

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 372 703**

51 Int. Cl.:
C12N 15/00 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

- 96 Número de solicitud europea: **07727141 .9**
96 Fecha de presentación: **20.03.2007**
97 Número de publicación de la solicitud: **1996705**
97 Fecha de publicación de la solicitud: **03.12.2008**

54 Título: **FRAGMENTOS DE ADN QUE AUMENTAN LA EXPRESIÓN, USO DE LOS MISMOS Y PROCEDIMIENTOS PARA ENCONTRARLOS.**

30 Prioridad:
20.03.2006 EP 06111381
20.03.2006 US 784133 P

45 Fecha de publicación de la mención BOPI:
25.01.2012

45 Fecha de la publicación del folleto de la patente:
25.01.2012

73 Titular/es:
CHROMAGENICS B.V.
ARCHIMEDES WEG 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: **Arias Sanz, Juan**

ES 2 372 703 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

Fragmentos de ADN que aumentan la expresión, uso de los mismos y procedimientos para encontrarlos

La invención hace referencia al campo de la biología molecular y la biotecnología. Más específicamente la presente invención hace referencia a medios y procedimientos para mejorar la expresión de ADN recombinante.

5 Las proteínas pueden producirse en diversas células huésped para un amplio intervalo de aplicaciones en biología y biotecnología, por ejemplo como productos biofarmacéuticos. Células huésped eucariotas y en particular células huésped de mamíferos se prefieren para este propósito para expresión de muchas proteínas, por ejemplo cuando tales proteínas tienen ciertas modificaciones postraduccionales tales como glucosilación. Los procedimientos para tal producción están bien establecidos y generalmente conllevan la expresión en una célula huésped de un ácido nucleico (también referido como "transgén") que codifica la proteína de interés. En general, el transgén junto con un gen marcador seleccionable se introduce dentro de una célula precursora, las células se seleccionan por la expresión del gen marcador y se identifican uno o más clones que expresan la proteína de interés a niveles altos y se usan para la expresión de la proteína de interés.

10 Se han descrito elementos de ADN que en general mejoran la expresión de ADN recombinante en células eucariotas. Ejemplos de tales elementos básicos son elementos anti-represores, tales como elementos STAR (documento WO 03/004704; Kwaks y cols., 2003), regiones de unión a matriz (MAR) (por ejemplo Phi-Van y cols., 1990; WO 02/074969, WO 2005/040377), elementos aislantes (por ejemplo West y cols., 2002; Chung y cols., 1993, 1997; Kellum y Schedl, 1991; WO 94/23046, WO 96/04390, WO 01/02553, WO 2004/027072), UCOE (documentos WO 00/05393, WO 02/24930, WO 02/099089, WO 02/099070). Estos elementos pueden al menos en parte funcionar contrarrestando los efectos represivos de la cromatina.

Otro ejemplo de elementos que mejoran la expresión recombinante son los así llamados Elementos de Secuencia que Aumentan Expresión (EASE) como se describe en el documento EP 0873405 B1. Tales elementos se obtuvieron a partir de ADN genómico de las células de ovario de hámster chino (CHO).

25 Varios de los elementos de ADN básicos que mejoran la expresión tienen una considerable longitud, siendo incluso más larga que 1 kb y a veces siendo de varias kb de longitud, lo que los hace en muchos casos menos fáciles de manipular y clonar en la configuración deseada en por ejemplo plásmidos de expresión.

Además, diferentes tipos de elementos podrían actuar de forma diferente y así puede haber potencial para mejorar expresión recombinante interviniendo a diferentes niveles moleculares.

30 Queda una necesidad en la técnica de incrementar expresión de ADN recombinante en células de mamíferos. La presente invención espera conseguir proporcionar medios y procedimientos para este propósito.

Sumario de la invención

Se reveló un concepto novedoso para seleccionar células huésped que expresen altos niveles de polipéptidos de interés en la solicitud internacional PCT/EP2005/055794 (publicada como documento WO 2006/048459), que se presentó antes pero se publicó después de la fecha de prioridad de la solicitud actual. Se reveló una alternativa en la solicitud de patente de EE.UU. n.º 11/359,953 (publicada como documento US 2006/0141577) y en la solicitud internacional PCT/EP2007/051696, también presentada antes pero publicada después de la fecha de prioridad de la solicitud actual. Las revelaciones de las solicitudes PCT/EP2005/055794, US 11/359,953 y PCT/EP2007/051696 se incorporan en su totalidad por referencia en el presente documento. Brevemente, aquellas solicitudes enseñan el uso de una secuencia que codifica un polipéptido marcador seleccionable con un codon de iniciación que no es ATG, por ejemplo un GTG o TTG. Esto dio como resultado la posibilidad de seleccionar clones con alta restrictividad y se usó para obtener clones de células huéspedes con niveles de expresión muy altos.

40 La presente invención explota la restrictividad de selección obtenida por esas secuencias codificando un polipéptido marcador seleccionable con un codon de iniciación GTG o TTG, para una revisión novedosa capaz de obtener fragmentos de ADN que aumentan la expresión. Se han descrito procedimientos para revisar en busca de elementos anti-represores en la técnica anterior, por ejemplo en el documento WO 00/09749, pero en contraste con los presentes procedimientos tales procedimientos usaban ADN que codifica proteínas marcadoras seleccionables que tienen un codon de iniciación normal, es decir un codon de iniciación ATG.

45 En un aspecto la presente invención proporciona un procedimiento de obtención de un fragmento de ADN que aumenta la expresión, procedimiento que comprende las etapas de: 1) proporcionar una diversidad de vectores que comprenden fragmentos, comprendiendo dichos vectores fragmentos de ADN que tienen un tamaño de entre aproximadamente 50 y 5000 pares de bases, estando localizados dichos fragmentos a una distancia de menos de aproximadamente 5000 pares de bases de una unidad de transcripción, comprendiendo dicha unidad de transcripción un promotor unido operativamente a la secuencia que codifica una proteína marcadora seleccionable que protege a una célula huésped de los efectos letales e inhibidores del crecimiento de un agente de selección, en el que la secuencia que codifica la proteína marcadora seleccionable tiene un codon de iniciación GTG o un codon de iniciación TTG; 2) introducir los fragmentos que comprenden vectores dentro de las células huésped; 3) cultivar

las células huésped en presencia del agente de selección; y 4) obtener el fragmento de ADN que aumenta la expresión a partir de las células huésped que aún pueden crecer en presencia del agente de selección después de al menos dos semanas.

5 La invención proporciona adicionalmente una genoteca que comprende una diversidad de vectores que comprenden fragmentos, teniendo dichos vectores que comprenden fragmentos un tamaño de entre aproximadamente 50 y 5000 pares de bases, estando localizados dichos fragmentos a una distancia de menos de aproximadamente 5000 pares de bases de una unidad de transcripción, comprendiendo dicha unidad de transcripción un promotor unido operativamente a secuencia que codifica una proteína marcadora seleccionable que protege una célula huésped de los efectos letales e inhibidores de crecimiento de un agente de selección, en la que la secuencia que codifica la proteína marcadora seleccionable tiene un codon de iniciación GTG o un codon de iniciación TTG.

10 En otro aspecto, la invención proporciona adicionalmente una molécula de ADN recombinante que comprende un fragmento de ADN que aumenta la expresión seleccionada del grupo constituido por: a) SEC. ID. N.º 3 (47D); b) SEC. ID. N.º 4 (44F); c) fragmentos de a) o b), en los que dicho fragmento tiene actividad que aumenta la expresión; y d) secuencias que son al menos idénticas al 70% en secuencia de nucleótidos a a), b) o c), en los que dichas secuencias tienen actividad que aumenta la expresión, dicha molécula de ADN recombinante comprende adicionalmente un casete de expresión, comprendiendo dicho casete de expresión un promotor heterólogo unido a un ácido nucleico de interés.

15 La invención también proporciona una célula que comprende una molécula de ADN recombinante de acuerdo con la invención.

20 La invención proporciona adicionalmente un procedimiento para producir una proteína de interés, que comprende cultivar una célula que comprende una molécula de ADN recombinante de acuerdo con la invención para expresar el ácido nucleico que codifica la proteína de interés en dicha célula.

Breve descripción de las figuras

25 FIG. 1. Resultados con fragmentos de ADN que aumentan expresión (EADF) 44F y 47D en construcciones de expresión, en las células CHO-K1. La construcción de expresión de control no contiene secuencias EADF o STAR. Como otro control, se muestra la expresión de una construcción con STAR7 flanqueando la construcción y STAR67 anteriormente en la cadena con respecto al promotor. Véase ejemplo 1 para detalles. Los puntos indican puntos de datos individuales; las líneas indican los niveles de expresión promedio; las construcciones usadas están indicadas en el eje horizontal y representadas esquemáticamente por encima de la gráfica; el eje vertical indica señal de d2EGFP en clones resistentes a zeocina.

30 FIG. 2. Como la Fig. 1, pero en las células CHO-DG44. Se usó el fragmento de ADN 47D que aumenta la expresión. Véase el ejemplo 1 para detalles.

Descripción detallada de la invención

35 El término "gen monocistrónico" se define como un gen capaz de proporcionar una molécula de ARN que codifica un polipéptido. Una "unidad de transcripción multicistrónica", también referida como gen multicistrónico, se define como un gen capaz de proporcionar una molécula de ARN que codifica al menos dos polipéptidos. El término "gen bicistrónico" se define como un gen capaz de proporcionar una molécula de ARN que codifica dos polipéptidos. Un gen bicistrónico está por lo tanto comprendido dentro de la definición de un gen multicistrónico.

40 Un "polipéptido" como se usa en el presente documento comprende al menos cinco aminoácidos unidos por enlaces peptídicos y puede por ejemplo ser una proteína o una parte, tal como una subunidad, de la misma. Mayoritariamente, los términos polipéptido y proteína se usan intercambiabilmente en el presente documento.

Un "gen" o una "unidad de transcripción" como se usa en la presente invención puede comprender ADN, ADNc, ADN artificial, combinaciones de los mismos y similares. Unidades de transcripción comprendiendo varios cistrones se transcriben como un ARNm individual.

