

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 388 199**

51 Int. Cl.:
C12N 15/10 (2006.01)
C12N 15/52 (2006.01)
C12N 15/62 (2006.01)
C12N 15/63 (2006.01)
C12N 5/10 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Número de solicitud europea: **08736881 .7**
 96 Fecha de presentación: **04.04.2008**
 97 Número de publicación de la solicitud: **2155875**
 97 Fecha de publicación de la solicitud: **24.02.2010**

54 Título: **Ácidos nucleicos y bibliotecas para uso en el análisis funcional de RNAs reguladores tales como microRNAs (miRNAs)**

30 Prioridad:
04.04.2007 GB 0706631

45 Fecha de publicación de la mención BOPI:
10.10.2012

45 Fecha de la publicación del folleto de la patente:
10.10.2012

73 Titular/es:
**KING'S COLLEGE LONDON
STRAND
LONDON WC2R 2LS, GB**

72 Inventor/es:
**GAKEN, Johannes Adrianus y
MOHAMEDALI, Azim**

74 Agente/Representante:
de Elzaburu Márquez, Alberto

ES 2 388 199 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

Ácidos nucleicos y bibliotecas para uso en el análisis funcional de RNAs reguladores tales como microRNAs (miRNAs)

Campo de la Invención

- 5 La invención se refiere a materiales tales como ácidos nucleicos y bibliotecas para uso en el análisis funcional de RNAs reguladores tales como microRNAs (miRNAs), y particularmente al test de o cribado de dianas de RNAs reguladores tales como secuencias de regiones no traducidas (UTR) 3.

Antecedentes de la Invención

10 Los microRNAs (miRNAs) están reconocidos actualmente como una nueva clase de pequeñas moléculas de RNA reguladoras que regulan la expresión de muchos genes. Se ha demostrado que los mismos median la angiogénesis, la adhesión celular, la proliferación celular y la supervivencia, y juegan un papel importante en la hematopoyesis. Son producidos a partir de transcritos primarios de RNA (pri-miRNAs) que son procesados por la enzima DROSHA en dúplex de aproximadamente 70 pares de bases que son procesados ulteriormente por DICER en dúplex de miRNA de 22 pares de bases. Una cadena del dúplex de 22 pares de bases se asocia con el complejo de silenciación inducido por RNA (RISC) que está direccionado a sitios dentro de la región no traducida (UTR) 3' del mRNA dando como resultado represión de la traducción, escisión del mRNA o inducción de desadenilación. Actualmente se cree que en los humanos, el complejo RISC actúa principalmente por inducción de inhibición específica de la traducción por la fijación a la UTR 3' del mRNA diana y en menor proporción degradación de las dianas de mRNA.

20 Los microRNAs (miRNAs) son una familia de pequeños RNAs maduros no codificantes de 21 a 25 nucleótidos de longitud. Los mismos regulan negativamente la expresión de los genes codificantes de proteínas. Los miRNAs son procesados secuencialmente a partir de transcritos precursores primarios de miRNA (pri-miRNA), y regulan la expresión génica a nivel posterior a la transcripción. La expresión de los miRNAs es altamente específica para la etapa tisular y del desarrollo, pero se sabe poco acerca del modo en que se regulan estos patrones de expresión. Se han identificado más de 541 genes miRNA humanos, pero los métodos bioinformáticos recientes predicen que el número está más próximo a 1000. Las estimaciones actuales sugieren que aproximadamente una tercera parte de los RNAs humanos parecen ser dianas de miRNA. Se ha demostrado que los mismos median la angiogénesis, la adhesión celular, la proliferación celular, la supervivencia y juegan un papel importante en la hematopoyesis y el cáncer.

30 Debido a la homología parcial entre un miRNA y su diana, y a la inhibición de la traducción en lugar de la degradación del mRNA, la identificación de las dianas es una tarea difícil. Se han desarrollado algoritmos bioinformáticos para la predicción de dianas de miRNA basados en la secuencia de "siembra". Los cuatro algoritmos principales predicen 101.031 pares miRNA/diana (como promedio 200 dianas por miRNA). Únicamente 0,01% (12) de estos pares son predichos por los 4 algoritmos, 2,8% por 3, 15,4% por 2 y 81,8% por un solo algoritmo. De los 465 miRNAs humanos identificados, sólo 57 tienen 103 sitios diana validados experimentalmente en 85 genes.

40 Hasta la fecha, el papel y las dianas específicas de la mayoría de los miRNAs son desconocidos en gran parte. Esto es debido principalmente a las dificultades en la identificación de dianas debido a que, contrariamente al RNA de interferencia corto (siRNA), la fijación de miRNAs debida sólo parcialmente a homología con la diana. Adicionalmente, la inhibición de la traducción impide estudios de redes de expresión de mRNA para el descubrimiento de dianas.

45 A fin de lograr una mejor comprensión acerca de la función de los miRNAs, se han dedicado grandes esfuerzos a la identificación por computadora de dianas de miRNA utilizando diversos algoritmos (v.g., miRBase (Instituto Sanger, <http://microrna.sanger.ac.uk/sequences/>), TargetScan (Whitehead Institute for Biomedical Research, <http://www.targetscan.org/>) and PicTar (New York University, <http://pictar.bio.nyu.edu/>)). Sin embargo, los inconvenientes de estas predicciones son que las mismas generan un número sustancial de positivos falsos. Adicionalmente, las predicciones tienen gran probabilidad de estar inherentemente sesgadas, dado que se basan principalmente en el conocimiento obtenido a partir del muy pequeño número conocido de interacciones miRNA:diana, un tamaño de muestra estadísticamente muy pequeño, que conduce casi con certeza a un sesgo en las predicciones.

50 El estudio en la técnica anterior de la regulación de los genes de miRNA carece de los instrumentos necesarios para identificación y validación de las dianas, particularmente en lo que respecta a estudios funcionales.

55 Es sabido que los siRNAs tienen efectos catalíticos y pueden descomponer los mRNAs. Por consiguiente, los siRNAs pueden estudiarse por utilización de un análisis de redes de patrones de expresión antes y después de la adición de siRNAs. Sin embargo, dado que la mayoría de los miRNAs no tienen actividad catalítica que conduzca a la descomposición de mRNAs, estos tipos de análisis no pueden aplicarse al estudio de los miRNAs.

Otra teoría acerca de la función de miRNA en la técnica anterior es que los mismos previenen la extensión del péptido. En este escenario, sería necesario considerar el producto proteínico a fin de analizar el comportamiento del mRNA.

5 Los métodos de la técnica anterior para la detección de miRNA han estado basados en redes de miRNA. Estas pueden producirse únicamente con el conocimiento de la secuencia del miRNA propiamente dicho. Adicionalmente, se han hecho intentos para estudiar estos fenómenos utilizando PCR en tiempo real para miRNAs específicos. Sin embargo, una vez más, este tipo de análisis está basado en el conocimiento de la secuencia precisa del miRNA.

10 A fin de comprender mejor la función de los miRNAs, se dedicado grandes esfuerzos a la identificación por computadora de las dianas de miRNA utilizando diversos algoritmos. No obstante, los inconvenientes de estas predicciones son que todas ellas generan un número sustancial de positivos falsos y pueden estar sesgadas dado que están basadas en su mayoría en el conocimiento obtenido a partir de las muy pocas interacciones miRNA:diana conocidas. Así pues, en este campo, la búsqueda de miRNAs candidatos se realiza directamente por técnicas de computadora. Sin embargo, las técnicas de computadora para la búsqueda de miRNAs adolecen de inconvenientes tales como estar inherentemente sesgadas hacia el pequeño número de miRNAs que han sido comprobados de hecho experimentalmente. Dado que el número de miRNAs comprobados es muy pequeño, el conjunto de secuencias de miRNA comprobadas a partir del cual pueden deducirse motivos o dominios conservados, es correspondientemente pequeño. En primer lugar, esto hace difícil extrapolar a partir de la superposición entre los pequeños números de secuencias conocidas para un conjunto más extenso de miRNAs candidatos. En segundo lugar, en cualquier muestra estadísticamente pequeña procedente de un grupo global grande se producirá un sesgo estadístico inherente por azar. Por tanto, dado que el número de miRNAs en el que se basan las predicciones por computadora es muy pequeño, es prácticamente seguro que existe un sesgo estadístico importante en las predicciones.

25 Adicionalmente, considerando los cuatro algoritmos de predicción principales, únicamente 0,01% de los pares miRNA/diana están predichos por cada uno de los algoritmos. De hecho, más del 80% de los pares están predichos por uno solo de los algoritmos. Así pues, la identificación o validación exacta de los apareamientos miRNA/diana es un problema en la técnica.

Una dificultad fundamental en el campo es el descubrimiento de una diana para un miRNA. Esto es especialmente difícil dado que se sabe que las dianas de miRNA no son necesariamente idénticas en secuencia a la secuencia de miRNA propiamente dicha.

30 Un método de la técnica anterior que intenta estudiar o cuantificar la acción de miRNAs el ensayo de la luciferasa de Ambion Inc's. Este ensayo implica la clonación de una diana y la combinación con el miRNA candidato, seguido por un ensayo de luciferasa diseñado para leer cualquier efecto, utilizando el plásmido de Ambion: pMIR-REPORT™, número de catálogo AM5795. En primer lugar, como se apreciará, es típicamente necesario conocer la diana o diana candidata antes que pueda realizarse este tipo de análisis. En segundo lugar, cada clon individual tiene que tratarse por separado, dado que no hay modo alguno de separar aquéllos que albergan un ácido nucleico de interés de los que no se encuentran en un escenario típico de cribado.

40 Otro modo de analizar los efectos de miRNAs por el uso de geles 2D para estudiar los patrones de expresión de proteínas. En este escenario, los patrones de expresión 2D de diversas proteínas se comparan entre un tratamiento con miRNA y un tratamiento sin miRNA. Sin embargo, la sensibilidad de esta técnica es muy baja. Es muy probable que no todas las proteínas sean detectadas por esta metodología más bien tosca. De hecho, se estima que sólo aproximadamente el 10% de las proteínas expresadas aparecen en el análisis de la expresión de proteínas de tipo gel 2D. Evidentemente, este enfoque no es suficientemente sensible para un estudio significativo de la acción de miRNA.

45 Adicionalmente, como se ha indicado arriba, dado que los miRNAs no degradan el RNA diana del mismo modo que lo hacen los siRNAs, tampoco es posible estudiar la acción de miRNA por monitorización de los niveles de mRNA.

WO 2004/097042 describe un método de selección de siRNA. Los siRNAs exhiben 100% de identidad con sus secuencias diana. Los clones utilizados comprenden únicamente un marcador por transcrito. El método se utiliza para seleccionar siRNA dirigido a cDNA clonado.

50 El Boletín Técnico Promega No. 329 "siCHECK™ Vectors" (12 febrero 2004), páginas 1-17, describe ciertos vectores para optimización de la interferencia del RNA (RNAi). En particular, se describen los vectores psiCHECK™-1 y psiCHECK™-2. Éstos comprenden luciferasa sintética de *Renilla* como gen informador. El documento describe también tablas de enzimas de restricción que tienen sitios en ellas, mapas de los vectores, y tampones para luminiscencia eficaz de luciferasa.

55 La técnica anterior adolece de inconvenientes como se indicado arriba. Adicionalmente, no existe ningún ensayo funcional para el descubrimiento de dianas en el campo del miRNA existentes en la técnica anterior. Además, existen ejemplos que exponen las limitaciones de los modelos de computadora. Por ejemplo, LED7 es un miRNA de *C. elegans*. Este gen (conocido como "letal 7") desactiva diversos genes y conduce a apoptosis en aquellos linajes de células en los cuales se expresa. Aplicando los modelos de computadora, es posible identificar secuencias

predichas a las que debería fijarse LED7. Sin embargo, se ha demostrado experimentalmente que muchas de estas dianas predichas no se fijan en absoluto a LED7. En contraste, se ha estudiado el gen ETWK3. En el curso de este estudio, se ha demostrado que el miRNA denominado miR143 es una diana auténtica de ETWK3. No obstante, mi143 no es predicho por la totalidad de los modelos de computadora arriba indicados, sino en el mejor de los casos por cierta proporción de los mismos. Por tanto, además de la predicción de dianas que no están fijadas de hecho por el miRNA, los modos de computadora no predicen tampoco con autenticidad los apareamientos de miRNA. Así pues, puede apreciarse que estos sistemas de computadora presentan en la técnica numerosos problemas e inconvenientes importantes asociados con ellos.

La presente invención trata de resolver los problemas asociados con la técnica anterior.

10 Sumario de la Invención

Los autores de la presente invención han diseñado ventajosamente un nuevo sistema que permite un ensayo funcional para la acción de RNA regulador tal como miRNA. La presente invención combina ventajosamente un marcador genético seleccionable con un sistema de clonación en el cual pueden insertarse UTR's 3' candidatas. De este modo, se hace posible estudiar los efectos de diversos miRNAs tanto en modalidad positiva como en modalidad negativa así como la expresión de RNAs particulares. El concepto fundamental es que los RNAs que están siendo estudiados (las UTR's 3' candidatas o los sitios diana para la acción del miRNA) están acoplados directamente a la secuencia codificante para el marcador seleccionable positivo y/o negativo.

Por tanto, siguiendo el marcador o marcadores seleccionables, puede obtenerse ventajosamente una lectura funcional directa del efecto de miRNAs particulares y dichos mRNAs. La presente invención está basada en este sorprendente descubrimiento. Una ventaja fundamental de la invención es que la misma proporciona una lectura funcional al nivel de las proteínas. Aunque algunos RNAs reguladores tales como siRNA producen escisión del RNA diana, lo cual permite el ensayo al nivel del RNA por ejemplo por monitorización de los niveles o escisión del RNA, otros RNAs reguladores tales como miRNAs no producen este efecto. Por ensayo de los efectos al nivel proteínico como se describe en esta memoria, pueden estudiarse funcionalmente numerosos tipos de RNA regulador, lo cual constituye una ventaja comparado con los métodos de la técnica anterior.

En un primer aspecto, la invención proporciona un ácido nucleico que comprende los elementos contiguos siguientes dispuestos en la dirección 5' a 3':

- a) un promotor;
- b) al menos dos marcadores seleccionables;
- c) un sitio de clonación para recepción de un segmento de ácido nucleico, comprendiendo dicho segmento una secuencia diana candidata de RNA regulador; y
- d) una señal de poliadenilación,

estando dichos elementos dispuestos de tal manera que un transcrito dirigido por dicho promotor comprende dichos al menos dos marcadores seleccionables, dicha secuencia diana candidata de RNA regulador, y dicha señal de poliadenilación por este orden.

