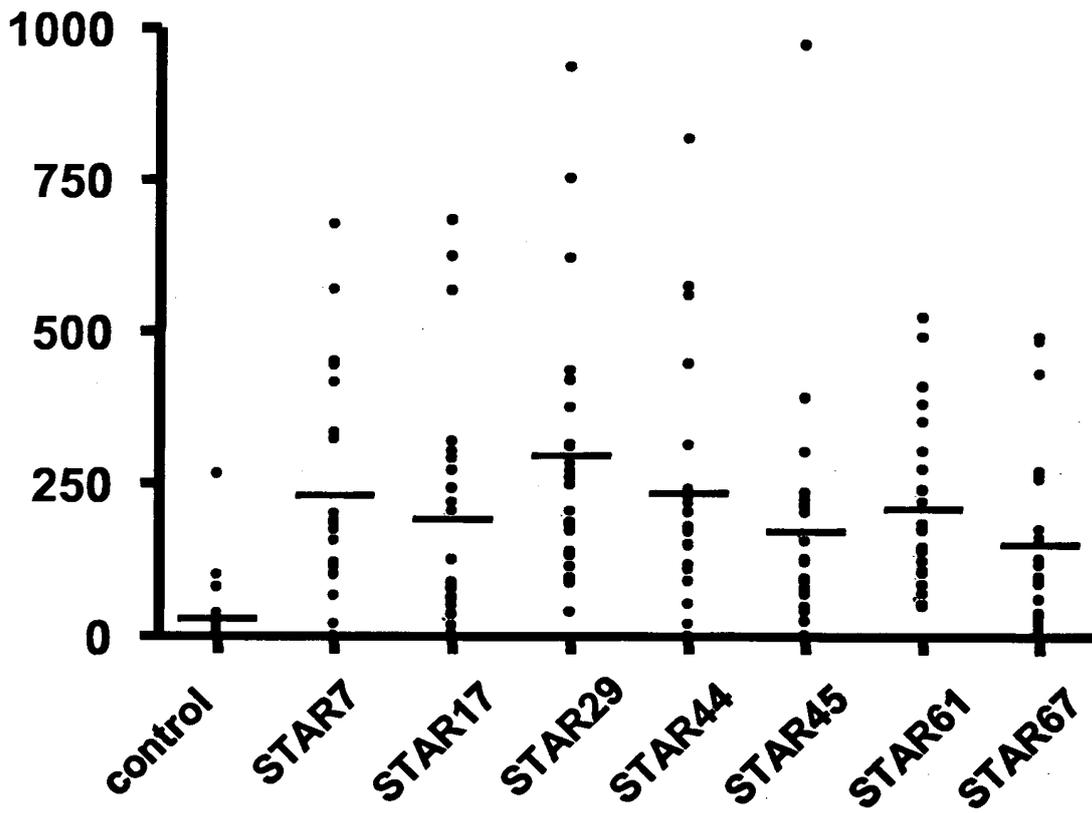
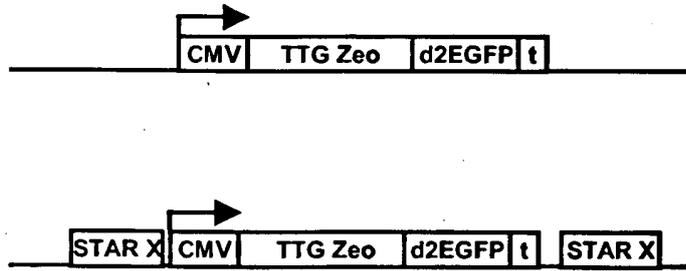


FIG 48