45 Las moléculas de ADN de la invención pueden estar presentes en forma de ADN de doble hebra, teniendo con respecto al polipéptido marcador seleccionable y/o al 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, salvo por la presencia de T en vez de U. Por lo tanto, un codon de iniciación AUG está codificado en la hebra codificante por una secuencia ATG y la hebra que contiene esta secuencia ATG que corresponde al codon de iniciación AUG en el ARN se refiere como la hebra codificante del ADN. Estará claro que los codones de iniciación o las secuencias de iniciación de traducción están de hecho presentes en una molécula de ARN, pero que éstas pueden considerarse igualmente expresadas en la molécula de ADN que codifica para una molécula de ARN tal; así, dondequiera que en la presente invención se hace referencia a un codon de iniciación o a una secuencia de iniciación de la traducción, se quiere incluir a la molécula de ADN correspondiente que tiene la misma secuencia que la molécula de ARN menos por la presencia de una T en vez de un U en la secuencia codificante de dicha molécula de ADN y viceversa, 50
55 salvo donde se especifique explícitamente de otra manera. En otras palabras, un codon de iniciación es por ejemplo

una secuencia AUG en ARN, pero la secuencia ATG correspondiente en la hebra codificante del ADN se refiere también como codon de iniciación en la presente invención. Lo mismo se usa para la referencia de las secuencias codificantes “en fase”, significado tripletes (3 bases) en la molécula de ARN que se traduce en un aminoácido, pero también se interpreta como las secuencias trinucleotídicas correspondientes en la hebra codificante de la molécula de ADN.

Una secuencia de inicio de traducción se refiere a menudo en el campo como “secuencia de Kozak” y una secuencia de Kozak óptima es RCCATGG, el codon de iniciación subrayado, siendo R una purina, es decir A o G (véase Kozak M, 1986, 1987, 1989, 1990, 1997, 2002). Así, además del codon de iniciación por sí mismo, el contexto del mismo, en particular nucleótidos -3 a -1 y +4, son relevantes y una traducción óptima de la secuencia de inicio comprende un codon de inicio óptimo (es decir ATG) en un contexto óptimo (es decir el ATG precedido directamente por RCC y seguido directamente por G). La traducción por los ribosomas comienza preferentemente en el primer codon de iniciación que encuentran cuando exploran a partir del casquete 5' del ARNm (mecanismo de barrido) y es más eficiente cuando está presente una secuencia de Kozak óptima (véase Kozak M, 1986, 1987, 1989, 1990, 1997, 2002). Sin embargo, en un pequeño porcentaje de eventos, se reconocen secuencias de iniciación de la traducción no óptimas y se usan por el ribosoma para empezar la traducción. La invención hace uso de este principio y permite disminuir la cantidad de traducción y así de expresión del polipéptido marcador seleccionable, que puede por lo tanto usarse para incrementar la restrictividad del sistema de selección y de ese modo pueden identificarse fragmentos de ADN que aumentan la expresión nuevos. Para este propósito, el codon de iniciación ATG en sí mismo del polipéptido marcador seleccionable está mutado en un codon de iniciación GTG o en un codon de iniciación TTG.

Las secuencias de codificación de la proteína marcadora seleccionable de los documentos WO 2006/048459 y US 2006/0141577 en las realizaciones preferidas tienen un codon de iniciación GTG o más preferentemente un codon de iniciación TTG. Como una consecuencia del sistema de selección restrictivo con un codon de iniciación GTG o TTG para la secuencia codificante marcadora seleccionable, ningún clon o casi ningún clon se desarrollará, a menos que el vector haya incorporado un fragmento de ADN que aumente la expresión. Por lo tanto, este sistema permite la detección y aislamiento de tales elementos novedosos con alta restrictividad y alta eficiencia.

En un aspecto, la presente invención proporciona por lo tanto un procedimiento de obtención de un fragmento de ADN que aumenta la expresión, procedimiento que comprende las etapas de: 1) proporcionar una diversidad de vectores que comprenden fragmentos, comprendiendo dichos vectores fragmentos de ADN que tienen un tamaño de entre aproximadamente 50 y 500 pares de bases estando dichos fragmentos de ADN localizados a una distancia de menos de aproximadamente 5000 pares de bases de una unidad de transcripción, comprendiendo dicha unidad de transcripción un promotor operativamente unido a secuencia que codifica una proteína marcadora seleccionable que protege a una célula hospedadora de los efectos letales o inhibidores del crecimiento de un agente de selección, en el que la secuencia que codifica la proteína marcadora seleccionable tiene un codon de iniciación GTG o un codon de iniciación TTG; 2) introducir los vectores que comprenden fragmentos dentro de las células huésped; 3) cultivar las células huésped en presencia del agente de selección; y 4) obtener el fragmento de ADN que aumenta la expresión de células huésped que aún pueden crecer en presencia del agente de selección después de al menos dos semanas. Este aspecto de la invención se refirió como el procedimiento de revisión de la invención.

Los fragmentos de ADN pueden por ejemplo obtenerse adecuadamente a partir del genoma de una célula, incluyendo una célula de mamífero, por ejemplo a partir de una célula humana, una célula de ratón, una célula de hámster, una célula de rata y similares. Los fragmentos pueden por ejemplo obtenerse también a partir de una genoteca de una parte de un genoma o de un genoma completo. Podrían usarse también genomas de otros organismos como la fuente para los fragmentos de ADN, incluyendo por ejemplo plantas, levaduras, *C. elegans*, pez cebra, *Drosophila*, etc., o incluso de bacterias o virus. En principio, los fragmentos de ADN sintético podrían ponerse también a prueba en el procedimiento de revisión de la invención. Los fragmentos pueden clonarse por técnicas de biología molecular convencionales dentro de los vectores, obteniendo así la diversidad de los vectores que contienen fragmentos. En las realizaciones preferidas, los fragmentos tienen un tamaño de entre aproximadamente 100 y 2000 pares de bases, más preferentemente entre aproximadamente 200 y 1000 pares de bases. La ventaja de usar fragmentos con un tamaño promedio menor es que los fragmentos de ADN que aumentan la expresión que se seleccionarán serán menores y así más fáciles de manipular debido a que proporcionarán menor carga sobre la capacidad, es decir el espacio disponible para clonar ADN de interés, de vectores de expresión en los que ellos se usarán.

En ciertas realizaciones, los fragmentos de ADN están localizados anteriormente en la cadena con respecto al promotor de la unidad de transcripción. También es posible examinar con vectores en los que los fragmentos estén en localizaciones diferentes, por ejemplo más adelante en la cadena con respecto a la unidad de transcripción. En realizaciones preferidas, los fragmentos se localizan a una distancia de menos de aproximadamente 2000 pares de bases a partir de la unidad de transcripción. Disminuir la distancia entre los fragmentos y la unidad de transcripción evita que ciertos fragmentos tuvieran actividad que aumente la expresión pero que no se detecten debido a que deberían trabajar mejor cuando se sitúan más cerca de la unidad de transcripción. En realizaciones más preferidas, los fragmentos de ADN están localizados a una distancia de menos de 800, 700, 600, preferentemente de menos de aproximadamente 500, 400, 300, 200, 100 o 50 pares de bases de la unidad de transcripción. En ciertas realizaciones, los fragmentos se localizan anteriormente en la cadena con respecto al promotor de la unidad de transcripción y el final de los fragmentos está separado por menos de 500, preferentemente por menos de 20, más

preferentemente por menos de 100 o menos de 50 pares de bases del comienzo de la secuencia del promotor.

El agente de selección puede ser cualquier agente de selección que confiera efectos letales o inhibidores del crecimiento a las células huésped, y se pueden usar por ejemplo los ejemplos de agentes de selección descritos en el documento WO 2006/048459. En una realización preferida, el agente de selección es zeocina. Se ha revelado en el documento WO 2006/048459 que las secuencias que codifican una proteína marcadora seleccionable proporcionan resistencia a un agente de selección con un codon de iniciación GTG o TTG hecho funcionar en procedimientos de selección restrictivos. Esta claro a partir de ahí que una amplia diversidad de posibilidades con respecto a proteínas marcadoras seleccionables y existen son funcionales y así podrían emplearse en principio en el procedimiento de revisión de la presente invención para encontrar fragmentos de ADN que aumenten la expresión. En una realización preferida, la secuencia que codifica el marcador seleccionable tiene un codon de iniciación TTG.

El término "marcador de selección" o "marcador seleccionable" se utiliza típicamente para hacer referencia a un gen y/o proteína cuya presencia se puede detectar directa o indirectamente en una célula, por ejemplo un polipéptido que inactiva un agente de selección y protege de los efectos letales o inhibidores del crecimiento del agente (por ejemplo un gen y/o una proteína de resistencia a antibióticos). Otra posibilidad es que dicho marcador de selección induzca fluorescencia o un depósito de color (por ejemplo proteína verde fluorescente (GFP) y derivados (por ejemplo d2EGFP), luciferasa, lacZ, fosfatasa alcalina, etc.), que puedan usarse para seleccionar células que expresen el polipéptido induciendo el depósito de color, usando por ejemplo un clasificador celular activado de fluorescencia (FACS) para seleccionar células que expresen GFP. Se conocen bien en la técnica polipéptidos marcadores seleccionables y se usan rutinariamente cuando hay que obtener clones de células huésped eucariotas y se proporcionan varios ejemplos de proteínas marcadoras seleccionables adecuadas en el documento WO 2006/048459. Se conocen las secuencias de ADN que codifican para tales polipéptidos marcadores seleccionables y se proporcionan varios ejemplos de secuencias de tipo silvestre de ADN que codifican proteínas marcadoras seleccionables en el documento WO 2006/048459 (por ejemplo Figs. 15-21 en él, incorporadas por referencia en el presente documento). Estará claro que mutantes o derivados de marcadores seleccionables se pueden usar también adecuadamente y están incluidos por lo tanto dentro del alcance del término "polipéptido marcador seleccionable", siempre que la proteína marcadora seleccionable esté aún funcional. Un polipéptido marcador seleccionable de acuerdo con la invención es una proteína que está codificada por ácido nucleico, polipéptido que puede usarse funcionalmente para selección, por ejemplo debido a que proporciona resistencia a un agente de selección tal como a un antibiótico. Así, cuando se usa un antibiótico como un agente de selección, el ADN codifica un polipéptido que confiere resistencia al agente de selección, polipéptido que es un polipéptido marcador seleccionable. En ciertas realizaciones, un marcador de selección usado para la invención es zeocina. La persona experta en la técnica sabrá que están disponibles y que pueden usarse otros marcadores de selección, por ejemplo blasticidina, neomicina, puromicina, bleomicina, higromicina, etc. En otras realizaciones, se usa kanamicina. En aún otras realizaciones, se usa el gen DHFR como un marcador seleccionable, que puede seleccionarse por metotrexato, especialmente incrementando la concentración de metotrexato las células pueden seleccionarse para números de copias incrementadas del gen DHFR (o, alternativamente, el gen DHFR puede seleccionarse en la ausencia de metotrexato en células que son deficientes en dhfr). De forma similar, se puede usar el gen de la glutamina sintetasa (GS), para el que la selección es posible en células que tienen GS insuficiente (por ejemplo, células NS-0) cultivando en medios sin glutamina, o alternativamente en células que tienen suficiente GS (por ejemplo células de CHO) añadiendo un inhibidor de GS, metionina sulfoximina (MSX). Otros genes marcadores seleccionables que se podrían usar, y sus agentes de selección, se describen por ejemplo en tabla 1 de patente de los Estados Unidos 5,561,053; véase también Kaufman, *Methods in Enzymology*, 185: 537-566 (1990), para una revisión de éstos. Estará claro que una "proteína marcadora de selección que protege una célula huésped de los efectos letales o inhibidores del crecimiento de un agente de selección" puede incluir también proteínas que protegen a las células de morir en ausencia de un compuesto que se requiera para el crecimiento (es decir el agente de selección puede en ciertas realizaciones ser un medio de cultivo que carezca de uno o más compuestos requeridos para el crecimiento). El término "selección" se define típicamente como el procedimiento de usar un marcador de selección/marcador seleccionable y un agente de selección para identificar células huésped con propiedades genéticas específicas (por ejemplo que la célula hospedadora contiene un transgén integrado dentro de su genoma). Por conveniencia y como se acepta generalmente por la persona experta, en muchas publicaciones así como en el presente documento, a menudo el gen y la proteína que codifican la resistencia a un agente de selección se refieren como el "gen de (resistencia a) agente seleccionable" o "proteína de (resistencia a) agente de selección", respectivamente, aunque los nombres oficiales pueden ser diferentes, por ejemplo el gen que codifica para la proteína confiriendo resistencia a neomicina (así como a G418 y kanamicina) se refiere a menudo como gen de (de resistencia a) neomicina (o neo^r), mientras que el nombre oficial es gen de aminoglicósido 3'-fosfotransferasa.