Convenientemente, el ácido nucleico comprende DNA, por ejemplo un plásmido de DNA. Cuando el ácido nucleico comprende DNA, las referencias a secuencias diana de RNA, microRNA y similares deben entenderse de acuerdo con la convención, es decir, que las mismas definen la secuencia de nucleótidos que se especifica, y no requieren necesariamente que el ácido nucleico sea RNA (o un híbrido DNA-RNA. El lector experto comprenderá por tanto que la secuencia de nucleótidos comprende T o U en la posición apropiada tal como venga dictada por la naturaleza del ácido nucleico como es convencional en la técnica.

Es una característica importante que los elementos están dispuestos de tal manera que un transcrito (un solo transcrito) dirigido por dicho promotor comprende dichos marcadores seleccionables, dicha secuencia diana candidata de RNA regulador, y dicha señal de poliadenilación por este orden, es decir como un transcrito simple de RNA 'fusionado'. Los plásmidos conocidos para aplicaciones no conectadas no admiten fusión del transcrito de esta manera; por ejemplo, las bibliotecas convencionales de cDNA no dirigen tales transcritos fusionados. Ésta es una ventaja particular de la invención.

Convenientemente, dicha secuencia diana candidata de RNA regulador es una secuencia diana de microRNA (miRNA) candidata o una secuencia diana de RNA de interferencia corto candidata (siRNA). Convenientemente, dicha secuencia diana candidata de RNA regulador es una secuencia diana de microRNA (miRNA) candidata.

El término 'marcador(es) seleccionable(s)' utilizado en conexión con los ácidos nucleicos de la invención tiene su significado ordinario en la técnica y hace referencia convenientemente a un ácido nucleico que comprende un marco de lectura abierto que codifica un marcador polipeptídico seleccionable, es decir un polipéptido que confiere una propiedad o actividad seleccionable.

Convenientemente, el ácido nucleico comprende además un codón de parada localizado entre dicho marcador seleccionable y dicho sitio de clonación. Convenientemente, dicho codón de parada es una caja de parada que comprende codones de parada en cada uno de los tres marcos directos.

El o los marcadores seleccionables pueden ser para selección positiva.

5 El o los marcadores seleccionables pueden ser para selección negativa.

Convenientemente, el ácido nucleico puede comprender además (e) una señal de terminación de la transcripción. Se considera que la señal de poliadenilación será típicamente suficiente para aplicaciones de la invención en eucariotas superiores tales como mamíferos, pero si la invención se aplica en procariotas inferiores tales como eucariotas unicelulares o incluso procariotas, entonces un terminador de la transcripción puede proporcionar un control adicional ventajoso de la transcripción del RNA.

10 El marcador seleccionable comprende convenientemente dos o más marcadores seleccionables, convenientemente dos marcadores seleccionables. Adecuadamente, dichos dos o más marcadores seleccionables se proporcionan como un solo polipéptido o marco de lectura abierto (es decir una 'proteína de fusión'). Así, adecuadamente dichos dos marcadores seleccionables se proporcionan como un marco de lectura abierto que codifica un solo polipéptido que comprende dichos dos marcadores seleccionables. Convenientemente, dichos marcadores seleccionables comprenden al menos un marcador para selección positiva y al menos un marcador para selección negativa. Adecuadamente, dicho marcador seleccionable es una proteína de fusión HSVTK/PURO.

Convenientemente, dicho sitio de clonación es un sitio de clonación direccional.

20 Convenientemente, dicho sitio de clonación tiene insertado en él un segmento de ácido nucleico que comprende una UTR 3' o una UTR 3' candidata. En otro aspecto, la invención proporciona una biblioteca UTR 3' candidata, comprendiendo dicha biblioteca una pluralidad de dichos ácidos nucleicos. Convenientemente, dichas secuencias diana de miRNA candidatas están constituidas por cDNA's. Convenientemente, dicha secuencia diana de miRNA candidata es menor que 6 kb. Convenientemente, dicha secuencia diana de miRNA candidata es aproximadamente de 2 kb.

25 Convenientemente, dichos cDNA's son cDNA's de cerebro, cDNA's de testículos o son cDNA's de células de leucemia mieloide aguda.

La invención proporciona también una o más células que comprenden un ácido nucleico como se ha descrito arriba, o que comprenden bibliotecas como las arriba descritas.

30 En otro aspecto, la invención proporciona una población de células, albergando conjuntamente dichas células al menos parte de una biblioteca como se descrito arriba.

En otro aspecto, la invención proporciona un método de construcción de una biblioteca UTR 3' que comprende proporcionar un ácido nucleico como se ha descrito arriba, e insertar en dicho sitio de clonación un ácido nucleico que comprende una UTR 3' o una UTR 3' candidata.

35 En otro aspecto, la invención proporciona un método de construcción de una biblioteca UTR 5' que comprende proporcionar un ácido nucleico como se ha descrito arriba, e insertar en dicho sitio de clonación un ácido nucleico que comprende una UTR 5' o una UTR 5' candidata.

40 En otro aspecto, la invención proporciona un vector que comprende un ácido nucleico como se ha descrito arriba. El vector puede ser cualquier vector basado en ácido nucleico tal como un vector de plásmido, vector de transposón, vector viral o retroviral, u otro vector. Convenientemente, el vector es un vector de plásmido. El vector se proporciona convenientemente con elementos 'lanzadera' que permiten la propagación y/o amplificación en organismos hospedadores. Convenientemente, dichos elementos lanzadera son para propagación en células de *E. coli* e incluyen un origen de replicación de *E. coli*.

En otro aspecto, la invención proporciona un método para identificación de una secuencia diana de miRNA que comprende los pasos de

45 (a) introducir un ácido nucleico como se ha descrito arriba que comprende una secuencia diana de miRNA candidata en una célula hospedadora;

(b) seleccionar una o más células hospedadoras que expresan al menos un marcador seleccionable de dicho ácido nucleico;

(c) introducir al menos un miRNA de interés en dicha o dichas células hospedadoras de (b), y

50 (d) ensayar la expresión de al menos un marcador seleccionable de dicho ácido nucleico en las células de (c),

en donde, si las células de (c) no exhiben la expresión de al menos un marcador seleccionable, entonces la secuencia diana de miRNA candidata se identifica como una secuencia diana de miRNA.

En otro aspecto, la invención proporciona un método para identificación de un miRNA activo contra una secuencia diana de miRNA que comprende los pasos de

5 (a) introducir un ácido nucleico como se ha descrito arriba que comprende dicha secuencia diana de miRNA en una célula hospedadora;

(b) seleccionar una o más células hospedadoras que expresan al menos un marcador seleccionable de dicho ácido nucleico;

(c) introducir al menos un miRNA de interés en dicha o dichas células hospedadoras de (b), y

10 (d) ensayar la expresión de al menos un marcador seleccionable de dicho ácido nucleico en las células de (c),

en donde, si las células de (c) no exhiben la expresión de al menos un marcador seleccionable, entonces el miRNA de interés se identifica como un miRNA activo contra dicha secuencia diana de miRNA.

15 El paso (d) puede comprender seleccionar en contra de las células que expresan al menos un marcador seleccionable.

El paso (d) puede comprender seleccionar a favor de las células que no expresan al menos un marcador seleccionable.

En otro aspecto, la invención proporciona un método para identificar un inhibidor de un RNA regulador que comprende los pasos de

20 (a) introducir al menos un RNA regulador de interés en una célula hospedadora;

(b) introducir un ácido nucleico como se ha descrito arriba que comprende una secuencia diana de RNA candidata en dicha célula hospedadora;

(c) seleccionar una o más células hospedadoras que no exhiben expresión de al menos un marcador seleccionable de dicho ácido nucleico;

25 (d) introducir en dichas células hospedadoras una sustancia o ácido nucleico de test ;

(e) ensayar la expresión de al menos un marcador seleccionable en las células de (d);

en donde, si las células de (d) exhiben la expresión de al menos un marcador seleccionable, entonces la sustancia o ácido nucleico de test se identifica como inhibidora de dicho RNA regulador.

30 En otro aspecto, la invención proporciona un método para identificar una secuencia diana de RNA regulador que comprende los pasos de

(a) introducir un ácido nucleico como se ha descrito arriba que comprende una secuencia diana candidata de RNA regulador en una célula hospedadora;

(b) seleccionar una o más células hospedadoras que expresan al menos un marcador seleccionable de dicho ácido nucleico;

35 (c) introducir al menos un RNA regulador de interés en dicha o dichas células hospedadoras de (b), y

(d) ensayar la expresión de al menos un marcador seleccionable de dicho ácido nucleico en células de (c),

en donde, si las células de (c) no exhiben la expresión de al menos un marcador seleccionable, entonces la secuencia diana candidata de RNA regulador se identifica como una secuencia diana de RNA regulador.

40 Convenientemente, dicho RNA regulador es un siRNA y dicha secuencia diana candidata de RNA regulador es una secuencia diana de siRNA candidata.

En otro aspecto, la invención proporciona un método como se ha descrito arriba que comprende adicionalmente el paso de comparar las secuencias diana identificadas con secuencias diana conocidas del RNA regulador de interés, identificando con ello nuevas secuencias diana de dicho RNA regulador.

45 En otro aspecto, la invención proporciona un ácido nucleico como se ha descrito arriba en donde dicho ácido nucleico comprende la secuencia de ácido nucleico de una o más de SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6 o SEQ ID NO:7.

En otro aspecto, la invención proporciona un ácido nucleico como se ha descrito arriba en donde dicho ácido nucleico se selecciona de los plásmidos p3'UTR3, p3'UTRTPuro, p3'UTRHyTK o p3'UTRTPzeo.

Descripción Detallada de la Invención

5 La invención proporciona ventajosamente un ensayo funcional para descubrimiento y validación de dianas de microRNA. Se entenderá que el microRNAs una clase de RNAs reguladores, tales como RNAs reguladores pequeños. Otras clases de RNA regulador pequeño pueden abordarse también en las realizaciones expuestas en esta memoria. En particular, puede emplearse RNA de interferencia pequeño (siRNA) en sustitución de microRNA. Pueden incluso combinarse aplicaciones tanto de miRNA como de siRNA. Por conveniencia, la invención se describe con referencia principal a miRNA como el RNA regulador.

10 El término 'secuencia de siembra' es bien conocido en la técnica y hace referencia típicamente al extremo 5' del RNA regulador (v.g. siRNA o miRNA). Esto se refiere típicamente a las 6 o 7 bases en el extremo 5' del RNA regulador. Éstas tienen típicamente un 100% de coincidencia con la secuencia diana.

15 El término 'UTR 3'' significa literalmente la región no traducida 3'. Ésta es la región de un mRNA que no se traduce y es a menudo una diana de regulación de la traducción, por ejemplo por miRNAs. El término se utiliza a menudo dentro del término más amplio 'secuencia diana de miRNA' en esta memoria, dado que es posible que una secuencia diana de miRNA pueda no haberse derivado de, o no haberse demostrado experimentalmente que es, una UTR 3', v.g. si la secuencia diana de miRNA se ha generado o derivado de una fuente distinta de miRNA. Típicamente, la mayor parte o la totalidad de las secuencias diana de miRNA se encuentran en UTRs 3'. Sin embargo, obviamente las secuencias diana de miRNA pueden derivarse de otras localizaciones, por ejemplo del genoma como un todo, o pueden incluso crearse artificialmente por generación de una biblioteca o secuencias aleatorias o semialeatorias que pueden comprender secuencias diana de miRNA. Así pues, es preciso tener en cuenta que la invención se aplica generalmente a secuencias diana de miRNA, y que por conveniencia se hace referencia a menudo a éstas como UTR's 3' en esta memoria, pero que dichas secuencias diana o secuencias diana candidatas pueden derivarse de hecho de una o más fuentes que son distintas de las UTRs 3' reales definidas experimentalmente. Convenientemente, la secuencia diana de miRNAs, o se deriva de, una UTR 3'.

25 El término 'sitio de clonación' tiene su significado ordinario en la técnica. En particular, el mismo se refiere a un elemento o secuencia de ácido nucleico que permite la digestión del ácido nucleico por una enzima de restricción o catalizador similar para permitir la inserción de ácido nucleico en dicho sitio digerido. Ejemplos de sitios de clonación son sitios de clonación múltiples ('MCS') que caracterizan sitios de reconocimiento que comprenden secuencias de ácido nucleico para enzimas de restricción de ácido nucleico múltiples, permitiendo con ello estrategias de clonación alternativas en un solo sitio de clonación. Convenientemente, el sitio de clonación de la invención comprende secuencias de ácido nucleico reconocibles por al menos una enzima de restricción, convenientemente una enzima de restricción que permite clonación direccional, convenientemente Sfil. Así, en una realización el sitio de clonación es simplemente un sitio de reconocimiento de Sfil.

35 La secuencia codificante de los polipéptidos a expresar de acuerdo con la presente invención puede estar ventajosamente optimizada en codones para la célula diana (célula hospedadora) en la que debe tener lugar la expresión. En particular, los marcadores seleccionables están convenientemente optimizados en codones para las células en las que debe tener lugar la selección. Convenientemente, la optimización de codones se refiere a criterios humanos para células humanas.

40 Ventajas de la Invención

Los métodos de la técnica anterior para analizar la acción del miRNA están basados en el uso de luciferasa. La luciferasa es una proteína cuya actividad puede medirse por monitorización de luminosidad o luz emitida. La luciferasa no proporciona selección positiva ni negativa. Utilizando un sistema basado en luciferasa, es sin duda muy intensivo en mano de obra el cribado para averiguar los efectos de miRNAs particulares. En primer lugar, si tuviera que aplicarse esta técnica a UTR's 3' o UTR's 3' candidatas, ello tendría que hacerse cada vez en un tratamiento independiente. Esto podría implicar cualquier número hasta 40-100.000 experimentos o tratamientos separados. Evidentemente, éste es un procedimiento muy engorroso y costoso de realizar. En contraste, de acuerdo con la presente invención, la acción del miRNA puede evaluarse utilizando técnicas de selección genética. Esto permite ventajosamente seleccionar células que expresen ciertos marcadores seleccionables, y selecciona genéticamente de manera directa los efectos del miRNA (sean positivos o negativos) sin recurrir a ningún ensayo de luminiscencia. Además de evitar ensayos de luminiscencia que consumen mucho tiempo, la presente invención ofrece la ventaja adicional de poder procesar análisis múltiples en paralelo, dado que solamente las células que alberguen (o expresen) ciertos constructos genéticos predeterminados sobrevivirán a los procedimientos de selección.