En principio, el uso de una proteína marcadora seleccionable es suficiente para obtener fragmentos de ADN que aumentan expresión, dado que el procedimiento de revisión de la presente invención se destina principalmente a identificar un fragmento de ADN que aumenta expresión adecuada. En contraste, los procedimientos descritos en los documentos WO 2006/048459 y US 2006/0141577 se destinan a proporcionar condiciones de selección para obtener células con expresión alta de un polipéptido de interés. Así, la diferencia es que para el presente procedimiento de revisión, un polipéptido de interés no juega necesariamente un papel. Por lo tanto, la unidad de transcripción puede ser en algunas realizaciones una unidad de transcripción monocistrónica, que sólo después codificará la proteína marcadora seleccionable.

Sin embargo, en ciertas realizaciones, la unidad de transcripción es una unidad de transcripción multicistrónica, tal como una unidad de transcripción bicistrónica, que comprende la secuencia que codifica la proteína marcadora seleccionable que proporciona resistencia a un inhibidor de crecimiento y que comprende adicionalmente una secuencia que codifica una proteína de la que se puede detectar la presencia. La ventaja de tales realizaciones multicistrónicas es que el número de "falsos positivos" puede reducirse analizando si la proteína de la que se puede detectar la presencia se expresa también (es decir si se expresa además suficiente proteína marcadora seleccionable para permitir el crecimiento de las células huésped en presencia del inhibidor de crecimiento). Una proteína de la que se puede detectar la presencia puede ser casi cualquier proteína, dado que para la mayoría de las proteínas la presencia se puede detectar por ejemplo por anticuerpos, por ejemplo en un ELISA. En las realizaciones preferidas sin embargo, la proteína de la que se puede detectar la presencia es una proteína comunicadora como se conoce bien por la persona experta en la técnica, por ejemplo una proteína fluorescente verde (GFP) o un derivado de la misma, luciferasa, beta-galactosidasa, SEAP, CAT, una enzima que cataliza una reacción que se puede detectar añadiendo sustrato y similares. Estas proteínas pueden detectarse fácilmente por ejemplo formando un depósito de color o por fluorescencia en ciertas condiciones bien conocidas. La proteína de la que se puede detectar la presencia puede ser también otra proteína marcadora seleccionable, es decir una que proporcione resistencia a otro agente de selección, de tal modo que la selección subsiguiente o concomitante con este segundo agente de selección detectará la presencia de esta proteína (indirectamente: la proteína debe estar presente de lo contrario las células no crecerían) y con eso reducirá el número de falsos positivos. En tales unidades de transcripción multicistrónica usadas para revisión en el procedimiento de la presente invención, se pueden usar ambas configuraciones como se revelan en el documento WO 2006/048459 y como se revelan en el documento US 2006/0141577. En una realización por lo tanto, la unidad de transcripción en el procedimiento de revisión de la invención es una unidad de transcripción multicistrónica que comprende en el siguiente orden: a) el promotor; b) la secuencia que codifica la proteína marcadora seleccionable que protege una célula huésped de los efectos letales o inhibidores del crecimiento de un agente de selección; y c) la secuencia que codifica una proteína de la que se puede detectar la presencia. Como se revela en el documento WO 2006/048459, en esta primera realización se prefiere fuertemente que la proteína marcadora seleccionable de b) esté desprovista de secuencias ATG, de tal forma que la primera ATG será el codón de iniciación de la proteína que está más adelante en la cadena en el cistron. En otra realización, la unidad de transcripción en el procedimiento de revisión de la invención es una unidad de transcripción multicistrónica que comprende en el siguiente orden: a) el promotor; b) la secuencia que codifica una proteína de la que la presencia puede ser detectada; y c) un sitio de entrada del ribosoma interno (IRES), unido operativamente a d) la secuencia que codifica la proteína marcadora seleccionable que proporciona resistencia al inhibidor de crecimiento. Las unidades de transcripción multicistrónica de esta segunda realización se han descrito en el documento US 2006/0141577.

El procedimiento de revisión de la presente invención tiene en cuenta obtener el fragmento de ADN que aumenta la expresión deseada a partir de células huésped que pueden crecer aún en la presencia del inhibidor del crecimiento después de al menos dos semanas, usando procedimientos de clonación moleculares de rutina bien conocidos para la persona experta en la técnica. Rondas adicionales de revisión pueden llevarse a cabo para reducir la complejidad de la genoteca que comprende la variedad de vectores que comprenden fragmentos, si es deseable. Usando la secuencia conocida de los vectores, los fragmentos de ADN con actividad que aumenta la expresión sospechada que resultan de una ronda de revisión se pueden clonar, usando por ejemplo la reacción en cadena de la polimerasa (PCR), empleando cebadores que circundan el sitio en el vector de revisión en el que se ha clonado la diversidad de los fragmentos. La secuencia del fragmento de ADN que aumenta la expresión se puede analizar, y la actividad que aumenta la expresión se puede confirmar en una amplia variedad de condiciones, si se desea.

Las células huésped con los fragmentos de ADN que aumentan la expresión identificados en el procedimiento de revisión de la presente invención serían capaces de crecer en presencia del agente de selección después de al menos una semana, preferentemente después de al menos dos semanas. Esto asegura que la proteína marcadora seleccionable se expresa establemente y así se aíslan los fragmentos de ADN que aumentan la expresión estable, mientras que descartan cualesquiera fragmentos que trabajen sólo muy poco. En determinadas realizaciones, esta etapa de selección puede continuarse durante al menos 3 semanas, o incluso (al menos) 4 semanas, 5 semanas, o más. El tiempo de selección mínimo óptimo puede depender en parte del agente de selección y la concentración de los mismos, pero en cualquier caso las células huésped que sobreviven después de un cierto periodo de selección permanecerán resistentes al agente de selección, de tal forma que en principio no hay límite superior al tiempo de selección. En la práctica, la persona especializada en la técnica encontrará empíricamente un tiempo de selección adecuado. Si por ejemplo las células huésped se cultivan en placas que contienen el agente de selección, alguien puede simplemente esperar hasta que se formen las colonias.

En realizaciones preferidas, los vectores no se replican en las células huésped y se integrarán (al azar) dentro del genoma de las células huésped. Esto está en contraste con por ejemplo el ensayo usado para obtener elementos de STAR, donde se usaron vectores episomales (documento WO 03/004704). Una ventaja de usar vectores de integración es que sólo un número limitado de copias del vector se integrarán usualmente dentro del genoma, de tal forma que los falsos positivos que resultan únicamente de números de copias altos se evitan. Además, los sistemas de expresión recombinantes usan en general preferentemente transgenes integrados debido a la estabilidad. Así otra ventaja en revisión de acuerdo con la presente invención con vectores integrados puede ser que los fragmentos de ADN que aumentan la expresión obtenidos después por definición funcionan cuando se integran en el genoma de

las células huésped.

En una realización, las células huésped en el procedimiento de rastreo de la invención son células de CHO. Esto asegura que se obtienen fragmentos de ADN que aumentan la expresión que son funcionales en una célula huésped ampliamente usada para propósitos de expresión recombinante, es decir células de CHO. Claramente, otras células huésped eucariotas, preferentemente de mamíferos, se pueden usar de acuerdo con el procedimiento de revisión de la invención.

En el procedimiento de rastreo, la invención emplea vectores que contienen fragmentos, que contienen una secuencia que codifica una proteína marcadora seleccionable con codón de iniciación GTG o TTG y que contienen adicionalmente fragmentos de ADN de entre aproximadamente 50 y 5000 pares de bases. Una recogida de tales vectores forma una genoteca, que es un producto intermedio adecuado que puede usarse en el procedimiento de rastreo de la invención. En un aspecto separado la invención proporciona por lo tanto una genoteca que comprende una diversidad de vectores que comprenden fragmentos, comprendiendo dichos vectores fragmentos de ADN que tienen un tamaño de entre aproximadamente 50 y 5000 pares de bases, estando dichos fragmentos de ADN localizados a una distancia de menos de aproximadamente 5000 pares de bases de una unidad de transcripción, comprendiendo dicha unidad de transcripción un promotor unido operativamente a secuencia que codifica una proteína marcadora seleccionable que protege a una célula huésped de los efectos letales o inhibidores de crecimiento de un agente de selección, en la que la secuencia que codifica la proteína marcadora seleccionable tiene un codón de iniciación GTG o un codón de iniciación TTG. Estará claro que la genoteca puede variarse a lo largo de las realizaciones descritas anteriormente para el procedimiento de revisión. Un genoteca como se usa en el presente documento es una colección de vectores, que comprenden el mismo vector armazón con un fragmento de ADN individual clonado en cada vector armazón y conteniendo en dicha colección al menos 5 fragmentos de ADN diferentes y en ciertas realizaciones al menos 10, 100, 1000, 10^4 , 10^5 , 10^6 , 10^7 , 10^8 , 10^9 , 10^{10} , 10^{11} , 10^{12} o más fragmentos de ADN diferentes clonados dentro de los vectores armazón. Mayor complejidad de la genoteca tiene la ventaja de incrementar la posibilidad de obtener fragmentos de ADN que aumenten la expresión. El procedimiento permite revisar partes de genomas o genomas completos en busca de fragmentos con actividad que aumente la expresión. Después de una ronda de revisión con el procedimiento de revisión, es posible crear una sub-genoteca que esté enriquecida por fragmentos con actividad que aumente la expresión. Estará claro que una sub-genoteca tal estará comprendida dentro del alcance del término genoteca como se usa en el presente documento.