55 Con objeto de comprender mejor esta ventaja, considérese la ilustración siguiente. En primer lugar, de acuerdo con la presente invención las células que albergan un constructo genético particular pueden seleccionarse en un primer paso de selección positiva. Esto da como resultado la pérdida de células que no albergan el ácido nucleico de interés. Así, todas las células supervivientes tienen, por deducción (por selección) que albergar el constructo genético de interés. Esta primera población de células seleccionada positivamente puede seguir luego al segundo paso del procedimiento. En el segundo paso del procedimiento, dichas células se tratan con miRNA, seleccionándose

aquellas células en las cuales el miRNA afecta a la expresión de la proteína del marcador de interés. De este modo, por realización de este segundo paso de selección, se aíslan genéticamente aquellas células que albergan un constructo genético que es sensible al miRNA particular que se estudia.

5 Es una ventaja de la invención que puede estudiarse una población de células por el procedimiento de selección múltiple. De hecho, en términos prácticos, es una ventaja de la invención que puede estudiarse en una sola operación una población de células, albergando individualmente dichas células constructos genéticos diferentes. Por supuesto, cuando se estudia una gran población de células, o por conveniencia dependiendo del formato del estudio, pueden ser ventajosas operaciones múltiples, pero una ventaja fundamental es que el procedimiento selectivo múltiple permite el procesamiento paralelo de células que albergan diferentes constructos genéticos en la
10 etapa de selección, en lugar de tener que procesar clones individuales por separado a lo largo de todo el procedimiento. Evidentemente, este tipo de aplicación no es posible con los análisis de la técnica anterior basados en luciferasa. Al menos una razón para esto es que no es viable el aislamiento de células que expresen un nivel particular de luciferasa de células comparables que difieran únicamente en alguna característica de su expresión de luciferasa.

15 **Marcadores Seleccionables**

El ácido nucleico de la presente invención comprende también un gen marcador seleccionable. Un gen marcador seleccionable permite que las células que llevan el gen se seleccionen específicamente o se excluyan, en presencia de un agente de selección correspondiente. Los marcadores seleccionables pueden ser positivos, negativos o bifuncionales. Los marcadores de selección positivos permiten seleccionar las células que llevan el marcador, mientras que los marcadores de selección negativos permiten que las células que llevan el marcador se eliminen selectivamente. Un marcador seleccionable bifuncional contiene medios para selección positiva o negativa de las células que contienen el gen marcador o gen de fusión seleccionable (véase Schwartz et al Proc. Natl. Acad. Sci USA 88:10416-10420 (1991)).

25 El uso de marcadores seleccionables en los ácidos nucleicos y las técnicas de la presente invención conduce a varias ventajas que se exponen en esta memoria. Una de dichas ventajas es que ello permite la selección de células que albergan constructos genéticos de interés. Adicionalmente, el uso de marcadores seleccionables múltiples puede permitir la implementación de un régimen de selección más complejo. Por ejemplo, por la utilización de dos marcadores seleccionables puede seleccionarse una primera población de células que albergan ácidos nucleicos de una biblioteca, y puede utilizarse un segundo marcador seleccionable para seleccionar aquellas células que regulan
30 en sentido decreciente la expresión por la vía de la UTR después de la adición de miRNA.

Típicamente, un gen marcador seleccionable conferirá resistencia a un fármaco (v.g. profármaco-convertasa) o compensará un defecto metabólico o catabólico en las células hospedadoras. Por ejemplo, marcadores seleccionables utilizados comúnmente con células de mamífero incluyen los genes para adenina-desaminasa (*ada*), higromicina B fosfotransferasa (*Hph*), dihidrofolato-reductasa (DHFR), timidina-quinasa (TK), timidilato-quinasa (que
35 convierte AZT y puede ser más potente que la timidina-quinasa), glutamina-sintetasa (GS), asparagina-sintetasa, y genes que codifican resistencia a neomicina (G418), puromicina, histidinol, zeocina (la zeocina puede sustituirse con bleomicina y/o thleomicina dado que el gen de resistencia es el mismo para las tres; la zeomicina es típicamente adecuada, debido a su menor coste) y Blastidicina S.

Los agentes de selección se utilizan de acuerdo con las recomendaciones del fabricante en caso apropiado. Como
40 guía, la selección con ZEO puede llevar aproximadamente 3 semanas, y la selección con PURO puede llevar aproximadamente una semana. Las concentraciones y condiciones, con inclusión del nivel de expresión del marcador seleccionable pueden ser manipuladas todas ellas por el técnico experto para modificar los tiempos de selección de acuerdo con las necesidades.

El gen marcador seleccionable puede ser cualquier gen que pueda complementar una deficiencia celular reconocible. Así, por ejemplo, podría utilizarse el gen para HPRT como la secuencia genética del marcador
45 seleccionable cuando se emplean células que carecen de actividad de HPRT. Así, este gen es un ejemplo de un gen cuyo producto de expresión puede utilizarse para seleccionar células mutantes, o para "seleccionar negativamente" células que expresen este producto génico. Otro ejemplo es el uso del gen marcador seleccionable puromicina-N-acetiltransferasa (Pac) que confiere resistencia al fármaco puromicina en las células portadoras del gen.

50 Otro gen marcador seleccionable común utilizado en sistemas de expresión de mamífero es la timidina-quinasa. Las células que no contienen una enzima timidina-quinasa (TK) activa son incapaces de crecer en medio que contenga timidina, pero son capaces de crecer en un medio que contenga análogos de nucleósidos tales como 5-bromodesoxiuridina, 6-tioguanina, 8-azapurina, etc. Inversamente, las células que expresan timidina-quinasa activa son capaces de crecer en medios que contienen hipoxantina, aminopterina, timidina y glicina (medio HATG) por son
55 incapaces de crecer en medio que contenga análogos de nucleósidos tales como 5-azacitindina (Giphart-Gassler, M et al Mutat. Res. 214:223-232 (1989), Sambrook et al, en: Molecular Cloning A Laboratory Manual, segunda edición, Cold Spring Harbour Laboratory Press, N.Y. (1989)). Las células que contienen un gen de timidina-quinasa del virus del herpes simplex (HSV-TK) como gen marcador seleccionable son incapaces de crecer en presencia de ganciclovir o agentes similares. Evidentemente, el agente utilizado para implementar la selección debe utilizarse de acuerdo con las instrucciones del fabricante. Es posible que la concentración o el modo/momento exacto de la
60

5 adición del agente a las células pudiera precisar optimización para los constructos o marcadores seleccionables particulares utilizados a fin de proporcionar la selección más robusta y fiable. Esta optimización está perfectamente dentro de las capacidades de la persona experta en la técnica. Es posible incluso que pudiera implementarse una estrategia de selección de nivel mixto, por ejemplo con niveles incrementados del agente de interés para seleccionar los clones expresantes máximos, o viceversa con un nivel inferior para seleccionar clones menos expresantes. Tales variaciones están plenamente dentro del alcance de la persona experta que lleve a la práctica la invención.

10 Además, se han creado mutantes de enzimas metabólicas que permiten una mayor sensibilidad a los fármacos. Por ejemplo, la timidilato-quinasa F105Y aumenta la sensibilidad de las células a AZT, lo cual puede permitir a su vez la utilización de menos AZT, o puede conducir a una destrucción más rápida para una concentración de AZT dada. También puede utilizarse el mutante R16GLL. Adicionalmente, se ha demostrado que un mutante de HSVTK denominado SC39 es significativamente más sensible a ganciclovir y/o agentes similares (Blumental et al, Mol. Therapy, 2007). Así pues, mutantes de marcadores seleccionables conocidos encuentran también aplicación en las realizaciones expuestas en esta memoria.

15 Así, para selección negativa pueden utilizarse HSVTK, timidilato-quinasa (tal como F105Y u otras). Para selección positiva pueden utilizarse PURO, ZEO, HYGRO o incluso NEO. Convenientemente, las fusiones de la invención comprenden un marcador positivo y un marcador negativo de estos grupos. Convenientemente, las fusiones pueden hacerse en cualquier orden. Las más preferibles son las citadas en la sección de ejemplos. De hecho, se ha demostrado que éstas funcionan satisfactoriamente como se ilustra, lo cual no puede asumirse a partir de una comprensión de su comportamiento en otros contextos.

20 Existen algunas fusiones con anterioridad a la invención, tales como TK/ZEO (Cayla/Invitrogen) o HYGRO/TK (Immunex). Éstas son conocidas únicamente para aplicaciones de tipo terapia génica, v.g. para destrucción de células que hayan recibido el vector después de la conclusión del tratamiento (es decir utilización como gen suicida). Se prefieren las combinaciones o fusiones descritas en esta memoria por primera vez. En cualquier caso, la fusión con RNAs reguladores como se expone en la invención no ha sido descrita ni sugerida con anterioridad.

25 Además, los marcadores seleccionables no siempre implican necesariamente destrucción de células, v.g. puede utilizarse proteína verde fluorescente (GFP)/PURO (al igual que otros agentes de fluorescencia o proteínas visualizables) para selección por clasificación de flujo, es decir marcadores seleccionables por clasificación de flujo.

Combinaciones particularmente adecuadas incluyen TK/PURO, wtThym/PURO, R16GLLThym/PURO, F105YThym/PURO, R16GLL-F105YThym/PURO, F105YThym/Zeo, Zeo/F105YThym, GFP/PURO.

30 Se describe que es posible que pueda utilizarse una estrategia seleccionable dual con un solo marcador seleccionable. En este escenario, sería necesario elegir el marcador seleccionable de tal manera que el mismo proporcione a la vez selección positiva y negativa. Por ejemplo, la enzima metabólica codificada por el gen URA puede proporcionar independencia de uracilo en ciertos sistemas eucariotas. Así, las células que albergan el gen URA pueden seleccionarse positivamente utilizando medio exento de uracilo – únicamente aquellas células que alberguen el gen URA serán capaces de crecer por fabricación de su propio uracilo. Exactamente el mismo gen es capaz de convertir el precursor ácido 5-fluoro-orótico (5-FA) en un metabolito tóxico. Así pues, las células que albergan el gen de uracilo pueden seleccionarse negativamente por inclusión de 5-FA en el medio de crecimiento - aquellas células que alberguen el gen URA convertirán el mismo en un metabolito tóxico y serán eliminadas por el procedimiento de selección. Así, en este escenario, un solo marcador seleccionable puede proporcionar de hecho a la vez pasos de selección positivos y negativos. Sin embargo, en la mayoría de los casos, los pasos de selección positivos y negativos serán proporcionados por la provisión de dos o más marcadores seleccionables.

40 De manera similar, puede utilizarse como marcador seleccionable la citosina-desaminasa. Las células normales de mamífero no contienen citosina-desaminasa. Las células que expresan el gen de citosina-desaminasa metabolizan el profármaco relativamente no tóxico 5-fluorocitosina a 5-fluorouracilo, altamente tóxico. Así, puede utilizarse citosina-desaminasa como marcador seleccionable permitiendo con ello selección negativa cuando se trata con 5-fluorocitosina en diferentes realizaciones.

45 Convenientemente, los marcadores seleccionables múltiples se proporcionan como fusiones en un solo marco de lectura abierto en el ácido nucleico de la invención.

50 Se utilizan convenientemente al menos dos marcadores seleccionables. Adecuadamente, se utilizan tres marcadores seleccionables. Adecuadamente, se utilizan cuatro marcadores seleccionables, o incluso más.

55 Se utilizan adecuadamente dos marcadores seleccionables, y convenientemente dichos dos marcadores seleccionables están fusionados. 'Fusionado' tiene su significado ordinario la técnica, es decir significa que convenientemente los marcadores pueden expresarse a partir de un solo marco de lectura abierto que codifica un polipéptido que tiene la secuencia de aminoácidos de cada uno de dichos marcadores. Así, 'fusionado' significa que convenientemente los dos o más marcadores seleccionables se proporcionan en un solo polipéptido (o un solo ácido nucleico o transcrito que codifica un solo polipéptido que comprende dichos dos o más marcadores seleccionables). En otras palabras, los marcos de lectura abiertos para los marcadores están 'fusionados' al nivel del ácido nucleico dando como resultado la expresión de una 'proteína de fusión' que comprende las secuencias de aminoácidos de

cada uno de los dos (o más) marcadores que se dice están 'fusionados'. Esto permite ventajosamente la realización de un procedimiento de cribado de selección dual, por ejemplo selección positiva respecto a la presencia del constructo genético seguido por selección negativa contra aquellas células que fracasan en cuanto a la regulación de la expresión en sentido decreciente en presencia del miRNA a testar.

- 5 Así pues, convenientemente el o los ácidos nucleicos que codifican los dos o más marcadores de selección proporcionados como un solo polipéptido de 'fusión' no tienen ningún codón de parada intercalado entre las partes del marco de lectura abierto que codifica los dos marcadores seleccionables.

Convenientemente, las fusiones de marcadores seleccionables se seleccionan a partir de las combinaciones de TK/PURO, TK/HYGRO, o TK/ZEO. Las fusiones de marcadores seleccionables enumeradas pueden invertirse típicamente, v.g. HYGRO/TK o TK/HYGRO pueden ser igualmente eficaces y debería entenderse que ambas están abarcadas por referencia a "HYGRO/TK" o "TK/HYGRO". En el caso de que se precise cualquier orientación adicional, convenientemente por defecto la fusión es tal como se escribe, v.g. HYGRO/TK significa término N-HYGRO/TK-término C, a no ser que el contexto indique otra cosa. Muy convenientemente, un marcador seleccionable es una fusión TK/PURO. Esto tiene la ventaja de que la puromicina es muy potente. Éste es posiblemente el mejor marcador seleccionable. La puromicina bloquea la síntesis de proteínas. Esto permite seleccionar una población pura de células transfectadas en aproximadamente una semana en condiciones de laboratorio.

La higromicina es también un marcador seleccionable muy potente. La higromicina es comparable a la puromicina en su potencia.

- 20 La zeomicina es un agente de intercalación. La zeomicina tiene un modo de acción más lento comparada con puromicina o higromicina. Esto puede ser ventajoso en ciertas situaciones.