Los fragmentos de ADN que aumentan la expresión que se pueden obtener por el procedimiento de revisión de la invención, se pueden usar para incrementar el nivel de expresión de un casete de expresión en una molécula de ADN. Para este fin, un fragmento de ADN que aumenta la expresión obtenido se clona en un casete de expresión, que se usa para expresión recombinante, de acuerdo con los principios y procedimientos bien conocidos por la persona experta. En ciertas realizaciones por ejemplo, el fragmento de ADN que aumenta la expresión se clona a una distancia de menos de 2 kb de una unidad de transcripción que comprende un promotor unido operativamente a secuencias de las que se desea expresión, por ejemplo secuencias que codifican proteínas. El casete de expresión resultante se usa para expresión recombinante en células.

Un "casete de expresión" como se usa en el presente documento es una secuencia de ácido nucleico que comprende al menos un promotor unido funcionalmente a una secuencia de la que se desea expresión. Preferentemente, un casete de expresión contiene adicionalmente secuencias de finalización de la transcripción y de poliadenilación. Otras secuencias reguladoras tales como potenciadores se pueden incluir también.

El promotor debe ser capaz de funcionar en una célula huésped eucariota, es decir debe ser capaz de dirigir transcripción de la unidad de transcripción. El casete de expresión puede contener adicionalmente opcionalmente otros elementos conocidos en la técnica, por ejemplo sitios de ajuste para comprender intrones y similares. En algunas realizaciones, un intrón está presente detrás del promotor y antes de la secuencia que codifica el polipéptido marcador seleccionable.

Para obtener expresión de secuencias de ácidos nucleicos que codifican proteína, se conoce bien por aquellos expertos en la técnica que las secuencias capaces de dirigir tal expresión, pueden unirse funcionalmente a 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 general, la secuencia promotora está colocada anteriormente en la cadena con respecto a las secuencias que se expresarían. Muchos vectores de expresión usados están disponibles en la técnica, por ejemplo las series 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 secuencias terminadoras de transcripción, secuencias de poli-A y similares. La persona especializada en la técnica es consciente de que se pueden usar diversos promotores para obtener expresión de un gen en células huésped. Los promotores pueden ser constitutivos o regulados y se pueden obtener a partir de varias fuentes, incluyendo virus, fuentes procariotas o eucariotas, o pueden diseñarse artificialmente. La expresión de ácidos nucleicos de interés puede ser a partir del promotor natural o de derivado del mismo o a partir de un promotor enteramente heterólogo (Kaufman, 2000). Algunos promotores bien conocidos y usados mucho para 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 temprano inmediato (IE) del CMV (referido en el presente documento como el promotor CMV) (obtenible por ejemplo desde pcDNA, Invitrogen), promotores

derivados del Virus de los Simios 40 (SV40) (Das y cols., 1985) y similares. Los promotores adecuados se pueden derivar también de células eucariotas, tales como promotores de metalotioneína (MT), promotor del factor de elongación 1 α (EF-1 α) (Gill y cols., 2001), promotor de ubiquitina C o promotor UB6 (Gill y cols., 2001; Schorpp y cols., 1996), promotor de actina, un promotor de inmunoglobulinas, promotores de choque térmico y similares.

5 Algunos promotores preferidos para obtener expresión en células eucariotas, que son promotores adecuados en la presente invención, son el promotor CMV, un promotor EF1-alfa de mamíferos, un promotor de ubiquitina de mamíferos tal como promotor de ubiquitina C, o un promotor de SV40 (por ejemplo obtenible a partir de pIRES, n.º de cat. 631605, BD Sciences). La realización de pruebas para la función de un promotor y la fuerza de un promotor es una materia de rutina para una persona experta en la técnica y en general puede por ejemplo abarcar clonación de un gen de prueba tal como LacZ, luciferasa, GFP, etc. detrás de la secuencia promotora y prueba para la expresión del gen de prueba. Por supuesto, los promotores se pueden alterar por delección, adición, mutación de secuencias en ellos, y se pueden poner a prueba para funcionalidad, para encontrar secuencias promotoras nuevas, atenuadas o mejoradas. Se prefieren promotores fuertes que dan niveles de transcripción altos en las células eucariotas de elección. Un promotor heterólogo como se usa en el presente documento se define como un promotor que no es el promotor natural de dicha secuencia de interés. En otras palabras, algunas formas de intervención humana, por ejemplo clonación molecular, se han usado en cualquier momento para hacer la combinación funcional de un promotor heterólogo con un ácido nucleico de interés y se entiende fácilmente en este contexto que un promotor heterólogo se puede derivar del mismo organismo o de un organismo diferente que la secuencia de interés.

20 Una ejemplo de una señal adecuada para finalización de la transcripción que puede usarse es el sitio de poliadenilación de SV40 (Kaufman y Sharp, 1982), o la señal de poliadenilación de hormona del crecimiento bovina (patente de los Estados Unidos 5,122,458).

En otro aspecto, la presente invención proporciona así una molécula de ADN recombinante que comprende un fragmento de ADN que aumenta la expresión seleccionada del grupo constituido por: a) SEC. ID. N.º: 3 (47D); b) SEC. ID. N.º: 4 (44F); c) fragmentos a) o b), en los que dicho fragmento es al menos de 50 pares de bases en longitud y tiene actividad que aumenta la expresión; y d) secuencias que son al menos idénticas al 70% en la secuencia de nucleótidos a a), b) o c), en las que dichas secuencias tienen actividad que aumenta la expresión; comprendiendo dicha molécula de ADN recombinante un casete de expresión, comprendiendo dicho casete de expresión un promotor heterólogo unido a un ácido nucleico de interés.

30 Las secuencias 47D y 44F donde se recuperaron a partir del genoma humano usando revisiones de la presente invención y donde confirmaron subsiguientemente tener actividad que aumenta la expresión. Esta claro para la persona experta en la técnica que estos fragmentos podrían usarse como un punto de partida para preparar y poner a prueba fragmentos o derivados de los mismos, por ejemplo por delecciones, adiciones, sustituciones de bases, combinaciones de las mismas y similares y así tales fragmentos o derivados pueden abarcarse también dentro de la presente invención. Una secuencia tiene actividad que aumenta la expresión como se usa en el presente documento cuando una secuencia proporciona de media al menos expresión 10% más alta según se compara con la situación idéntica pero sin que esté presente dicha secuencia. En realizaciones preferidas, un fragmento de ADN que aumenta la expresión confiere una expresión más alta al menos el 20%, preferentemente al menos el 30%, más preferentemente al menos el 50%, aún más preferentemente al menos el 100%, o incluso más, según se compara con la situación donde dicho fragmento no está presente. Esto puede ponerse a prueba de acuerdo con procedimientos de rutina disponibles para la persona experta. Por ejemplo, esto puede ponerse a prueba adecuadamente en una construcción de expresión según se usa por el procedimiento de revisión revelado en el presente documento: un promotor de CMV unido operativamente a una unidad de transcripción bicistónica que comprende una secuencia codificante TTG Zeo (que codifica resistencia a zeocina y que tiene un codon de iniciación TTG) seguida por una secuencia que codifica una proteína fluorescente verde (d2EGFP) y una secuencia de finalización de la transcripción (véase por ejemplo ejemplo 1 a continuación). Un fragmento tiene actividad que aumenta la expresión cuando, si está situado anteriormente en la cadena con respecto al promotor de CMV en una construcción tal, la fluorescencia media en las colonias resistentes a zeozina es al menos el 10% más alta según se compara con la situación de control sin tal fragmento en condiciones aparte de eso idénticas. Las secuencias 47D y 44F tienen un tamaño de menos de 1 kb, lo que es ventajoso porque no requieren tanto espacio en un vector de expresión, y así queda más capacidad en ello para otras secuencias de interés. No obstante, incluso fragmentos más pequeños de estos fragmentos (por ejemplo fragmentos de entre aproximadamente 50-750 pares de bases, 100-700, 200-700, 300-700, 400-700, 500-700 pares de bases, etc., por ejemplo fragmentos de aproximadamente 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700 pares de bases) pueden conferir ya actividad que aumenta la expresión. También es posible alargar estos elementos añadiendo secuencias hacia atrás del genoma, por ejemplo un total de aproximadamente 1 kb, 1,5 kb, 2 kb y poner a prueba los fragmentos resultantes para actividad aumentadora de la expresión que posiblemente podría ser más alta que, pero también igual a, o menor que, la actividad de los elementos revelados. Es probable que la variación será posible en los fragmentos de ADN que aumentan la expresión, dando como resultado fragmentos derivados que tienen aún actividad que aumenta la expresión. Tales fragmentos menores y/o fragmentos derivados pueden prepararse y ponerse a prueba de forma rutinaria y podrían obtenerse fragmentos alterados con actividad disminuida, similar o incrementada. Preferentemente, dichas secuencias son al menos idénticas al 80%, más preferentemente idénticas al menos al 90% y aún más preferentemente idénticas al menos al 95% a la secuencia nativa de referencia o a fragmento funcional

de la misma. Para fragmentos de una secuencia dada, la identidad en porcentaje se refiere a aquella parte de la secuencia nativa de referencia que se encuentra en el fragmento. También puede ser posible encontrar fragmentos homólogos a fragmentos 47D o 44F, a partir de genomas de otros organismos, por ejemplo organismos mamíferos, por ejemplo mono, rata, ratón, hámster, etc. Tales fragmentos ortólogos o parálogos y derivados de los mismos están también comprendidos en la presente invención, cuando son al menos el 70% idénticos a una secuencia según se revela en el presente documento y cuando tienen actividad que aumenta la expresión según se define en el presente documento.