Así pues, convenientemente el marcador seleccionable es una fusión de las proteínas HSVTK y PURO. De manera conveniente, dicha fusión comprende SEQ ID NO: 1, y convenientemente dicha fusión consiste en SEQ ID NO: 1.

- 25 Pueden utilizarse otras profármaco-convertasas en lugar de HSVTK, v.g. beta-glucosidasa u otras mencionadas en esta memoria, particularmente como se han citado arriba (genes marcadores seleccionables).

En una realización amplia, podrían utilizarse otras vías de selección de células tales como selección con perlas respecto a la presencia o ausencia de marcadores tales como LNGFR en la superficie celular.

Promotores

- 30 El ácido nucleico de la presente invención comprende un promotor enlazado operativamente a una secuencia codificante que codifica, por ejemplo, un gen marcador seleccionable. El término "enlazado operativamente" significa que los componentes descritos se encuentran en una relación que permite que los mismos funcionen de su manera propuesta. Un promotor enlazado operativamente a una secuencia codificante está posicionado de tal manera que la expresión de la secuencia codificante se logra en condiciones en las cuales el promotor es activo.

- 35 El término "promotor" hace referencia a una secuencia polinucleotídica que controla la transcripción de un gen o secuencia al (o a la) cual está enlazada operativamente. Un promotor incluye señales para la fijación de la RNA-polimerasa e iniciación de la transcripción. El término promotor es bien conocido en la técnica y abarca secuencias de polinucleótidos comprendidas en tamaño y complejidad desde promotores mínimos a promotores que incluyen elementos e intensificadores aguas arriba.

- 40 Un promotor está posicionado usualmente, pero no con carácter necesario, aguas arriba de la secuencia codificante cuya expresión regula. Adicionalmente, los elementos reguladores que comprenden un promotor están posicionados usualmente dentro de 2 kilobases del sitio de comienzo de la transcripción de un gen.

- 45 Una persona con especial ordinaria en la técnica entenderá que la selección de un promotor útil particular depende de las líneas de células exactas y otros diversos parámetros del vector de expresión a utilizar para expresar la secuencia codificante. Un gran número de promotores, que incluyen promotores constitutivos, inducibles y reprimibles de una diversidad de fuentes diferentes son bien conocidos en la técnica y pueden identificarse en bases de datos tales como GenBank, estando disponibles como o dentro de polinucleótidos clonados, por ejemplo de depositarias tales como ATCC y otras fuentes comerciales o individuales.

- 50 Promotores adecuados para uso en los ácidos nucleicos de la presente invención incluyen los derivados de genes de mamífero, microbios, virus o insectos. Secuencias promotoras de células de mamífero utilizadas comúnmente se derivan de poliomavirus, adenovirus, retrovirus, virus de la hepatitis B, virus 40 de los simios (SV40) y citomegalovirus. También pueden utilizarse promotores mínimos tales como el promotor timidina-quinasa del virus del herpes simple (HSVtk). Son también adecuados promotores de mamífero tales como el promotor beta-actina para uso en los ácidos nucleicos de la presente invención. Pueden ser adecuados también promotores procedentes de la célula hospedadora o una especie afín.

Puede utilizarse el promotor temprano inmediato constitutivo de citomegalovirus para obtener un nivel alto de expresión génica en células de mamífero. Tales promotores están ampliamente disponibles y pueden obtenerse por ejemplo de Stratagene (por ejemplo el vector pCMV-Script®). Otro promotor constitutivo, el intensificador/promotor SV 40 (que incluye el promotor SV 40 tardío o temprano), es utilizado comúnmente en la técnica y permite un nivel moderadamente alto de expresión génica en células de mamífero.

Puede ser también ventajoso que los promotores sean inducibles. Con los promotores inducibles, la actividad del promotor aumenta o disminuye en respuesta a una señal. Por ejemplo, el promotor tetraciclina (tet) que contiene la secuencia del operador tetraciclina (tetO) puede ser inducido por una proteína transactivadora regulada por tetraciclina (tTA). La fijación de la tTA al tetO se inhibe en presencia de tet. Los Sistemas de Expresión de Genes Tet-On y Tet-Off (Clontech) utilizan un elemento sensible a tetraciclina para mantener la expresión de proteína recombinante en un modo "encendido" (constitutivamente "apagado", pero inducido con tetraciclina) o "apagado" (constitutivamente "encendido", pero reprimido con tetraciclina o doxiciclina). Detalles de otros promotores inducibles adecuados que incluyen jun, fos y los promotores metalotioneína y del choque térmico, pueden encontrarse en Sambrook et al, en: Molecular Cloning A Laboratory Manual, segunda edición, Cold Spring Harbour Laboratory Press, N.Y. (1989) y Gossen et al Curr Opin Biotech 5:516-520 (1994).

Adicionalmente, cualquiera de estos promotores puede modificarse por la adición de secuencias reguladoras adicionales, por ejemplo secuencias intensificadoras enlazadas operativamente a la secuencia codificante. Un intensificador es un elemento de DNA con acción cis que actúa sobre un promotor para aumentar la transcripción. Puede ser necesario un intensificador que funcione en conjunción con el promotor a fin de aumentar el nivel de expresión obtenido con un promotor solo. Los intensificadores enlazados operativamente pueden estar localizados aguas arriba, dentro de o aguas abajo de las secuencias codificantes, y pueden encontrarse a distancias considerables del promotor.

Terminador de la transcripción

Los ácidos nucleicos de la presente invención pueden comprender también un terminador de la transcripción. Un "terminador de la transcripción" hace referencia a una secuencia de nucleótidos representada normalmente en el extremo 3' de un gen de interés o al tramo de secuencias a transcribir que hace que la RNA-polimerasa termine la transcripción.

Un elemento genético separado es la señal de poliadenilación, que facilita la adición de secuencias poliadenilato al extremo 3' de un transcrito primario. La secuencia de la señal de poliadenilación incluye la secuencia AATAAA localizada aproximadamente 10-30 nucleótidos aguas arriba del sitio de escisión, más una secuencia aguas abajo. La señal de poliadenilación puede estar localizada muy próxima al terminador de la transcripción (cuando está presente) o puede incluso superponerse con el mismo en algunas circunstancias.

Por regla general, la mayoría de los terminadores de la transcripción incluyen una secuencia rica en GC que precede al sitio de terminación, y una secuencia de residuos T en la cadena no-molde del DNA unida al sitio de terminación. La RNA-polimerasa atraviesa la secuencia rica en GC para producir mRNA que puede formar una estructura estable de tallo y bucle apareada en bases dentro del mRNA. La transcripción termina entonces por lo general precisamente aguas abajo de la estructura de tallo y bucle donde los residuos T dan como resultado un RNA que termina con una secuencia que comprende fundamentalmente residuos uridilato (Brennan y Geiduschek, 1983, Nucleic Acids Res. 11:4157).

Un ejemplo de una secuencia terminadora es el del gen de la hormona de crecimiento de los bovinos. Este elemento terminador puede proporcionar también la señal de poliadenilación. Pueden obtenerse también secuencias terminadoras de suministradores comerciales muy conocidos tales como el Sistema Vector ZAP Express® (Stratagene) y el pCMV-V5-His6 (disponible de Clontech Laboratories (Palo Alto, Calif.). Terminadores activos en sistemas de expresión de mamífero se describen en la bibliografía y pueden ser obtenidos fácilmente por las personas expertas en la técnica.

Transfección/Transducción

La expresión "transfección celular" se refiere a la introducción de ácido nucleico extraño en una célula. Existen varios métodos de introducción de DNA y RNA en una célula, que incluyen métodos químicos de transfección (mediados por liposomas, lípidos no liposómicos, dendrímeros), métodos físicos de suministro (electroporación, microinyección, choque térmico), y transferencia de genes basada en virus (retrovirus, virus adeno-asociados, y lentivirus). El método de elección dependerá usualmente del tipo de célula y la aplicación de clonación, y métodos alternativos son bien conocidos por los expertos en la técnica. Tales métodos se describen en muchos manuales estándar de laboratorio tales como Davis et al, Basic Methods In Molecular Biology (1986).

El material genético transfectado puede expresarse en la célula con carácter transitorio o permanente. En la transfección transitoria, el DNA se transfiere y está presente en la célula, pero los ácidos nucleicos no se integran en los cromosomas de la célula hospedadora. Típicamente, la transfección transitoria da como resultado niveles de expresión altos del RNA introducido 24-72 horas después de la transfección, y del DNA 48-96 horas después de la

transfección. La transfección estable se consigue por integración del vector de DNA en el DNA cromosómico y se expresa de manera permanente en el genoma de la célula.

Puede utilizarse transfección utilizando liposomas disponibles comercialmente tales como Lipofectinamina 2000, electroporación o cualquier otra forma de transducción. Adicionalmente, el ácido nucleico tal como el microRNA de interés puede clonarse en plásmidos de expresión virales o no virales que pueden introducirse luego por infección (vectores virales) o transfección (no virales). Esto dará como resultado la transducción estable de las células. Tales detalles son comunes y bien conocidos por las personas expertas en la técnica. En particular, dichas técnicas pueden practicarse de acuerdo con Sambrook, E.F. Fritsch, y T. Maniatis, 1989, *Molecular Cloning: A Laboratory Manual*, segunda edición, volúmenes 1-3, Cold Spring Harbor Laboratory Press; Ausubel, F.M. et al. (1995 y suplementos periódicos; *Current Protocols in Molecular Biology*, capítulos 9,13, y 16, John Wiley and Sons, Nueva York, N.Y.).

Los medios químicos de transfección de células con ácido nucleico extraño incluyen el uso de DEAE-dextrano, fosfato de calcio o liposomas artificiales. DEAE-dextrano es un polímero catiónico que se asocia con ácidos nucleicos cargados negativamente. Un exceso de carga positiva, aportada por el polímero en el complejo DNA/polímero permite que el complejo entre en asociación más estrecha con la membrana celular cargada negativamente. Se cree que la captura subsiguiente del complejo por la célula es por endocitosis. Este método tiene éxito en el caso de suministro de ácidos nucleicos en células para expresión transitoria. Otros polímeros catiónicos sintéticos pueden utilizarse para la transferencia de ácido nucleico a células, con inclusión de polibreno, polietiliminina y dendrímeros.

La transfección utilizando un método de co-precipitación con fosfato de calcio puede utilizarse para transfección transitoria o estable de una diversidad de tipos de células. Este método implica mezclar el ácido nucleico a transfectar con cloruro de calcio, añadir esto de manera controlada a una solución tamponada salina/fosfato y dejar que la mixtura se incube a la temperatura ambiente. Este paso genera un precipitado que se dispersa en las células cultivadas. El precipitado que incluye ácido nucleico es capturado por las células por endocitosis o fagocitosis. Esto se aplicado en gran escala para células de mamífero, por ejemplo como se expone en J R Rayner y T J Gonda ("A simple and efficient procedure for generating stable expression libraries by cDNA cloning in a retroviral vector." *Mol Cell Biol.*, febrero 1994; 14(2): 880-887).

Puede utilizarse transfección utilizando liposomas artificiales para obtener expresión transitoria o a largo plazo de ácido nucleico extraño en una célula hospedadora. Este método puede ser útil también para transfectar ciertos tipos de células que son intransigentes al fosfato de calcio o al DEAE-dextrano.

Los liposomas son pequeños cuerpos fijados a la membrana que pueden fusionarse realmente con la membrana celular, liberando ácido nucleico en la célula. Un lípido con carga global positiva neta a pH fisiológico es el componente lipídico sintético más común de los liposomas desarrollados para métodos de transfección que utilizan liposomas artificiales. A menudo, el lípido catiónico está mezclado con un lípido neutro tal como L-dioleoilfosfatidil-etanolamina (DOPE). La porción catiónica de la molécula lipídica se asocia con los ácidos nucleicos cargados negativamente, dando como resultado la compactación del ácido nucleico en un complejo liposoma/ácido nucleico. Después de la endocitosis, los complejos aparecen en los endosomas, y más tarde en el núcleo. Reactivos de transfección que utilizan lípidos catiónicos para el suministro de ácidos nucleicos a células de mamífero están ampliamente disponibles y pueden obtenerse por ejemplo de Promega (TransFast™ Transfection Reagent).

40 **Ventajas Adicionales**

El uso de un marcador seleccionable en el estudio de la función del miRNA no ha sido descrito con anterioridad. Como se ha indicado arriba, el análisis en este campo ha estado confinado típicamente al uso de marcadores cuantificables tales como luciferasa. En el intento de cuantificar los efectos sobre la expresión de proteínas de miRNAs particulares, la luciferasa es particularmente atractiva. Esto permite hacer medidas directamente comparables de luminiscencia y compararlas para diferentes tratamientos. En contraste acusado, los marcadores seleccionables operan sobre una base más binaria. El concepto fundamental de un marcador seleccionable es que puede hacerse que las células que albergan el marcador sobrevivan, y las células que carecen del marcador (o que no expresan el marcador) puedan eliminarse. Así, puede considerarse que el uso de marcadores seleccionables en el campo del análisis del miRNA es contrario a la intuición. Adicionalmente, comparado con el uso de luciferasa de la técnica anterior, el uso de marcadores seleccionables representa una pérdida de información. Esto es debido, como se ha indicado arriba, a que la luciferasa se adapta muy bien para cuantificación y para comparación de los niveles de expresión entre tratamientos, información que está disponible o puede medirse raras veces utilizando marcadores seleccionables. Es decir que los métodos y materiales de la presente invención pueden considerarse como contrarios a la intuición en relación con la técnica anterior. Evidentemente, en un campo tal como el análisis de miRNA, que está tan estrechamente basado en niveles de expresión comparativos, la idea de conversión a un sistema que permita únicamente análisis binario a partir de los antecedentes de un sistema que permita la realización de medidas proporcionales y deducciones directas de amplio alcance en relación con la expresión de proteínas podría descartarse sin más. *A priori*, parecería ciertamente que esto es un paso atrás en términos de la información que puede extraerse útilmente de un análisis de este tipo. No obstante, como se muestra en esta

memoria, es sorprendentemente útil de hecho emplear técnicas de selección genéticas en el análisis de la función del miRNA, y particularmente para la identificación de dianas de dichos miRNAs.