Un casete de expresión de la invención puede contener una unidad de transcripción monocistrónica o multicistrónica. Ello contiene un promotor y una secuencia de la que se desea expresión y usualmente una señal de terminación de la transcripción. En una realización, un casete de expresión de la invención contiene una unidad de transcripción monocistrónica que comprende el ácido nucleico de interés. En otras realizaciones, el ácido nucleico de interés está presente en una unidad de transcripción multicistrónica que codifica adicionalmente una proteína marcadora seleccionable. En ciertas realizaciones de la misma, la unidad de transcripción multicistrónica en la hebra codificante para la proteína marcadora seleccionable comprende una secuencia de inicio de traducción para la proteína marcadora seleccionable elegida a partir del grupo constituido por: a) un codon de iniciación ATG en un contexto no óptimo para la iniciación de la traducción, comprendiendo la secuencia (C/T)(A/T/G)(A/T/G)ATG(A/T/C) en el que el codon de iniciación está subrayado; b) un codon de iniciación GTG; c) un codon de iniciación TTG; d) un codon de iniciación CTG; e) un codon de iniciación ATT; y f) un codon de iniciación ACG. En tales realizaciones, la secuencia que codifica la proteína marcadora seleccionable puede situarse bien i) anteriormente en la cadena con respecto al ácido nucleico de interés, o bien ii) más adelante en la cadena con respecto al ácido nucleico de interés. En la situación i), la secuencia que codifica la proteína marcadora seleccionable está preferentemente desprovista de ATG internos (y es en el efecto por lo tanto como una secuencia en su totalidad desprovista de cualquier ATG, dado que el codon de iniciación es un GTG o un TTG), de tal manera que se expresa el ácido nucleico de interés (véase el documento WO 2006/048459 para este concepto y para procedimientos y ejemplos de crear ADN que codifica proteínas marcadoras seleccionables funcionales desprovistas de secuencias ATG). En la situación ii) el ácido nucleico de interés está anteriormente en la cadena y está preferentemente seguido por un sitio de entrada al ribosoma interno (IRES) que está operativamente unido a la secuencia que codifica la proteína marcadora seleccionable (que tiene el codon de iniciación GTG o TTG), de modo que la proteína marcadora seleccionable se traduce desde el IRES (véase el documento US 2006/0141577 para este concepto).

Como se usa en el presente documento, un "sitio de entrada a ribosoma interno" o "IRES" hace referencia a un elemento que promueve entrada directa al ribosoma para el codon de iniciación, tal como normalmente un ATG, pero en esta invención preferentemente GTG o TTG, de un cistrón (una región codificante de la proteína) que conduce por lo menos a la traducción independiente del casquete del gen. Las secuencias IRES y el uso de las mismas para expresión se conocen bien por la persona experta en la técnica, como se enseña en los documentos US 2006/0141577 y PCT/EP2007/051696, incorporados por referencia en el presente documento. Véase también, por ejemplo, Jackson R J, Howell M T, Kaminski A (1990) Trends Biochem Sci 15 (12): 477-83), Jackson R J y Kaminski, A. (1995) RNA 1 (10): 985-1000, Martínez-Salas, 1999, Venkatesan & Dasgupta, 2001, Rees y cols., 1996 y Mizuguchi y cols., 2000.

En ciertas realizaciones, el ácido nucleico de interés codifica todo o parte de un polipéptido de interés.

En ciertas realizaciones, los fragmentos de ADN que aumentan la expresión están situados anteriormente en la cadena con respecto al promotor en el casete de expresión.

En ciertas realizaciones, los fragmentos de ADN que aumentan la expresión están separados por menos de 5 kb, preferentemente menos de 2 kb, más preferentemente menos de 1 kb del promotor cuando se localizan anteriormente en la cadena con respecto al mismo, o de la señal de terminación de transcripción del casete de expresión cuando se localizan más adelante en la cadena con respecto a la misma. En las realizaciones preferidas, los fragmentos de ADN que aumentan la expresión están anteriormente en la cadena con respecto al promotor y separados por menos de 800, 700, 600, preferentemente por menos de aproximadamente 500, 400, 300, 200, 100 ó 50 pares de bases de (el extremo 5' de) la secuencia del promotor en el casete de expresión.

No se conoce si las secuencias que aumentan la expresión obtenidas se clasificarían también como secuencias STAR usando el ensayo según se describe en el documento WO 2006/048459. En cualquier caso, dado que se han obtenido por un procedimiento fundamentalmente diferente, se pueden combinar con otros elementos que influyen en la expresión, tales como elementos de control de cromatina como se discute en el documento WO 2006/048459 y después se pueden poner a prueba en su actividad con respecto a los niveles de expresión que en ciertos casos podría ser aditiva o que incluso podría ser sinérgica con al menos algunos de aquellos elementos de control de cromatina. Por lo tanto, en ciertas realizaciones, las moléculas de ADN proporcionadas por la presente invención además de un fragmento de ADN que aumenta la expresión de acuerdo con la invención comprenden adicionalmente un elemento de control de cromatina, tal como un elemento MAR, o un elemento STAR (véase el documento WO 2006/048459 para ejemplos de tales elementos y uso de los mismos). Incluso si ellos no fuesen aditivos o sinérgicos, los fragmentos de ADN que aumentan la expresión de la presente invención pueden reemplazar algunos de los elementos de cromatina y usarse como una alternativa a éstos, con la ventaja de que tienen un tamaño menor.

En ciertas realizaciones, una molécula de ADN de acuerdo con la invención es parte de un vector, por ejemplo de un plásmido. Tales vectores pueden manipularse fácilmente por procedimientos bien conocidos por la persona experta en la técnica y pueden por ejemplo estar diseñados para ser capaces de replicación en células procariotas y/o eucariotas. Además, muchos vectores pueden usarse directamente o en forma de fragmento deseado aislado de los mismos para transformación de las células eucariotas y se integrarán en su totalidad o en parte en el genoma de tales células, dando como resultado células huésped 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 de un genoma viral recombinante. El plásmido o el genoma vírico se introduce dentro de las células (huéspedes eucariotas) y se integra preferentemente en sus genomas por procedimientos conocidos en la técnica. En realizaciones preferidas, la presente invención usa también estos tipos de moléculas de ADN para suministrar su sistema de expresión de transgenes mejorado. Una realización preferida de la invención es el uso de ADN de plásmido para la administración del sistema de expresión.

El vector usado puede ser cualquier vector que sea adecuado para clonar ADN y que pueda usarse para transcripción de un ácido nucleico de interés. Cuando se usan células huésped se prefiere que el vector sea un vector de integración. Alternativamente, el vector puede ser un vector que se replique episomalmente.

Un ácido nucleico de interés en realizaciones preferidas codifica un polipéptido de interés. Un polipéptido de interés de acuerdo con 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. Ejemplos no limitantes de una proteína de interés de acuerdo con la invención son enzimas, hormonas, cadenas de inmunoglobulina, proteínas terapéuticas como proteínas anticancerígenas, proteínas de coagulación de la sangre tales como Factor VIII, proteínas multifuncionales, tales como eritropoyetina, proteínas diagnósticas o proteínas o fragmentos de las mismas útiles para propósitos de vacunación, todos conocidos por la persona experta en la técnica.

En determinadas 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 las mismas. En una realización se proporciona una unidad de expresión de proteína de acuerdo con la invención, en la que dicha proteína de interés es una cadena pesada de inmunoglobulina. En otra realización se proporciona una unidad de expresión de proteína de acuerdo con la invención, en la que dicha proteína de interés es una cadena ligera de inmunoglobulina. Cuando estas dos unidades de expresión de proteínas están presentes dentro de la misma célula (huésped), se ensambla una proteína multimérica y más específicamente una inmunoglobulina. Así, en ciertas realizaciones, la proteína de interés es una inmunoglobulina, tal como un anticuerpo, que es una proteína multimérica. Preferentemente, un anticuerpo tal es un anticuerpo humano o humanizado. En ciertas realizaciones de la misma, es un anticuerpo IgG, IgA, o IgM. Un inmunoglobulina puede estar codificada en sus cadenas pesada y ligera en diferentes casetes de expresión, o en un casete de expresión individual.

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 condensación o una proteína mutada) y preferentemente es una proteína humana.

Obviamente, las configuraciones de los casetes de expresión de la invención pueden usarse también cuando la meta definitiva no es la producción de un polipéptido de interés, sino el ARN por sí mismo, por ejemplo para producir cantidades incrementadas de ARN a partir de un casete de expresión, que pueden usarse para propósitos de regular otros genes (por ejemplo, ARNi, ARN antisentido), terapia génica, producción proteica *in vitro*, etc.

La invención proporciona adicionalmente células que comprenden una molécula de ADN recombinante de acuerdo con la presente invención. En ciertas realizaciones, dicha célula es una célula de mamífero, tal como una célula humana, una célula de hámster, una célula de ratón y similares. En ciertas realizaciones, dicha célula es una célula de CHO.

Las moléculas de ADN que comprenden unidades de transcripción multicistrónica y/o casetes de expresión de acuerdo con la invención se pueden usar para mejorar la expresión de ácido nucleico, preferentemente en células huésped. Los términos "célula"/"célula huésped" y "línea celular"/"línea celular huésped" se definen típicamente respectivamente como una célula y como poblaciones homogéneas de la misma que pueden mantenerse en cultivo celular por procedimientos conocidos en la técnica y que tiene la capacidad de expresar proteínas homólogas o heterólogas.

Las células huésped procariotas pueden usarse para propagar y/o llevar a cabo ingeniería genética con las moléculas de ADN de la invención, especialmente cuando estaban presentes en plásmidos capaces de replicar en células huésped procariotas tales como bacterias.

Una célula huésped de acuerdo con la presente invención preferentemente es una célula eucariota, más preferentemente una célula de mamífero, tal como una célula de roedor o una célula humana o fusión entre células diferentes. En ciertas realizaciones no limitantes, dicha célula hospedadora es una célula de osteosarcoma U-2 OS,

de CHO (ovario de hámster chino), de HEK 293, de mieloma HuNS-1, de retinoblastoma WERI-Rb-1, de BHK, de COS, de Vero, de mieloma de ratón no secretor Sp2/0-Ag 14, de mieloma de ratón no secretor NS0, carcinomal de glándula adrenal NCI-H295R o una célula de PER.C6[®]. Las células de PER.C6 para el propósito de la presente invención significan células de un paso anteriormente en el procedimiento o posteriormente en el procedimiento o de un descendiente de un paso anteriormente en el procedimiento o posteriormente en el procedimiento de células según se depositan en ECACC n.º 96022940 (documento US 5,994,128), es decir que tienen las características de esas células. Se ha mostrado previamente que tales células son capaces de expresión de proteínas a niveles altos (por ejemplo documento WO 00/63403 y Jones y cols., 2003). En ciertas realizaciones preferidas, las células huésped son células de CHO, por ejemplo CHO-K1, CHO-S, CHO-DG44, CHO-DUKXB11 y similares. En ciertas realizaciones, dichas células de CHO tienen un fenotipo dhfr⁻.