Es una ventaja la invención que se utilice una estrategia de clonación direccional. En una realización preferida, se utiliza clonación con Sfil. Esta es una enzima de restricción de corte raro. Sfil corta en una secuencia de reconocimiento de 8 pares de bases. Adicionalmente, Sfil corta dejando un saliente simétrico en los dos extremos del corte. Esto permite ventajosamente estrategias de clonación direccionales después de la digestión con Sfil. Estos métodos se describen en la técnica anterior, por ejemplo en la patente US No. 5.595.895, que se incorpora en esta memoria por referencia. Evidentemente, la invención abarca cualquier sistema de clonación direccional adecuado para uso en un constructo de ácido nucleico tal como la clonación con BstXI. La o las enzimas de restricción utilizadas para clonación direccional pueden ser BstXI. Esto se describe también en la patente US número 5.595.895. Se prefiere la clonación direccional con Sfil debido a su simplicidad. Una ventaja adicional del uso de la clonación con Sfil es que una secuencia de reconocimiento de 8 pares de bases es relativamente rara en el genoma. Por ejemplo, si se utilizan enzimas de restricción de corte más frecuente tales como SpeI o HindIII, entonces existe un riesgo correspondientemente mayor de que las mismas digieran las secuencias diana durante la operación de clonación, riesgo que se reduce con el uso de una secuencia de reconocimiento más larga, tal como una secuencia de reconocimiento de 8 pares de bases.

En contraste con los vectores de expresión en otros campos, se prefiere que los ácidos nucleicos de la presente invención exhiban codones de parada después del marcador seleccionable. En esta realización, cuando el marcador seleccionable es un polipéptido codificado por el ácido nucleico, la traducción de dicho polipéptido marcador seleccionable termina en el codón de parada. Así, tanto si cualquier secuencia presente en el ácido nucleico invención como una UTR 3' o un candidato para UTR 3' codifica cualquier polipéptido como si no lo hace, ello no afectaría a la operación de la invención. De hecho, puede ser una ventaja de la invención que, idealmente, alguna de tales secuencias codificantes presentes en la UTR 3' o la UTR 3' candidata no esté fusionada con el polipéptido del marcador seleccionable. Así, el codón de parada o codones de parada (convenientemente una caja de parada) presente directamente después del marco de lectura abierto que codifica el polipéptido del marcador seleccionable presenta la ventaja de impedir o al menos oponerse a la traducción de cualesquiera secuencias adicionales de ácido nucleico aguas abajo.

Una caja de parada es un elemento genético conocido comúnmente en la técnica. En resumen, una caja de parada comprende al menos tres codones de parada, que están dispuestos en un formato solapante o no solapante de tal modo que entre el extremo 5' y el extremo 3' de la caja de parada está presente un codón de parada en cada una en cada uno de los tres marcos de lectura directos posibles. Los codones de parada pueden estar superpuestos, o pueden estar separados por un pequeño número de nucleótidos, por ejemplo separados por 1, 2, 4, 5 o más nucleótidos. Obviamente, es improbable que los codones de parada estén separados por 3, 6, 9 o cualquier otro número de nucleótidos divisible por 3, dado que los codones de parada dispuestos de esta manera no se presentarían en marcos de lectura diferentes. Sin embargo, debe observarse por supuesto que dos o más codones de parada en marco son también útiles, por ejemplo para prevenir contra la lectura completa, y pueden emplearse por tanto en realizaciones adecuadas, utilizando por ejemplo codones de parada repetidos o duplicados, o incluso cajas de parada, en caso apropiado. Tales detalles son bien conocidos por una persona experta en la técnica.

Bibliotecas de cDNA

Convenientemente, las UTR's 3' o UTR's 3' candidatas se derivan de bibliotecas de cDNA. Convenientemente, los cDNA's son cDNA's de mamífero. Convenientemente, los cDNA's son de un tejido o enfermedad de interés. Por ejemplo, los cDNA's pueden ser de cerebro. Éste presenta la ventaja de ser un tejido que presenta los cDNA's más diversos. De esta manera, pueden prepararse cDNA's de un solo tejido, pero tienen la probabilidad máxima de representar el mayor número posible de genes diferentes. En otra realización, los cDNA's pueden ser de una enfermedad de interés. Un ejemplo de una enfermedad de este tipo es la leucemia mieloide aguda. En esta realización, convenientemente los cDNA's se derivan todos ellos de las células de leucemia mieloide aguda. Esto tiene la ventaja de presentar UTR's 3' o UTR's 3' candidatas que son probablemente relevantes para la enfermedad seleccionada.

En principio, las UTR's 3' o UTR's 3' candidatas pueden derivarse de cualquier fuente genética adecuada. Las bibliotecas de cDNA son una fuente particularmente conveniente a partir de la cual acceder a UTR's 3' o UTR's 3' candidatas. La utilización de cDNA's como la fuente para las UTR's de interés presenta varias ventajas. En primer lugar, las bibliotecas de cDNA pueden seleccionarse con oligo-dT, por ejemplo solos o en combinación con hexámeros aleatorios. Esto tiene el efecto de hacer que las bibliotecas sean más robustas en el extremo 3', que está unido finalmente a la cola poli A. Debido a su método de preparación, las bibliotecas de cDNA tienen cierta tendencia a estar infra-representadas en el extremo 5', particularmente para los transcritos de cDNA más largos. Sin embargo, esto tendrá un efecto mínimo (en todo caso) en el uso de una biblioteca de cDNA como fuente de UTR's 3' o UTR's 3' candidatas, dado que el extremo 3' de las bibliotecas de cDNA es típicamente el mejor representado con las secuencias más intactas y diversas.

Por supuesto, pueden existir también sitios diana de miRNA presentes en la UTR 5' de los genes o en otras posiciones. Por tanto, el uso de una combinación de oligo-dT y hexámeros aleatorios permite ventajosamente una mayor cobertura de sitios diana de miRNA candidatos por una biblioteca de cDNA así producida.

5 Dado que las bibliotecas de cDNA se utilizan tradicionalmente para el estudio de los polipéptidos codificados, es en sí mismo sorprendente que tales materiales puedan utilizarse como fuente de diversas UTR's o UTR's candidatas.

Opcionalmente, las UTR's 3' candidatas pueden seleccionarse por tamaños. Esto tiene la ventaja de optimizar el tamaño del ácido nucleico total. Esto tiene la ventaja adicional de permitir la optimización de las probabilidades de inclusión del mayor número posible de UTR's 3' intactas basándose en el conocimiento de los tamaños más comunes de UTR's 3' en el organismo o tejido de interés del que se derivan las UTR's 3'.

10 **Células Hospedadoras**

Los ensayos de la invención se llevan a cabo ventajosamente en (o sobre) células hospedadoras. Convenientemente, éstas son células eucariotas. Convenientemente, éstas son células de un organismo multicelular. Convenientemente, las células proceden de insectos o vertebrados. Cuando las células proceden de vertebrados, las mismas son convenientemente células de mamífero. Convenientemente, las células son 'cognadas' respecto al miRNA o a la UTR 3' que se estudia. Ser cognado significa preferiblemente que se deriva del mismo organismo. Esto tiene la ventaja de que la maquinaria de procesamiento celular, por ejemplo para procesamiento de los miRNAs o para traducción de los mRNAs, será común y proporcionará por tanto las condiciones biológicamente más relevantes para estudiar o testar la función de la UTR 3' del miRNA.

En algunas realizaciones, es deseable que las células hospedadoras sean diferentes de la fuente del miRNA y/o UTR 3' que se estudia. Un ejemplo de una aplicación de este tipo es cuando existen mRNAs endógenos que podrían interferir con o interactuar con la secuencia diana (v.g. UTR 3' o UTR 3' candidata) objeto del estudio. En esta realización, puede ser deseable utilizar células o líneas de células que sean de un organismo diferente del o de los organismos de los que se deriva el miRNA y/o la secuencia diana. Por ejemplo, cuando se estudian miRNAs humanos puede ser deseable utilizar células de insecto tales como células Sf9. De esta manera, puede ser posible evitar 'interferencia' o complicación del estudio o cribado por los miRNAs existentes naturalmente o endógenos. Por supuesto, es sencillo testar si existen o no mRNAs endógenos interferentes en las células o líneas de células de interés por introducción de ácido nucleico que lleva la o las secuencias diana en la célula o células y testado respecto a la expresión del o de los marcadores seleccionables. Si no se observa expresión alguna incluso en ausencia de la adición o introducción de miRNAs de interés, ello puede ser una indicación de que miRNAs existentes o endógenos están impidiendo o regulando en sentido decreciente la expresión de los marcadores seleccionables. Una observación de este tipo es una indicación de que es preciso abordar este problema antes de emprender un estudio o cribado significativo, por ejemplo por testado de una célula o línea de células alternativa hasta que se establezcan las condiciones para expresión fiable del o de los genes marcadores seleccionables en ausencia de miRNA exógeno. Esto es evidentemente una cuestión de rutina para el técnico experto dadas las orientaciones proporcionadas en esta memoria.

Convenientemente, las células hospedadoras contienen al menos los aparatos necesarios para procesamiento del miRNA y para la expresión de las proteínas. Una vez más, esto puede testarse fácilmente por introducción de uno o más ácidos nucleicos de la invención y monitorización de la expresión de los genes marcadores como se ha indicado arriba.

40 Células adecuadas incluyen células 3T3 tales como fibroblastos de ratón NIH 3T3 (aunque estas células expresan MIR10a y MIR130); células humanas HL60 o Jurkat (las cuales no expresan ventajosamente MIR10a o MIR130 en grado significativo); células HeLa humanas (que expresan ventajosamente MIR10a y MIR130 muy bajos); células Cos (que son ventajosamente transfectables con facilidad).

Las células NIH3T3 y HeLa tienen la ventaja adicional de ser fácilmente transfectables.

45 De modo muy conveniente, las células son células MCF7.

En el formato de biblioteca o cribado, las líneas de células pueden considerarse como 'autopurificadoras' en el sentido de que las UTR's no llegarán a pasar la primera tanda de cribado/selección si su miRNA se expresa endógenamente en las células hospedadoras utilizadas.

Son particularmente adecuadas las células o líneas de células indicadas en la sección de ejemplos.

50 **Aplicaciones Adicionales**

Los microRNA's juegan un papel en muchos procesos biológicos tales como diferenciación, angiogénesis, adhesión celular, proliferación celular y supervivencia, y juegan un papel importante en la hematopoyesis. Se ha demostrado también que los mismos juegan papeles importantes en el cáncer. Por consiguiente, la invención puede aplicarse ventajosamente en muchas áreas diferentes de la industria.

55 Se describe un ensayo funcional desarrollado para la identificación de dianas de microRNA que puede identificar

dianas múltiples para un microRNA específico en un procedimiento. Esto encuentra aplicación en campo en expansión del estudio del microRNA.

5 La adaptación del procedimiento de selección puede hacer ventajosamente utilizable esta invención en conexión con miRNA y/o dianas de miRNA de diversos organismos. Además, la identificación de dianas de microRNAs es importante en enfermedades tales como el cáncer en las que los microRNA's juegan papeles importantes. Las dianas identificadas pueden constituir nuevas dianas para desarrollo de moléculas pequeñas (v.g. BCR/ABL, glivec, y otras).

Adicionalmente, la invención proporciona uno o más plásmidos nuevos para la clonación de UTR's detrás de HSVTK/puro, por ejemplo como se muestra en la Figura 2.

10 La invención proporciona también nuevas fusiones de marcadores seleccionables.

Pueden existir también sitios diana de miRNA presentes dentro de la UTR 5' de los genes. Por tanto, el uso de una combinación de oligo-dT y hexámeros aleatorios puede permitir ventajosamente una mayor cobertura de sitios diana por la biblioteca de cDNA en comparación con el uso de oligo-dT solo.

Reguladores de RNA regulador

15 Utilizando una diana para un microRNA específico, el sistema puede utilizarse para identificar reguladores de este microRNA. Una población de células que expresen la secuencia diana (v.g. la UTR diana) enlazada a marcadores seleccionables (tales como la fusión TKpuro) y microRNA será resistente a ganciclovir y sensible a puromicina. Si se introduce luego una sustancia o biblioteca de cDNA en estas células y se selecciona en puromicina será posible identificar genes que regulen esta expresión de microRNAs, es decir genes o sustancias que impidan o inhiban la acción del miRNA y permitan o aumenten por tanto la expresión de marcadores seleccionables (que está reprimida en ausencia del gen o sustancia).

20 Dicho de otro modo, este sistema puede utilizarse para estudiar o cribar genes, productos químicos, moléculas pequeñas u otras entidades que regulen el RNA regulador tal como microRNA. Por ejemplo, si mirX regula la diana Y, entonces para identificar entidades o tratamientos que regulen en sentido decreciente la expresión de mirX, podría introducirse en las células la sustancia o gen (v.g. una biblioteca de cDNA o biblioteca de moléculas pequeñas). La regulación decreciente de mirX daría como resultado la expresión de un marcador seleccionable tal como puroTK y conferiría resistencia a la puromicina a estas células. Cuando la entidad de test es cDNA, la identificación del cDNA introducido revelará uno o más genes que regulen la expresión y/o función de mirX. Cuando la entidad es una molécula pequeña, tales bibliotecas de molécula pequeña pueden aplicarse ventajosamente en experimentos simples o agrupaciones de compuestos múltiples como es bien conocido en la técnica y está en muchos casos automatizado ventajosamente, v.g. por el uso de procesamiento robótico de las muestras.

Cribado fuera de la diana

35 Muchos RNAs reguladores, tales como moléculas siRNA, se encuentran en desarrollo o en pruebas clínicas. Las realizaciones de la invención pueden utilizarse para cribar estas moléculas de siRNA respecto a los efectos fuera de la diana del siRNA. Ésta es una aplicación y utilidad industrial adicional importante de este sistema.