Tales células eucariotas pueden expresar polipéptidos deseados y se usan a menudo para ese propósito. Ellos pueden obtenerse por introducción de una molécula de ADN de la invención, preferentemente en forma de un casete de expresión, dentro de las células. Preferentemente, el casete de expresión está integrada en el genoma de las células huésped, lo que puede ser en diferentes posiciones en diversas células huésped, y la selección mantendrá un clon donde el transgén está integrado en una posición adecuada, conduciendo a un clon de célula huésped con propiedades deseadas en términos de niveles de expresión, estabilidad, características de crecimiento y similares. Alternativamente la unidad de transcripción puede estar dirigida o puede seleccionarse al azar para integración dentro de una región cromosómica que está transcripcionalmente activa, por ejemplo detrás de un promotor presente en el genoma.

Preferentemente las células huésped son de un clon estable que se puede seleccionar y propagar de acuerdo con procedimientos estándar conocidos por la persona experta en la técnica. Un cultivo de un clon tal es capaz de producir polipéptido de interés, si las células comprenden la unidad de transcripción que lo codifica. Las células de acuerdo con la invención preferentemente son capaces de crecer en cultivo en suspensión en medio libre de suero.

En realizaciones preferidas, la molécula de ADN de la invención está integrada dentro del genoma de la célula huésped eucariota de acuerdo con la invención. Esto mantendrá herencia estable en la molécula de ADN.

La invención proporcionará adicionalmente un procedimiento para producir un polipéptido de interés, que comprende cultivar una célula comprendiendo una molécula de ADN recombinante de acuerdo con la invención, para expresar el ácido nucleico que codifica la proteína de interés en dicha célula. En las realizaciones preferidas, la proteína de interés se recogió de dicha célula o del medio de cultivo o de ambos. En realizaciones preferidas, dicha célula es una célula de mamífero, por ejemplo una célula de CHO. En realizaciones preferidas de este procedimiento, la molécula de ADN recombinante de la invención está integrada dentro del genoma de la célula. En ciertas realizaciones, el ácido nucleico que codifica la proteína de interés está presente en una unidad de transcripción multicistrónica. En determinadas realizaciones de las mismas, una unidad de transcripción multicistrónica tal es una de acuerdo con la revelación del documento WO 2006/048459 o del documento US 2006/0141577.

La introducción de ácido nucleico que está para expresarse en una célula, se puede hacer por uno de varios procedimientos, que como tales se conocen por la persona experta en la técnica, también son dependientes del formato del ácido nucleico a introducirse. Dichos procedimientos incluyen pero no se limitan a transfección, infección, inyección, transformación y similares.

En ciertas realizaciones, el agente de selección está presente en el medio de cultivo al menos parte del tiempo durante el cultivo, bien en concentraciones suficientes para seleccionar las células que expresan el polipéptido marcador seleccionable o bien en concentraciones más bajas. En realizaciones preferidas, el agente de selección no estará presente más tiempo en el medio de cultivo durante la fase de producción cuando se expresa el polipéptido.

Cultivar una célula se hace para permitirle metabolizar, y/o crecer y/o dividirse y/o producir proteínas recombinantes de interés. Esto puede lograrse por procedimientos bien conocidos por personas expertas en la técnica e incluye pero no se limita a proporcionar nutrientes para la célula. Los procedimientos comprenden crecimiento adhiriéndose a superficies, crecimiento en suspensión, o combinaciones de los mismos. Cultivar puede hacerse por ejemplo en platos, botellas enrolladas o en biorreactores, usando sistemas por lotes, sistemas semicontinuos, sistemas continuos tales como sistemas de perfusión y similares. Con el fin de conseguir producción a gran escala (continua) de proteínas recombinantes a través de la célula de cultivo se prefiere en la técnica tener células capaces de crecer en suspensión, y se prefiere tener células capaces de cultivarse en ausencia de suero derivado de animal o ser humano o de componentes de suero derivados de animal o de ser humano.

Las condiciones para cultivar o multiplicar células (véase por ejemplo Tissue Culture, Academic Press, Kruse and Paterson, editores (1973)) y las condiciones para expresión del producto recombinante se conocen por la persona experta en la técnica. En general, los principios, protocolos y técnicas prácticas para maximizar la productividad de cultivos celulares de mamíferos se pueden encontrar en Mammalian Cell Biotechnology: a Practical Approach (M. Butler, ed., IRL Press, 1991).

En una realización preferida, la proteína expresada se recoge (se aísla), bien de las células o bien del medio de cultivo o bien de ambos. Puede entonces purificarse adicionalmente usando procedimientos conocidos, por ejemplo

filtración, cromatografía en columna, etc., por procedimientos generalmente conocidos por la persona experta en la técnica.

La práctica de esta invención empleará, a menos que se indique lo contrario, técnicas convencionales de inmunología, biología molecular, microbiología, biología celular y de ADN recombinante, que están dentro de la capacidad de la técnica. Véase por ejemplo Sambrook, Fritsch y Maniatis, *Molecular Cloning: A Laboratory Manual*, 2ª edición, 1989; *Current Protocols in Molecular Biology*, Ausubel FM, y cols., 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, eds., 1988.

La invención se explica adicionalmente en el siguiente ejemplo. El ejemplo no limita la invención de ninguna manera. Sirve meramente para aclarar la invención.

Ejemplo

El Ejemplo 1 describe la revisión en busca de fragmentos de ADN que aumentan la expresión de la presente invención y está claro que las variaciones descritas en los ejemplos 1-19 del documento WO 2006/048459 pueden aplicarse y ponerse a prueba para la revisión para los fragmentos de ADN que aumentan la expresión y que los fragmentos de ADN que aumentan la expresión obtenidos podrían estar incorporados en casetes de expresión, incluidos aquellos que contienen unidades de transcripción multicistrónicas del documento WO 2006/048459 y del documento US 2006/0141577.

Ejemplo 1: Uso del sistema STAR-Select identificando fragmentos de ADN que aumentan la expresión

El sistema de selección del documento WO 2006/048459 en el que un codón de iniciación GTG o TTG se usa para el ARNm de proteína marcadora seleccionable es muy restrictivo. La aplicación del sistema de selección no da resultados en casi ninguna colonia, a menos que los elementos STAR se incorporen en la construcción potenciando el promotor CMV (véase el documento WO 2006/048459). Este sistema puede en principio usarse para una revisión sistemática hallando fragmentos de ADN genómicos, potencialmente distintos de los elementos STAR, que potencian el promotor CMV de una manera tal que las colonias sobreviven fácilmente en una presión de selección dada. Los autores de la presente invención describen aquí una búsqueda sistemática de tales fragmentos de ADN que aumentan la expresión.

Resultados

Se cortaron aproximadamente 10 µg de ADN genómico humano (Promega, G147A 16040303) con la enzima de restricción Sau3AI durante 10 minutos a 37°C obteniendo fragmentos al azar entre 200 y 1000 pares de bases. El ADN digerido se corrió en un gel de agarosa al 1,2% y la mancha que contenía fragmentos entre 200 y 1000 pares de bases se escindió y el ADN se purificó y se aisló. Esta reserva de fragmentos de ADN se clonó en el sitio de restricción de BglIII anteriormente en la cadena con respecto al promotor de CMV, que estaba seguido por los casetes de TTG Zeo y de d2EGFP. La mezcla de ligazón se transformó a *E. coli* XL10 Gold (Stratagene). Aproximadamente se almacenaron 20.000 colonias, se cultivaron en medio de caldo líquido y el ADN se aisló usando una columna de maxiprep Qiagen. Aproximadamente se aislaron 300 µg de ADN. Por aproximadamente 700.000 células CHO-K1 en un pocillo de una placa de 6 pocillos, se transfectaron 3 µg de ADN usando Lipofectamina 2000 (Invitrogen). Este procedimiento se llevó a cabo para 36 pocillos. Como control, se transfectaron 12 pocillos con una construcción, que contenía el promotor CMV, TTG Zeo y d2EGFP, pero ningún ADN genómico insertado. Tanto células que se transfectaron con plásmidos de control como células que se transfectaron con la genoteca se cultivaron en presencia de 100 µg/ml de Zeocina en medio HAM-F12 (Invitrogen) + FBS al 10% (Invitrogen). Las células se dejaron crecer durante tres semanas, momento en el que se formaron algunas colonias. Fuera del plásmido control transfectado en 12 pocillos, se formaron aproximadamente 5 colonias. Sin embargo, la inspección en un microscopio de fluorescencia reveló que estas colonias no fueron verdes, implicando que produjeron sólo poco d2EGFP. En cambio, se formaron 72 colonias como resultado de las transfecciones con la genoteca. Todas las 72 colonias se aislaron y propagaron, antes de medir los niveles de expresión de d2EGFP usando un citómetro de flujo (Beckman Coulter). De las 25 colonias de CHO de expresión más alta, se aisló ADN genómico y se analizó por PCR, para la presencia de fragmentos de genoteca. Para la PCR se usaron los siguientes cebadores, que reconocieron las secuencias flanqueantes de los sitios de clonación.

Cebador A: gatcggcgcgcccgaaggccgtaccttaattaag (SEQ ID NO: 1)

Cebador B: agcggcgcgcccgcgaattaggcaaggaattatcag (SEQ ID NO: 2).

Por clon de CHO original se observaron productos de PCR de diversas longitudes, indicando que fragmentos de ADN múltiples de la genoteca se habían integrado en tal clon de CHO original. Para identificar fragmentos de ADN que aumentan expresión potenciales, se clonaron los productos de PCR dentro del sitio *Ascl* del vector TTG Zeo d2EGFP. La mezcla de ligazón se transformó a *E. coli* y se analizaron hasta 10 colonias bacterianas diferentes para insertos. Cuando se encontraron más de cinco insertos de ADN con diferentes tamaños por clon de CHO original, el análisis no se llevó a cabo. Seis clones de CHO originales albergaron menos de cinco insertos de ADN, indicando una complejidad limitada. De estos seis clones de CHO originales, se recuperaron en total 16 construcciones que

5 contenían fragmentos de ADN, anteriormente en la cadena con respecto a TTG Zeo. Estas 16 construcciones se retransfectaron a CHO-K1. La transfección de una construcción (construcción 47D; SEQ ID NO: 3) dio como resultado la generación de más de 25 clones por transfección que mostraron niveles elevados de expresión de d2EGFP en un microscopio de fluorescencia. La transfección de una segunda construcción (44F; SEQ ID NO: 4) dio como resultado la generación de 7 clones que mostraron expresión de d2EGFP más alta en un microscopio de fluorescencia. Las otras 14 construcciones cualquiera dieron niveles de expresión de d2EGFP muy bajos según se supervisó en un microscopio de fluorescencia y no se continuaron adicionalmente.

10 El análisis de los 7 clones elegidos que se transfectaron con construcción 44F, mostró que los niveles de expresión de d2EGFP estaban significativamente elevados, cuando se compararon con las colonias control (Fig. 1). Además el análisis de 12 clones elegidos al azar que se transfectaron con la construcción 47D, mostró que los niveles de expresión de d2EGFP estaban significativamente elevados, cuando se compararon con las colonias control (Fig. 1). Como control positivo los autores de la presente invención compararon una construcción TTG Zeo d2EGFP, flanqueada con elementos STAR 7 y 67 anteriormente en la cadena y con STAR 7 posteriormente en la cadena. Según se compara con el STAR 7/67/7 el elemento 47D indujo valores de d2EGFP comparables, aunque ligeramente menos altos. Dado que la construcción 47D indujo los valores de d2EGFP más altos (Fig. 1), los autores de la presente invención continuaron el análisis de esta construcción en el contexto de otra línea celular de CHO, CHO-DG44.

20 La construcción que contiene el elemento 47D se transfectó también a células CHO-DG44. La construcción se transfectó usando Lipofectamina 2000 (Invitrogen). Las células se cultivaron en presencia de Zeocina a 100 µg/ml en medio HAM-F12 (Invitrogen) + FBS al 10% (Invitrogen) y DMEM (1:1). Las células se dejaron crecer durante tres semanas, tiempo en el que se formaron las colonias. El análisis de 18 clones elegidos al azar que se transfectaron con construcción 47D, demostró que los niveles de expresión de d2EGFP estaban significativamente elevados, cuando se compararon con las colonias transfectadas de CHO-DG44 controles (Fig. 2). Como control positivo los autores de la presente invención compararon una construcción TTG Zeo d2EGFP, flanqueada con elementos STAR 45 y 67 anteriormente en la cadena y con STAR 45 posteriormente en la cadena. Según se comparó con el STAR 45/67/45 el elemento 47D novedoso indujo valores de d2EGFP comparables, aunque ligeramente menos altos.

30 Estos resultados muestran que es posible aislar elementos de ADN que aumentan la expresión empleando el marcador de selección TTG Zeo restrictivo como metodología de revisión. La persona experta entenderá que los vectores usados para revisión y en particular el ADN que codifica la proteína marcadora seleccionable, pueden variarse a lo largo de las líneas discutidas en los ejemplos del documento WO 2006/048459.

Referencias

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*. *Célula* 74: 505-514.

35 Chung JH, Timbre AC, Felsenfeld G. (1997). Characterization of the chicken beta-globin insulator. *Proc Natl Acad Sci EE.UU.* 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.

40 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.

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.

Kaufman, R.J. (2000) Overview of vector design for mammalian gene expression *Mol Biotechnol* 16, 151-160.

45 Kaufman, R.J, 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.

50 Kozak M. (1986) Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell* 44: 283-292.

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.

- Kozak M. (1990) Downstream secondary structure facilitates recognition of initiator codons by eukaryotic ribosomes. *Proc Natl Acad Sci EE.UU.* 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.
- 5 Kozak M. (2002) Pushing the limits of the scanning mechanism for initiation of translation. *Gen* 299: 1-34.
- 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).
- 10 Phi-Van L, Von Kreis JP, Ostertag W, y Strätling 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.
- Martínez-Salas, E. (1999) Internal ribosome entry site biology and its use in expression vectors *Curr Opin Biotechnol* 10, 458-64.
- 15 Mizuguchi, H, Xu, Z, Ishii-Watabe, A, Uchida, E, y Hayakawa, T. (2000) IRES-dependent second gene expression is significantly lower than cap-dependent first gene expression in a bicistronic vector *Mol Ther* 1, 376-82.
- 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.
- 20 Schorpp, M, Jager, R, Schellander, K, Schenkel, J, Wagner, EF, Weiher, H, y Ángel, P. (1996) The human ubiquitin C promoter directs high ubiquitous expression of transgenes in mice *Nucleic Acids Res* 24, 1787-8.
- Venkatesan, A, y Dasgupta, A. (2001) Novel fluorescence-based screen to identify small synthetic internal ribosome entry site elements *Mol Cell Biol* 21, 2826-37.
- West AG, Gaszner M, Felsenfeld G (2002) Insulators: many functions, many mechanisms. *Genes Dev.* 16: 271-288.
- 25

Listado de secuencias

<110> ChromaGenics B.V.

Otte, Arie P.

Kwaks, Theodorus H.J.

5 Sewalt, Richard G.A.B.

van Blokland, Rik

<120> Fragmentos de ADN que aumentan la expresión, uso de los mismos y procedimientos para encontrar los mismos

10

<130> 0129 WO 00 ORD

<150> EP 06111381

<151> 20-3-2006

15

<150> US 60/784,133

<151> 20-3-2006

<160> 4

20

<170> PatentIn version 3.3

<210> 1

<211> 38

25 <212> ADN

<213> Artificial

<220>

<223> cebador A (para obtener fragmentos de ADN que aumentan la expresión a partir de
30 revisión)

<400> 1

gatcggcgcg cccgaaaggg cccgtacctt aattaaag

38

35 <210> 2

<211> 39

ES 2 372 703 T3

<212> ADN

<213> Artificial

<220>

5 <223> cebador B (para obtener fragmentos de ADN que aumentan la expresión a partir de
revisión)

<400> 2

aggcggcgcg cccgcgaaat taggcaaagg aattatcag

39

10 <210> 3

<211> 656

<212> ADN

<213> Homo sapiens

15 <220>

<221> fragmento de ADN que aumenta la expresión 47D

<222> (1)..(656)

<400> 3

gatctaataga tgaaaataga caaagtggca ggatggaaaa agacactcca tgcaaattgtt 60

aaccaaacga gaggtggaga gctaattata ttaagtcaaa agctgtcata ttttataaaa 120

tatagttaa gtcaaaactc acaagagaca aacaaggaca ttatagtata ataaaatggt 180

20 tcattcactg ggaacctatt tatatctgct accaggggtc ccaaatatth aaagcaagca 240

ttgacagaaa tgaatataac aatgggagag tacttcaata ctttcagtaa tgaataataa 300

agcatgacag aacattaata agggaacaaa gaacttgaaa gcagtgtaaa gcaattacac 360

ctaaaagacg tatacaggac acaccacaca acagcagaat ccacaaactt ttcgaaagct 420

cataagaaaa ttctcctgga tacaacacct tttatctcac aagacaagtc ttaatccatg 480

tttaaactgg aatcttacgc attattatth ttggccaaat tgtgtgaaac taaaaatcac 540

taacaacagg aaaacaaaa aaacacaaaa tatgaaaatt aaacaataca ctcttaagca 600

tgctcttggt caaaggttgg aagactgtga agatggccat actgaccaa ttgatc 656

<210> 4

<211> 729

25 <212> ADN

ES 2 372 703 T3

<213> Homo sapiens

<220>

<221> fragmento de ADN que aumenta la expresión 44F

5 <222> (1)..(729)

<400> 4

```
gatccctaat gctgacccaa ttaaatttgt ggttttggca tagcaagggtt tattgattga      60
tcaatgctgt gtgtatgggtg gatatgcttg ccttgtctca cagaaaggca aacaagcagc      120
ccattatatt aaagcatgtg gattacgttt tagtccctat ggaaaaataa tataacttcc      180
ttgttcttac tccaatcttt aaggatctca aataagagaa aaataattgc ataattcatc      240
gtcttttgtg atttgagaca tacatgatca tgttattggc agaaaatgaa ttggttttat      300
acattttggt gataagtgtt cccaaacatt tgtaattagt tgaccatatt cttgtttctc      360
caccttcaga agaactgggtt attatgattg ctttttagga aggtggtaat tacaaaatag      420
aatgagctg  aatgaaatta aagagaaact gaaaaaagaa attccttgcg aggaaataat      480
ttcatttgaa cgagtacaaa gttctttgca attcttgctt agctgaagtg tatttctagt      540
tgtgtttcca gatttatatc attatgaaaa tagccacttc ctaggttctc atttttcata      600
tacaatgttc actaggaata aattttctcc ctggattcaa tttctgaata ttcacaaggt      660
gaaaattgtg ttcacctctt ttcagatgga gattctctga actgtaattg ggtgactatt      720
tcatagatc                                     729
```

10

REIVINDICACIONES

1. Una molécula de ADN recombinante que comprende un fragmento de ADN que aumenta la expresión seleccionado del grupo constituido por:
 - a) SEQ ID NO 3;
 - 5 b) SEQ ID NO 4;
 - c) fragmentos de a) o b), en los que dichos fragmento tienen actividad que aumenta la expresión; y
 - d) secuencias que son al menos un 70% idénticas en secuencia de nucleótidos a a), b) o c), en las que dichas secuencias tienen actividad que aumenta la expresión;
- 10 comprendiendo adicionalmente dicha molécula de ADN recombinante un casete de expresión, comprendiendo dicho casete de expresión un promotor heterólogo unido a un ácido nucleico de interés.
2. Una molécula de ADN recombinante de acuerdo con la reivindicación 1, en la que dicho ácido nucleico de interés codifica todo o parte de una proteína de interés.
3. Una molécula de ADN recombinante de acuerdo con la reivindicación 1 ó 2, en la que dicho fragmento de ADN que aumenta la expresión está situado anteriormente en la cadena con respecto a dicho promotor en dicho casete de expresión.
- 15 4. Una molécula de ADN recombinante de acuerdo con la reivindicación 3, en la que dicho fragmento de ADN que aumenta la expresión y dicho promotor están separados por menos de 2 kb.
5. Una molécula de ADN recombinante de acuerdo con una cualquiera de las reivindicaciones 1-4, en la que dicho ácido nucleico de interés está presente en una unidad de transcripción multicistónica que codifica adicionalmente una proteína marcadora seleccionable.
- 20 6. Una molécula de ADN recombinante de acuerdo con la reivindicación 5, en la que la unidad de transcripción que codifica la proteína marcadora seleccionable en la hebra codificante comprende una secuencia de iniciación de traducción para la proteína marcadora seleccionable elegida del grupo constituido por: a) un codon de iniciación ATG en un contexto no óptimo para iniciación de traducción, que comprende la secuencia (C/T)(A/T/G)(A/T/G)ATG(A/T/C) en la que el codon de iniciación está subrayado; b) un codon de iniciación GTG; c) un codon de iniciación TTG; d) un codon de iniciación CTG; e) un codon de iniciación ATT; y f) un codon de iniciación ACG.
7. Una célula que comprende una molécula de ADN recombinante de acuerdo con una cualquiera de las reivindicaciones 1-6.
- 30 8. Una célula de acuerdo con la reivindicación 7, que es una célula de mamífero.
9. Una célula de acuerdo con la reivindicación 7 u 8, que es una célula de CHO.
10. Un procedimiento para producir una proteína de interés, que comprende cultivar una célula que comprende una molécula de ADN recombinante de acuerdo con una cualquiera de las reivindicaciones 2-6 para expresar dicho ácido nucleico que codifica la proteína de interés en dicha célula.
- 35 11. Un procedimiento de acuerdo con la reivindicación 10, que comprende adicionalmente cosechar la proteína de interés.
12. Un procedimiento de acuerdo con la reivindicación 10 u 11, en el que dicha célula es una célula de mamífero.
13. Un procedimiento de acuerdo con una cualquiera de las reivindicaciones 10-12, en el que dicha célula es una célula de CHO.
- 40 14. Un procedimiento de acuerdo con una cualquiera de las reivindicaciones 10-13, en el que dicha molécula de ADN recombinante está integrada en el genoma de dicha célula.
15. Un procedimiento de obtención de un fragmento de ADN que aumente la expresión comprendiendo el procedimiento las etapas de:
 - 45 1) proporcionar una diversidad de vectores que comprenden fragmentos, teniendo dichos vectores que comprenden fragmentos un tamaño de entre aproximadamente 50 y 5000 pares de bases, estando localizados dichos fragmentos a una distancia de menos de aproximadamente 5000 pares de bases de una unidad de transcripción, comprendiendo dicha unidad de transcripción un promotor unido operativamente a secuencia que codifica una proteína marcadora seleccionable que protege una célula huésped de los efectos letales o