40 En estas realizaciones, el sistema puede utilizarse para estudiar los efectos fuera de la diana del RNA regulador tal como RNA de interferencia pequeño (siRNA). Muchas moléculas de siRNA se encuentran en desarrollo en pruebas clínicas para desactivación de genes tales como oncogenes (v.g. BCL2) en el cáncer y/o genes mutantes implicados en otras enfermedades genéticas. Un problema con los siRNAs individuales son los efectos fuera de la diana debidos a la secuencia de siembra (secuencia hexámera en el extremo 5' del siRNA o microRNA). Es prácticamente imposible diseñar siRNA sin una secuencia de siembra que, excepto con respecto a la diana propuesta, esté ausente en el genoma humano. Esto es debido simplemente al tamaño del genoma humano y a la probabilidad de que una secuencia corta de este tipo (v.g. un 6-mero) sea única en el genoma. Podía esperarse que esta secuencia de siembra existiera centenares de veces en el genoma humano. El siRNA con una o más secuencias de siembra fuera de la diana podría actuar como un microRNA (sólo con homología parcial con la diana en lugar de 100% de homología como en el caso del siRNA) en estos sitios inadecuados o fuera de la diana. El sistema descrito en esta memoria podría utilizarse para testar moléculas de siRNA propuestas respecto a posibles efectos fuera de la diana. Podrían utilizarse adecuadamente bibliotecas de RNA de longitud total como fuente de secuencias diana de RNA regulador candidatas en los ácidos nucleicos de la invención. Esto tiene la ventaja de una mayor probabilidad de abarcar todas las posibles secuencias de siembra en comparación con cDNAs truncados u otras fuentes, aunque por supuesto los mismos podrían utilizarse igualmente en caso deseado.

La invención se describe a continuación mediante ejemplos. Estos ejemplos tienen por objeto ser ilustrativos, y no deben considerarse como limitantes de las reivindicaciones adjuntas.

Breve Descripción de las Figuras

55 La figura 1 muestra un diagrama del método o métodos de la invención.

- La figura 2 muestra un diagrama de un ácido nucleico de invención.
- La figura 3 muestra un diagrama de un ácido nucleico de invención.
- La figura 4 muestra un diagrama de un ácido nucleico de invención.
- La figura 5 muestra un diagrama de un ácido nucleico de invención.
- 5 La figura 6 muestra un gráfico de barras de la regulación decreciente de la expresión de luciferasa/MAFBUTR y una fotografía de la expresión de la proteína MAFB.
- Las figuras 7 y 8 muestran gráficos de barras de la sensibilidad a GCV el día 10.
- La figura 9 muestra un gráfico de barras de la expresión de mir-10a y mir-130a.
- La figura 10 muestra un gráfico de barras de TKZEO ganciclovir 7d.
- 10 La figura 11 muestra un gráfico de barras de TKZEO ganciclovir 13d.
- La figura 12 muestra un gráfico de barras de la sensibilidad a AZT el día 7.
- La figura 13 muestra la expresión de Mir-10a y mir-130a a partir de células MF7 transitorias (superior) y estables (inferior).
- La figura 14 muestra una fotografía de una biblioteca UTR representativa de cerebro de la invención.
- 15 La figura 15A muestra cDNA seleccionado por tamaños; la figura 15B muestra una biblioteca clonada digerida con Sfil.
- La figura 16 muestra el análisis PCR de la biblioteca.

Ejemplos

Ejemplo 1: Ácidos Nucleicos

- 20 Se construye un ácido nucleico que comprende los elementos contiguos siguientes dispuestos en la dirección 5' a 3': un promotor; un marcador seleccionable; un sitio de clonación para recepción de un segmento de ácido nucleico, comprendiendo dicho segmento una secuencia diana de miRNA candidata; y una señal de poliadenilación.
- Los elementos están dispuestos de tal manera que un transcrito dirigido por dicho promotor comprende dicho marcador seleccionable, dicha secuencia diana de miRNA candidata, y dicha señal de poliadenilación por este orden.
- 25

Ejemplo 2: Marcadores Seleccionables Duales

- Como se explica en esta memoria, el marcador seleccionable es una parte fundamental de la presente invención. En ciertas realizaciones, el marcador seleccionable puede comprender ventajosamente más de una actividad. Este ejemplo demuestra la producción de marcadores seleccionables con más de una actividad. En este ejemplo, ello se consigue por fusión de los ORFs para dos marcadores seleccionables individuales diferentes en un solo segmento de ácido nucleico. Esto da ventajosamente como resultado la producción de un solo polipéptido que comprende dos dominios de polipéptido diferentes, cada uno de los cuales tiene su actividad específica (seleccionable).
- 30 En este ejemplo, los dos marcadores individuales utilizados son HSVTK y PURO. Éstos se fusionan para formar un marcador seleccionable dual TK/PURO.
- 35 Se estudian los marcos de lectura abiertos de HSVTK y PURO. Se selecciona un punto de fusión adecuado con consideración a la naturaleza de los productos polipeptídicos a fin de maximizar las probabilidades de que se conserve su actividad en el producto fusionado. En esta etapa, puede tomarse una decisión acerca de si incluir o no un enlazador (v.g. una región enlazadora o un 'inmovilizador' u otra unión de este tipo) en la junta entre los dos polipéptidos. Se presta también atención a cuestiones prácticas tales como el escaneo de las secuencias de ácido nucleico para los uno o más sitios de reconocimiento por enzimas de restricción que podrían interferir con el procedimiento o con el uso de la fusión en la invención (v.g. los sitios Sfil, BstXI, u otras enzimas de restricción propuestas para ser utilizadas para la inserción de UTR en el ácido nucleico final de la invención deberían eliminarse ventajosamente en esta etapa). La eliminación de tales sitios puede realizarse adecuadamente por mutagénesis orientada o técnicas similares.
- 40
- 45 Se producen luego las secuencias de ácido nucleico y se unen en caso necesario. Esto puede realizarse por cualquier medio adecuado conocido en la técnica. Por ejemplo, ello puede hacerse por digestión con enzimas de restricción y ligación de los diferentes elementos para formar la fusión (con inclusión del relleno selectivo o terminación roma de cualesquiera fragmentos intermedios en caso requerido). Alternativamente, ello puede realizarse por amplificación PCR de los fragmentos deseados seguido por clonación/ligación en caso apropiado.

Alternativamente, la secuencia completa de ácido nucleico diseñada puede sintetizarse directamente en forma completa, por ejemplo por síntesis química.

En este ejemplo, se produce una fusión Hygro/TK. Esta fusión tiene la secuencia que se muestra en SEQ ID NO: 3.

Ejemplo 3: Marcador Dual Seleccionable HSVTK/PURO

5 En este ejemplo, los dos marcadores seleccionables se fusionan para producir un solo producto de traducción que comprende ambas actividades/polipéptidos.

En este ejemplo, los dos marcadores individuales utilizados son HSVTK y PURO. Éstos se fusionan para formar un marcador TK/PURO dual seleccionable.

10 Se estudian los marcos de lectura abiertos de HSVTK y PURO. Los marcadores se fusionan luego como se describe en el ejemplo 2.

El marcador seleccionable resultante se muestra en SEQ ID NO: 1. Éste es un marcador dual seleccionable. Esta es una fusión TK-PURO de acuerdo con la presente invención.

Ejemplo 4: Ácido Nucleico con Marcadores Duales Seleccionables

En este ejemplo, se incorporan dos marcadores seleccionables en el ácido nucleico de la invención

15 En este ejemplo, se produce un ácido nucleico con HSVTK/puro como marcador seleccionable.

Los dos marcadores seleccionables se fusionan para producir un solo producto de traducción que comprende ambas actividades/polipéptidos como en el ejemplo 3.

20 Esta secuencia de nucleótidos que codifica el marcador dual seleccionable se introduce luego en el ácido nucleico la invención después (es decir aguas abajo o 3') del promotor y antes (es decir aguas arriba o 5') del sitio para la inserción UTR 3'.

Ejemplo 5: Bibliotecas de UTR 3'

Se producen bibliotecas de UTR 3' de acuerdo con la presente invención.

25 Se prepara una biblioteca de UTR 3' proporcionando un ácido nucleico como se ha descrito arriba, tal como se describe en el ejemplo 1, e insertando en dicho sitio de clonación un ácido nucleico que comprende una secuencia diana de miRNA candidata. En este ejemplo, la secuencia diana de miRNA candidata es una UTR 3' o una UTR 3' candidata.

De modo más detallado, el ácido nucleico en el cual se insertan las UTR's 3' o UTR's 3' candidatas está constituido por el ácido nucleico del ejemplo 4. Específicamente, el ácido nucleico está constituido por el plásmido p3' UTR3 (véase la figura 2).

30 En este ejemplo, los segmentos de ácido nucleico que llevan las UTR's 3' o las UTR's 3' candidatas son o se derivan de cDNAs. En este ejemplo específico, los cDNAs se derivan de cerebro. El cerebro tiene el número máximo de transcritos singulares comparado con cualquier otro órgano. Esto permite ventajosamente la creación de bibliotecas con diversidad maximizada. Evidentemente, pueden utilizarse cDNAs de cualquier tejido, o puede utilizarse de hecho una mezcla de cDNAs de tejidos diferentes a fin de maximizar la diversidad.

35 Se utiliza una biblioteca de cDNA de cerebro humano cebada con oligo-dT (como se ha indicado arriba, el cerebro expresa el número máximo de miRNA's diferentes). En esta biblioteca de cDNA, los cDNAs se han clonado direccionalmente en dos sitios Sfil con salientes 3' diferentes (GGCCNNNNNGGCC).

40 Por término medio, una UTR 3' humana tiene una longitud aproximada de 1000 nucleótidos. Por esta razón, la biblioteca se digiere con Sfil y se selecciona opcionalmente por tamaños, es decir se aísla la fracción inferior a 1500 pares de bases para asegurar la captura de la mayoría de las UTR's 3'. Este cDNA se clona luego direccionalmente en el sitio Sfil del vector UTR de 3' aguas abajo de TKpuro.

De este modo, se produce una biblioteca de UTR 3' de acuerdo con la presente invención.

Ejemplo 6: Bibliotecas de AML

Se aplica la técnica del ejemplo 5 a la construcción de una biblioteca de UTR 3' específica de una enfermedad.

45 Las UTR's 3' (UTR's 3' candidatas) se derivan de una biblioteca de cDNA. En este ejemplo, dicha biblioteca se deriva de células de leucemia mieloide aguda.

Los cDNAs se seleccionan usualmente por tamaños. En este ejemplo, los mismos se seleccionan por tamaños con un tamaño máximo de aproximadamente 1500 nucleótidos.

Éste cDNA se clona luego direccionalmente en el sitio Sfil del vector UTR p3' aguas abajo de TKpuro.

Se construye así una biblioteca UTR 3' de acuerdo con la presente invención.

Ejemplo 7: Bibliotecas Basadas en Células

5 Se construye una biblioteca de plásmidos de acuerdo con el ejemplo 5 o el ejemplo 6 anterior y se introduce en gran escala en células hospedadoras. En este ejemplo, las células son células no humanas y la introducción de la biblioteca en las células se realiza como se describe en Mourtada et al 2005 (Mourtada-Maarabouni M, Kirkham L, Farzaneh F, Williams GT. Functional expression cloning reveals a central role for the receptor for activated protein kinase C 1 (RACK1) in T cell apoptosis. J Leukoc Biol. 2005.2:503).

10 Las células que contienen la biblioteca de plásmidos se seleccionan luego en presencia de puromicina a fin de que únicamente puedan desarrollarse las células que han capturado la biblioteca de plásmidos.

Las células se expanden luego mientras que se preserva la diversidad de la colección. Las células expandidas se agrupan luego. Se preservan después partes alícuotas de las células expandidas agrupadas para uso futuro, por ejemplo por congelación y almacenamiento a -196 °C en nitrógeno líquido.

15 En caso requerido, se descongelan las células y se devuelven al cultivo para uso en cribado/análisis. Puede aplicarse selección con puromicina en cualquier momento a fin de asegurar que únicamente se mantienen las células que albergan el plásmido diana. Una colección de células que comprenden la biblioteca de plásmidos de esta manera se considera como una biblioteca basada en células de acuerdo con la presente invención.

Ejemplo 8: Cribado

20 La invención proporciona instrumentos y métodos para identificación y validación de dianas en la regulación de los genes miRNA. Se proporcionan también ensayos funcionales para la identificación de dianas de miRNA, por ejemplo por cribado de bibliotecas.

Estudio de Selección (Estudio de Cribado)

25 En este ejemplo, se aplica un método nuevo de selección para la identificación de la regulación decreciente de proteínas debida a la fijación de miRNA a UTR's 3'. A este fin, se utilizan UTR's 3' clonadas aguas abajo de un gen de fusión HSVTK/Puro que, cuando se expresa, confiere resistencia a puromicina y sensibilidad a ganciclovir a las células. La regulación decreciente de la traducción debida a la fijación de miRNA a la UTR 3' convierte estas células en sensibilidad a puromicina y resistencia a ganciclovir (véase Fig.1 para vista de conjunto).

30 A fin de demostrar este método, se clonaron sitios diana de miRNA validados y las UTR's 3' de longitud total para los genes HOXA1 y MAFB aguas abajo de TKpuro en los sitios Sfil de p3' UTR (véase Fig. 2). HOXA1 y MAFB tienen integración conocida con los miRNAs mir-10a y mir-130a respectivamente (Garzon R, Pichiorri F, Palumbo T, et al. MicroRNA fingerprints during human megakaryocytopoiesis. PNAS 2006;103:5078-5083).

Se transfectan células de murino o de insecto con los plásmidos de expresión p3' UTR y se seleccionan en puromicina para obtener una población de células transfectadas.

35 El miRNA precursor (mir-10a, mir-130a; Ambion) y oligos de RNA de control enmarañados se transfectan luego y las células se expanden en presencia de ganciclovir para aislar clones en los cuales el miRNA ha regulado en sentido decreciente la expresión de la proteína TKpuro, convirtiendo estas células a resistencia a ganciclovir.

Las células supervivientes se clonan y se verifica la presencia de los sitios diana HOXA1 y MAFB o las UTR's por PCR y secuenciación.

40 Los niveles de expresión de TKpuro en presencia de los miRNAs pueden investigarse por transferencia Western para HSVTK utilizando anticuerpos disponibles comercialmente (Insight Biotechnology).

Cribado de Bibliotecas

45 Se produce una biblioteca de plásmidos de acuerdo con el ejemplo 5 anterior y se introduce en gran escala en células hospedadoras. En este ejemplo, las células son células no humanas y la introducción de la biblioteca en las células se revisa como describe en Mourtada et al 2005 (Mourtada-Maarabouni M, Kirkham L, Farzaneh F, Williams GT. Functional expression cloning reveals a central role for the receptor for activated protein kinase C 1 (RACK1) in T cell apoptosis. J Leukoc Biol. 2005.2:503).

Después de selección con puromicina, se introduce(n) el miRNA de interés y/o el o los controles. En este ejemplo, se introducen mir-10a, mir-130a u oligos enmarañados.