- inhibidores de crecimiento de un agente de selección, en la que la secuencia que codifica la proteína marcadora seleccionable tiene un codon de iniciación GTG o un codon de iniciación TTG;
- 2) introducir los vectores que comprenden fragmentos desde la genoteca en las células huésped;
- 3) cultivar las células huésped en presencia del agente de selección; y
- 5 4) obtener el fragmento de ADN que aumenta la expresión a partir de células huésped que aún pueden crecer en presencia del agente de selección después de al menos dos semanas.
16. Un procedimiento de acuerdo con la reivindicación 15, en el que el agente de selección es zeocina.
17. Un procedimiento de acuerdo con la reivindicación 15 ó 16, en el que la secuencia que codifica la proteína marcadora seleccionable tiene un codon de iniciación TTG.
- 10 18. Un procedimiento de acuerdo con una cualquiera de las reivindicaciones 15-17, en el que la unidad de transcripción es una unidad de transcripción multicistrónica, que comprende la secuencia que codifica una proteína marcadora seleccionable que protege una célula huésped de los efectos letales o inhibidores del crecimiento de un agente de selección y una secuencia que codifica una proteína de la que se puede detectar la presencia.
- 15 19. Un procedimiento de acuerdo con la reivindicación 18, en el que la unidad de transcripción multicistrónica comprende en el siguiente orden:
- a) el promotor;
- b) la secuencia que codifica una proteína marcadora seleccionable que protege una célula huésped de los efectos letales o inhibidores de crecimiento de un agente de selección, en la que dicha secuencia está desprovista de secuencias ATG; y
- 20 c) la secuencia que codifica una proteína de la que se puede detectar la presencia.
20. Un procedimiento de acuerdo con la reivindicación 18, en el que la unidad de transcripción multicistrónica comprende en el siguiente orden:
- a) el promotor;
- b) la secuencia que codifica una proteína de la que se puede detectar la presencia; y
- 25 c) un sitio de entrada en ribosoma interno (IRES), unido operativamente a
- d) la secuencia que codifica una proteína marcadora seleccionable que protege una célula huésped de los efectos letales o inhibidores del crecimiento de un agente de selección.
21. Un procedimiento de acuerdo con una cualquiera de las reivindicaciones 15-20, en el que vectores que comprenden fragmentos se integran en el genoma de las células huésped.
- 30 22. Una genoteca que comprende una diversidad de vectores que comprenden fragmentos, comprendiendo dichos vectores fragmentos de ADN que tienen un tamaño de entre aproximadamente 50 y 5000 pares de bases, estando localizados dichos fragmentos de ADN a una distancia de menos de aproximadamente 5000 pares de bases de una unidad de transcripción, comprendiendo dicha unidad de transcripción un promotor unido operativamente a
- 35 una secuencia que codifica una proteína marcadora seleccionable que protege una célula hospedadora de los efectos letales o inhibidores del crecimiento de un agente de selección, en la que la secuencia que codifica la proteína marcadora seleccionable tiene un codon de iniciación GTG o un codon de iniciación TTG.

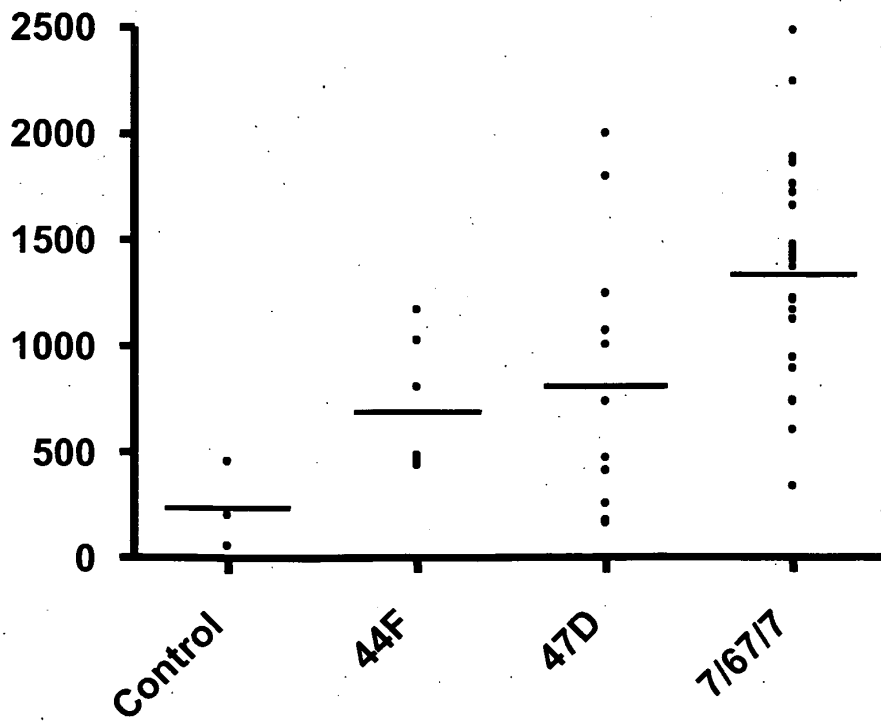
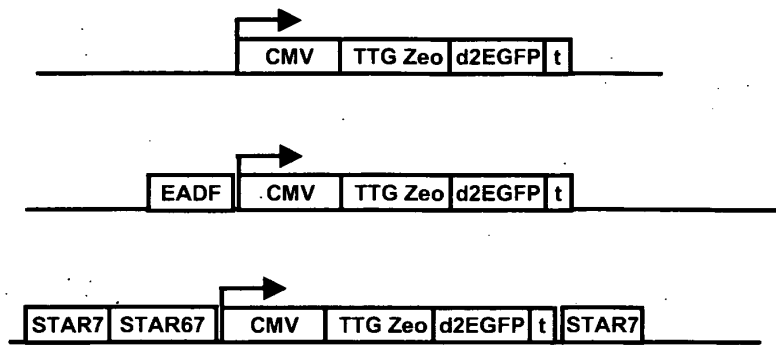


FIG. 1

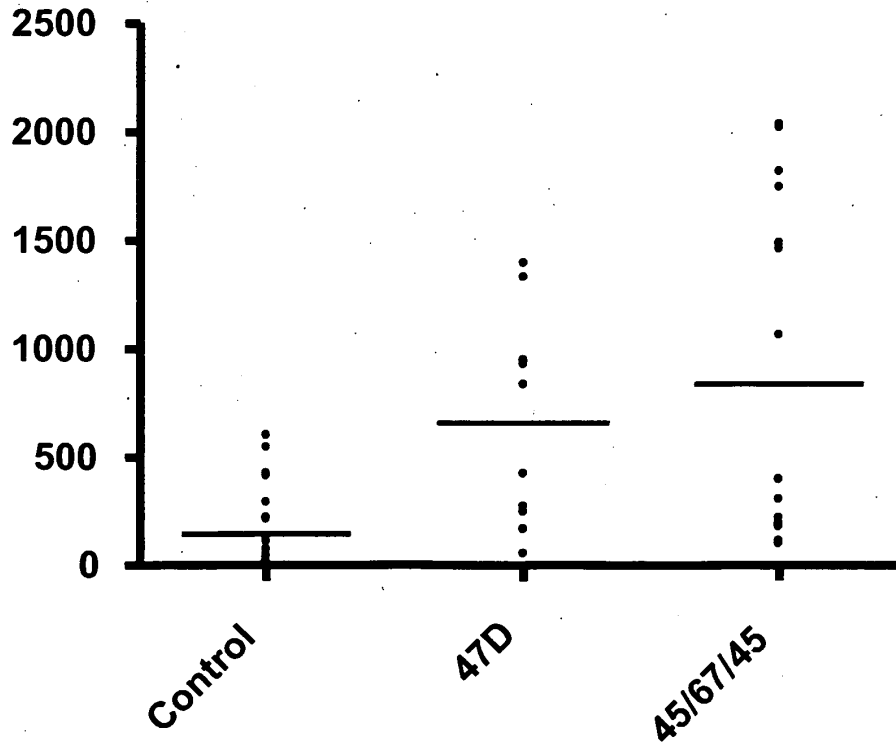
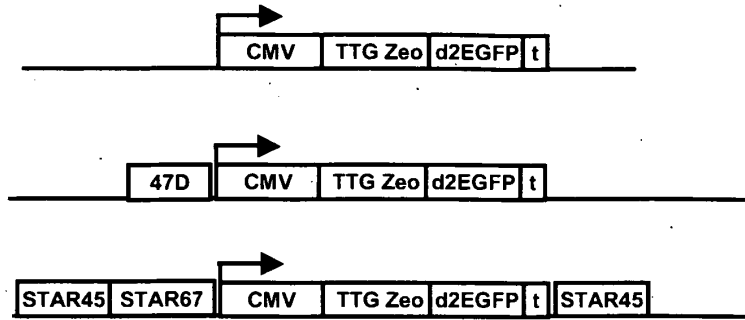


FIG. 2