50 Se utiliza transfección utilizando liposomas disponibles comercialmente tales como Lipofectamina 2000, electroporación o cualquier otra forma de transducción.

Se dejan crecer luego las células que contienen la biblioteca en presencia de ganciclovir y los clones resistentes se testan en cuanto a la presencia de la UTR 3' de HOXA1 o MAFB en estos clones.

Este procedimiento identifica también cierto número de otras dianas para mir-10a y mir-130a. Éstas se verifican por análisis de transferencia Western de la expresión de TK/puro en estos clones. Se demuestra así que esta técnica de cribado de bibliotecas es un instrumento inestimable para la identificación y validación de dianas para miRNA's tanto conocidos como no identificados todavía.

Ejemplo 9: Cribado Fuera de la Diana

En este ejemplo, el RNA regulador de interés es siRNA para desactivación de un gen implicado en el cáncer de hígado. Convenientemente, éste puede direccionarse específicamente al hígado *in vivo*.

Para investigar los efectos fuera de la diana de este RNA regulador, podría testarse una biblioteca 3' UTR de cerebro o hígado acoplada a un marcador seleccionable como TKpuro.

Se introduce en las células El siRNA objeto de investigación.

Las secuencias diana candidatas de colonias resistentes a ganciclovir se someten luego a PCR y se secuencian. Si se recuperan genes distintos del gen diana X deseado, ello es indicativo de efectos fuera de la diana del RNA regulador. Éstos pueden evaluarse luego o estudiarse ulteriormente según sea apropiado.

Estos resultados ayudan a la decisión respecto a proseguir con o diseñar un RNA regulador diferente tal como siRNA.

Ejemplo 10: Cribado Ilustrativo de Bibliotecas

A) Se han transfectado MCF7 y MCF7mir130A con una biblioteca de UTR impurificada con 20% de MAFBUTR. Se seleccionan luego los mismos en zeocina y se destruyen todos los controles, obteniéndose muchas colonias. Se introduce mir130A en las células MCF7 transfectadas y se selecciona luego en puromicina (7-10 días), seleccionándose después en ganciclovir. Posteriormente se secuencian los clones.

B) Se hicieron dos transfecciones adicionales en MCF7 y MCF7mir130A que expresan de hecho mir130A. Dado que MCF7 no expresa naturalmente mir130A después de selección con zeocina, los clones recuperados deberían contener una MAFBUTR en aproximadamente el 20% de los clones. Sin embargo, en MCF7mir130A, la MAFBUTR debería estar silenciada, lo cual da como resultado la pérdida de resistencia a zeocina. Los clones recuperados después de selección con zeocina de esta segunda transfección en MCF7mir130A deberían carecer totalmente de inserciones MAFBUTR o exhibir un número muy pequeño de las mismas.

Se aísla luego el DNA a partir de una población mixta de células de ambas transfecciones y se someten a PCR las inserciones UTR (población mixta). Estas inserciones se clonan en el vector de clonación TA y los clones individuales se envían para secuenciación en formato de 96 pocillos. Se secuencian aproximadamente 48 clones de cada transfección.

Se efectúa luego el recuento de la frecuencia con que está presente la MAFBUTR en los clones de la transfección de MCF7 y MCF7mir130A. De este modo se demuestra el principio del procedimiento.

Al mismo tiempo, el procedimiento puede seguirse también con selección de GCV.

Ejemplo 11: Cribado de Bibliotecas Adicionales

Se han transfectado células MCF7 y células MCF7(mir130). MCF7 no expresa mir130, y en MCF7(mir130) se ha introducido mir130 y se ha comprobado la expresión de mir130 por qPCR.

En un experimento en pequeña escala (10 placas de cada una) se ha introducido una biblioteca que se había clonado en el vector p3'TKzeo. La biblioteca se impurificó con 20% de MAFBUTR que es una diana para mir130.

Se seleccionaron ambas líneas de células en 1 mg/ml de zeocina, lo cual dio como resultado 200-300 colonias para cada línea de células. Debido a la ausencia de mir130 en MCF7, la MAFBUTR no debería estar regulada en sentido decreciente. La regulación decreciente de la MAFBUTR debería dar como resultado la ausencia de proteína TKzeo, lo cual debería conducir a la muerte de estas células en zeocina. En MCF7(mir130) la MAFBUTR debería estar regulada en sentido decreciente, lo cual debería dar como resultado la muerte de las células que contuviesen la MAFBUTR.

En conclusión; en las células MCF7 después de selección en zeocina, aproximadamente el 20% de los clones deberían contener la MAFBUTR, mientras que en MCF7(mir130) el porcentaje debería ser mucho menor. Para investigar esto se diseñaron cebadores que amplificarían sólo el DNA de MAFBUTR presente en la biblioteca y no el MAFB endógeno. Los resultados se presentan en la figura 16. Existe una diferencia aproximada de 10X en la cantidad de DNA de MAFBUTR entre las dos células diferentes, lo cual es una indicación clara de la validez del procedimiento.

Se amplificaron también por PCR las UTR's completas presentes en las dos células, y se clonaron estos productos PCR en plásmidos. Se han enviado 24 clones de cada célula para secuenciación. Existe todavía una reducción de 4 veces en el número de clones que contienen MAFB.

5 Nota explicativa: Esto puede ser una subestimación. Sin desear quedar ligados por la teoría, es posible que las preparaciones de plástido no sean igualmente puras. La MAFBUTR utilizada para impurificación era una Maxiprep™ de Sigma™ y la preparación de la biblioteca era una Gigaprep™ de Qiagen™. La preparación de Sigma™ puede ser más pura, dando como resultado más células transfectadas. Evidentemente, esto puede ser optimizado por el técnico experto purificando la preparación de la biblioteca de acuerdo con cualquier método adecuado conocido en la técnica.

10 Adicionalmente, se han introducido ahora mir 10, mir 130 y un RNA de horquilla corta (shRNA) contra MABF UTR en las células MCF7 que contienen la biblioteca. Estas células se pondrán bajo selección con zeocina, lo cual debería eliminar la MABF UTR de las células que expresen mir130 o el shRNA, pero no en las células que expresen mir10. En un experimento separado, estas células pueden ponerse bajo selección con ganciclovir, que debería rescatar la MABF UTR de las células que expresen mir130 o el shRNA, pero no de las células que expresen mir10.

15 **Ejemplo 12: Ensayo Funcional**

La selección es más potente que el cribado convencional (en el que los desaciertos se mantienen presentes en lugar de perderse o eliminarse por selección); así, se empleó un cribado basado en selección como sigue:

Selección de Fármaco

Selección positiva/negativa

20 Fusionar la proteína de un marcador seleccionable (v.g., puro, hygro, zeo u otro adecuado) con una profármaco-convertasa (v.g. HSVtk - GCV, citosina-desaminasa - 5FC, timidilato-quinasa - AZT u otra adecuada).

Fusionar GFP-puro para cribado y clasificación FACS.

Ejemplo 13: Biblioteca 3' UTR

Se construye una biblioteca de acuerdo con lo siguiente:

25 La longitud mediana de la 3' UTR es 1kB

Material de partida: Biblioteca de cDNA de cerebro

Cebado con oligo-dT: la mayoría de las inserciones contendrán al menos 3' UTR parcial

Clonación direccional utilizando sitios Sfil diferentes

Tamaño seleccionado > 2,5 kilobases.

30 **Ejemplo 14: Cribado**

El HoxA1 y MAFB se regulan en sentido decreciente por mir10a y mir130 respectivamente.

Se clonan HoxA1 y MAFBUTR's y los sitios diana predichos en un vector de selección positiva/negativa y un vector de luciferasa.

35 MAFB es una diana de mir-130a (véase Figura 6); se ha establecido también la regulación decreciente de HOXA1 por mir10a.

La sensibilidad a GCV el día 10 se muestra en las Figuras 7 y 8.

La expresión de mir-10 y mir-130a se muestra en la figura 9.

Se muestra TKZEO ganciclovir'7d' en la figura 10, y'13d' en la figura 11.

La sensibilidad a AZT el día 7 se muestra en la figura 12.

40 La expresión de mir10a y mir130a por las células MCF7 transitoria (superior) y estable (inferior) se muestra en la figura 13.

Ejemplo 15: Construcción Detallada de la Biblioteca

La biblioteca se construye como sigue:

Tamaño seleccionado del cDNA digerido con Sfil > 2,5 kilobases

Se clona en TKzeo Sfil

15 µl ligación de 1 µg TKzeo + 200 ng biblioteca transformada 1,0 µl

Se extienden placas 1 µl y 100 µl a partir de 1000 µl

> 500 colonias de 1 µl

5 500 × 1000 × 15 = 7,5 millones

50 minipreparaciones, 50 inserciones diferentes

Se recogen ± 600.000 clones independientes

7,5 mg de Giga prep

Una biblioteca de UTR's de cerebro representativa de acuerdo con la presente invención se muestra en la figura 14.

10 El cDNA seleccionado por tamaños y la biblioteca clonada digerida con Sfil se muestran en las figuras 15A y 15B respectivamente. La biblioteca se impurificó con 20% del plásmido MAFB-UTR.

Cribado de la Selección:

Se transfectó en células MCF7 y células MCF7 que expresaban mir130A.

Las células transfectadas se seleccionaron con zeocina

15 (~ 2000 colonias).

Se aisló el DNA genómico y se determinó la cantidad de plásmido MAFB-UTR por qPCRPCR de las UTR's presentes en la biblioteca MCF7+ y la biblioteca MCF7mir130A+ y clonación en Topo TA para secuenciación de los clones individuales.

La biblioteca MCF7+ con adición de 20% de MAFB se transfectó con mir10A, mir130A y shRNA contra MAFB.

20 Selección en zeocina (reducción en MAFB).

Selección en ganciclovir (enriquecimiento en MAFB e identificación de las dianas mir10A y mir130A).

25 Aunque la presente invención se ha descrito en conexión con realizaciones específicas preferidas, debe entenderse que la invención tal como se reivindica no debe considerarse indebidamente limitada a dichas realizaciones específicas. De hecho, debe entenderse que diversas modificaciones de los modos de realización descritos para realización de la invención que son evidentes para los expertos en la técnica están dentro del alcance de las reivindicaciones siguientes.

Listado de Secuencias

SEQ ID NO: 1

Secuencia de ácido nucleico de la fusión TK-PURO

ATGGCCTCGTACCCCGGCCATCAACACGCGTCTGCGTTGACCAGGCTGCGCGTTCTCGCGGCCATAGC
AACCGACGTACGGCGTTGCGCCCTCGCCGGCAGCAAGAAGCCACGGAAGTCCGCCCGGAGCAGAAAATG
CCCACGCTACTGCGGGTTTATATAGACGGTCCCCACGGGATGGGGAAAACCACCACCACGCAACTGCTG
GTGGCCCTGGGTTGCGCGGACGATATCGTCTACGTACCCGAGCCGATGACTTACTGGCGGGTGTGGGG
GCTTCCGAGACAATCGCGAACATCTACACCACACAACACCGCCTCGACCAGGGTGAGATATCGGCCGGG
GACGCGGCGGTGGTAATGACAAGCGCCAGATAACAATGGGCATGCCTTATGCCGTGACCAGCGCGTT
CTGGCTCCTCATATCGGGGGGAGGCTGGGAGCTCACATGCCCGCCCCCGGCCCTCACCTCATCTTC
GACCGCCATCCCATCGCCGCCCTCCTGTGCTACCCGGCCGCGCGGTACCTTATGGGCAGCATGACCCCC
CAGGCCGTGTGGCGTTTCGTGGCCCTCATCCCGCCGACCTTGCCCGGCACCAACATCGTGCTTGGGGCC
CTTCCGGAGGACAGACACATCGACCGCCTGGCCAAACGCCAGCGCCCCGGCGAGCGGCTGGACCTGGCT
ATGCTGGCTGCGATTGCGCCGCTTTACGGGCTACTTGCCAATACGGTGCGGTATCTGCAGTGCGGCGGG
TCGTGGCGGGAGGACTGGGGACAGCTTTCGGGGACGGCCGTGCCGCCACAGGTTGCCGAGCCCCAGAGC
AACCGGGCCACGACCCCATATCGGGGACAGCTTATTTACCCTGTTTCGGGCCCCGAGTTGCTGGCC
CCCAACGGCGACCTGTATAACGTGTTTGCCTGGGCCTTGACGTCTTGGCCAAACGCCCTCCGTTCCATG
CACGTCTTTTATCCTGGATTACGACCAATCGCCCCGCGGCTGCCGGACGCCCTGCTGCAACTTACCTCC
GGGATGGTCCAGACCCACGTACCACCCCGGCTCCATACCGACGATATGCGACCTGGCGCGCACGTTT
GCCCGAGAAATGAAGCTTACCATGACCGAGTACAAGCCACGGTGCGCCTCGCCACCCGCGACGACGTC
CCCAGGGCCGTACGCACCCTCGCCGCGCGTTGCGCGACTACCCCGCCACGCGCCACACCGTCGATCCG
GACCGCCACATCGAGCGGGTCAACGAGCTGCAAGAACTTCTCCTCACGCGCGTTCGGGCTCGACATCGGC
AAGGTGTGGGTCGCGGACGACGGCGCCGCGGTGGCGGTCTGGACCACGCCGGAGAGCGTCGAAGCGGGG
GCGGTGTTCGCCGAGATCGGCCCGCGCATGGCCGAGTTGAGCGGTTCGCCGTGGCCGCGCAGCAACAG
ATGGAAGGCCCTCCTGGCGCCGACCGGCCAAGGAGCCCGCGTGGTTCTTGGCCACCGTCGGCGTCTCG
CCCACCACAGGGCAAGGTTCTGGGCAGCGCCGTGCTGCTCCCCGAGTGGAGGCGGCGAGCGCGCC
GGGGTGCCCGCTTCTTGGAGACCTCCGCGCCCGCAACCTCCCTTCTACGAGCGGCTCGGCTTACC
GTCACCGCCGACGTGAGGTGCCCGAAGGACCGCGCACCTGGTGCATGACCCGCAAGCCCGTGCCTGA

5

SEQ ID NO: 2

Secuencia de ácido nucleico de la cadena principal del plásmido

GCTAGCATCGATAAGAATTCCGGATCCTTAGGCCATTAAGGCCGGCCGCTCGGCCCACTTCG
TGGGGTACCGAGCTCGAATTCAGTGGCCGTCGTTTACAACGTCGTGACTGGGAAAACCTGGCGTTAC
CCAACTTAATCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGCCCCGCACCGA
TCGCCCTTCCCAACAGTTGCGTGGCCGAGGAGCAGGACTGACACGTGCTACGAGATTTTCGATTCCACCG
CCGCTTCTATGAAAGGTTGGGCTTCGGAATCGTTTTCCGGGACGCCGGCTGGATGATCCTCCAGCGCG
GGGATCTCATGCTGGAGTTCTTCGCCACCCCAACTGTTTTATTGCAGCTTATAATGGTTACAAATAAA
GCAATAGCATCACAAATTTACAAATAAAGCATTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAC
TCATCAATGTATCTTATCATGTCTGTATACCGTCGACCTTAGCTAGAGCTTGGCGTAATCATGGTCAT
AGCTGTTTTCTGTGTGAAATTGTTATCCGCTCACAAATCCACACAACATACGAGCCGGAAGCATAAAGT
GTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCC
AGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGCGGTTTTCGTA
TTGGGCGCTCTTCCGCTTCTCGCTCACTGACTCGCTGCGCTCGGTGCTTCGGCTGCGGCGAGCGGTAT
CAGCTCACTCAAAGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAG
CAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAGGCCGCGTTCGTTGGCGTTTTTCCATAGGCTCCGCC
CCCCTGACGAGCATCACAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGAT

ACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACC
 TGTCGCCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCAATGCTCACGCTGTAGGTATCTCAGTTCCG
 TGTAGGTCGTTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTTCAGCCCGACCCTGCGCCTTAT
 CCGGTAACCTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTA
 ACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCT
 AACTAGAAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTA
 GCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTTGTTTTGCAAGCAGCAGATTACGC
 GCAGAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAAACGAAA
 ACTCACGTTAAGGGATTTTGGTCAAGATTATCAAAAAGGATCTTACCTAGATCCTTTTAAATTAAA
 AATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCA
 GTGAGGCACCTATCTCAGCGATCTGTCTATTTTCGTTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGA
 TAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGTGCAATGATACCGCGAGACCCACGCTCAC
 CGGCTCCAGATTTATCAGCAATAAACAGCCAGCCGGAAGGGCCGAGCGCAGAAAGTGGTCTGCAACTT
 TATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTGCCAGTTAATAGTT
 TGCACAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTACGCTCGTCTTTGGTATGGCTTCATCA
 GCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCATGTTGTGCAAAAAAGCGTTAGCTCCT
 TCGGTCTCCGATCGTTGTGAGAAGTAAGTTGGCCGAGTGTATCACTCATGGTTATGGCAGCACTGC
 ATAATCTCTTACTGTATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAAGTACTCAACCAAGTCA
 TCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCAC
 ATAGCAGAACTTTAAAGTGTCTATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTAC
 CGTGTGTGAGATCCAGTTCCGATGTAACCCACTCGTGCACCCAAGTATCTTCAGCATCTTTTACTTTCA
 CCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAAATGCCGCAAAAAGGGAATAAGGGCGACACGGA
 AATGTTGAATACTCATACTCTTCCCTTTTCAATATTATTGAAGCATTATCAGGGTTATTGTCTCATGA
 GCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAG
 TGCCACCTGACGTCGACGGATCGGGAGATCTCCCGATCCCCATGGTTCGACTCTCAGTACAATCTGCTC
 TGATGCCGCATAGTTAAGCCAGTATCTGCTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTAGTGCCGAG
 CAAAATTTAAGCTACAACAAGGCAAGGCTTGACCGACAATTGCATGAAGAATCTGCTTAGGGTTAGGCG
 TTTTGCCTGCTTCGCGATGTACGGGCCAGATATACGCGTTGACATTGATTATTGACTAGTTATTAATA
 GTAATCAATTACGGGGTCATTAGTTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAA
 TGGCCCGCTGGCTGACCGCCCAACGACCCCCGCCATTGACGTCAATAATGACGTATGTTCCCATAGT
 AACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGACTATTTACGGTAAACTGCCACTTGGCAGT
 ACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCTGGCA
 TTATGCCAGTACATGACCTTATGGGACTTTCCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTAT
 TACCATGGTGTATGCGTTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTACTCACGGGGATTTCC
 AAGTCTCCACCCCATGACGTCAATGGGAGTTTGTGTTTGGCACCAAAAATCAACGGGACTTTCCAAAATG
 TCGTAACAACCTCCGCCCCATTGACGCAAAATGGGCGGTAGGCGGTGACGGTGGGAGGTCTATATAAGCAG
 AGCTCTCGGCTAACTAGAGAACCCACTGCTTACTGGCTTATCGAAATTAATACGACTCACTATAGGGA
 GACCCAAGCTG

SEQ ID NO: 3

Secuencia de ácido nucleico de la fusión Hygro/TK

ATGGGTAAAAAGCCTGAACTCACCGCGACGTCTGTCGAGAAGTTTCTGATCGAAAAGT
TCGACAGCGTCTCCGACCTGATGCAGCTCTCGGAGGGCGAAGAATCTCGTGCTTTTCAG
CTTCGATGTAGGAGGGCGTGGATATGTCCTGCGGGTAAATAGCTGCGCCGATGGTTTC
TACAAAGATCGTTATGTTTATCGGCACTTTGCATCGGCCGCGCTCCCGATTCCGGAAG
TGCTTGACATTGGGAATTCAGCGAGAGCCTGACCTATTGCATCTCCCGCCGTGCACA
GGGTGTCACGTTGCAAGACCTGCCTGAAACCGAACTGCCCGCTGTTCTGCAGCCGGTC
GCGGAGGCCATGGATGCGATCGCTGCGGCCGATCTTAGCCAGACGAGCGGGTTCGGCC
CATTGCGACCGCAAGGAATCGGTCAATACACTACATGGCGTGATTTTCATATGCGCGAT
TGCTGATCCCCATGTGTATCACTGGCAAACCTGTGATGGACGACACCGTCAGTGCGTCC
GTCGCGCAGGCTCTCGATGAGCTGATGCTTTGGGCCGAGGACTGCCCGAAGTCCGGC
ACCTCGTGCACGCGGATTTCCGGCTCCAACAATGTCTGACGGACAATGGCCGCATAAC
AGCGGTCAATGACTGGAGCGAGGCGATGTTCCGGGGATTCCCAATACGAGGTGCGCAAAC
ATCTTCTTCTGGAGGCCGTGGTTGGCTTGTATGGAGCAGCAGACGCGCTACTTCGAGC
GGAGGCATCCGGAGCTTGCAAGATCGCCGCGGCTCCGGGGCGTATATGCTCCGCATTGG
TCTTGACCAACTCTATCAGAGCTTGGTTGACGGCAATTCGATGATGCAGCTTGGGCG
CAGGGTCGATGCGACGCAATCGTCCGATCCGGAGCCGGGACTGTGCGGCGTACACAAA
TCGCCCCGAGAAGCGCGGCCGTCTGGACCGATGGCTGTGTAGAAGTCGCGTCTGCGTT
CGACCAGGCTGCGCGTTCTCGCGGCCATAGCAACCGACGTACGGCGTTGCGCCCTCGC
CGGCAGCAAGAAGCCACGGAAGTCCGCCCCGAGCAGAAAATGCCACGCTACTGCGGG
TTTATATAGACGGTCCCCACGGGATGGGGAAAACCACCACCACGCAACTGCTGGTGGC
CCTGGGTTCGCGCGACGATATCGTCTACGTACCCGAGCCGATGACTTACTGGCGGGTG
CTGGGGGCTTCCGAGACAATCGCGAACATCTACACCACACAACACCGCCTCGACCAGG
GTGAGATATCGGCCGGGGACGCGGCGGTGGTAATGACAAGCGCCAGATAACAATGGG
CATGCCTTATGCCGTGACCGACGCGCTTCTGGCTCCTCATATCGGGGGGGAGGCTGGG
AGCTCACATGCCCCGCCCCCGGCCCTCACCTCATCTTCGACCGCCATCCCATCGCCG
CCCTCCTGTGCTACCCGGCCGCGCGGTACCTTATGGGCAGCATGACCCCCAGGCCGT
GCTGGCGTTTCGTGGCCCTCATCCCGCCGACCTTGCCCGGCACCAACATCGTGCTTGGG
GCCCTTCCGGAGGACAGACACATCGACCGCCTGGCCAAACGCCAGCGCCCCGGCGAGC
GGCTGGACCTGGCTATGCTGGCTGCGATTGCGCGGTTTACGGGCTACTTGCCAATAC
GGTGCGGTATCTGCAGTGCGGCGGGTCTGGCGGGAGGACTGGGGACAGTTTTCGGGG
ACGGCCGTGCCGCCAGGGTGCCGAGCCCCAGAGCAACGCGGGCCACGACCCCAT
TCGGGGACACGTTATTTACCCTGTTTCGGGGCCCCGAGTTGCTGGCCCCAACGGCGA
CCTGTATAACGTGTTTGCCTGGGCCTTGACGTCTTGGCCCAAACGCCTCCGTTCCATG
CACGTCTTTATCCTGGATTACGACCAATCGCCCGCCGGCTGCCGGGACGCCCTGCTGC
AACTTACCTCCGGGATGGTCCAGACCCACGTACCACCCCCGGCTCCATACCGACGAT
ATGCGACCTGGCGCGCACGTTTGCCCGAGAAATGAAGCTTCGATAA

SEQ ID NO: 4

5 Secuencia de ácido nucleico de la fusión TKzeo

ATGGCTTCGTACCCCGGCCATCAACACGCGTCTGCGTTCGACCAGGCTGCGCG
TTCTCGCGGCCATAGCAACCGACGTACGGCGTTGCGCCCTCGCCGGCAGCAAGAAGCC
ACGGAAGTCCGCCCGGAGCAGAAAATGCCACGCTACTGCGGGTTTATATAGACGGTC
CCCACGGGATGGGGAAAACCACCACCACGCAACTGCTGGTGGCCCTGGGTTCGCGCGA

CGATATCGTCTACGTACCCGAGCCGATGACTTACTGGCGGGTGCTGGGGGCTTCCGAG
ACAATCGCGAACATCTACACCACACAACACCCGCTCGACCAGGGTGAGATATCGGCCG
GGGACGCGGGCGGTGTAATGACAAGCGCCAGATAACAATGGGCATGCCTTATGCCGT
GACCGACGCCGTTCTGGCTCCTCATATCGGGGGGAGGCTGGGAGCTCACATGCCCCG
CCCCGGCCCTCACCTCATCTTCGACCGCCATCCCATCGCCGCCCTCCTGTGCTACC
CGGCCGCGCGGTACCTTATGGGCAGCATGACCCCCAGGCCGTGCTGGCGTTCGTGGC
CCTCATCCCGCCGACCTTGCCCGGCACCAACATCGTGCTTGGGGCCCTTCCGGAGGAC
AGACACATCGACCGCTGGCCAAACGCCAGCGCCCCGGCGAGCGGCTGGACCTGGCTA
TGCTGGCTGCGATTGCGCGGTTTACGGGCTACTTGCCAATACGGTGCGGTATCTGCA
GTGCGGGCGGGTTCGTGGCGGGAGGACTGGGGACAGCTTTCGGGGACGGCCGTGCCGCC
CAGGGTGCCGAGCCCCAGAGCAACCGGGGCCACGACCCCATATCGGGGACACGTTAT
TTACCCTGTTTCGGGCCCCGAGTTGCTGGCCCCAACGGCGACCTGTATAACGTGTT
TGCCTGGGCCTTGGACGTCTTGGCCAAACGCCTCCGTTCCATGCACGTCTTATCCTG
GATTACGACCAATCGCCCGCCGGCTGCCGGGACGCCCTGCTGCAACTTACCTCCGGGA
TGGTCCAGACCCACGTCACCACCCCCGGCTCCATACCGACGATATGCGACCTGGCGCG
CACGTTTGGCCGAGAGATGATCAGCGGAGCTAATGGCGTCATGGCCAAGTTGACCAGT
GCCGTTCCGGTGCTCACCGCGCGGACGTCGCCGGAGCGGTTCGAGTTCTGGACCGACC
GGCTCGGGTTCTCCCGGGACTTCGTGGAGGACGACTTCGCCGGTGTGGTCCGGGACGA
CGTGACCCTGTTTCATCAGCGCGGTCCAGGACCAGGTGGTGCCGGACAACACCCTGGCC
TGGGTGTGGGTGCGCGGCCTGGACGAGCTGTACGCCGAGTGGTCGGAGGTCGTGTCCA
CGAACTTCCGGGACGCCTCCGGGCGGCCATGACCGAGATCGGCGAGCAGCCGTGGGG
GCGGGAGTTCGCCCTGCGCGACCCGGCCGCAACTGCGTGCACTTCGTGGCCGAGGAG
CAGGACTGA

SEQ ID NO: 5

p3'HYTK

CCTAGGCTTTTGCAAAAAGCTTGGCCACATGGGTAAAAAGCCTGAA
CTCACCGCGACGTCTGTCGAGAAGTTTCTGATCGAAAAGTTCGACAGCGTC
TCCGACCTGATGCAGCTCTCGGAGGGCGAAGAATCTCGTGCTTTCAGCTTC
GATGTAGGAGGGCGTGGATATGTCCTGCGGGTAAATAGCTGCGCCGATGG
TTTCTACAAAGATCGTTATGTTTATCGGCACTTTGCATCGGCCGCGCTCCCG
ATTCCGGAAGTGCTTGACATTGGGGAATTCAGCGAGAGCCTGACCTATTGC
ATCTCCCGCCGTGCACAGGGTGTACGTTGCAAGACCTGCCTGAAACCGAA

CTGCCCCTGTTCTGCAGCCGGTCGCGGAGGCCATGGATGCGATCGCTGCG
 GCCGATCTTAGCCAGACGAGCGGGTTCGGCCCATTCGGACCGCAAGGAAT
 CGGTCAATACACTACATGGCGTGATTTTCATATGCGCGATTGCTGATCCCCA
 TGTGTATCACTGGCAAACCTGTGATGGACGACACCGTCAGTGCGTCCGTGCG
 GCAGGCTCTCGATGAGCTGATGCTTTGGGCCGAGGACTGCCCGAAGTCCG
 GCACCTCGTGCACGCGGATTTTCGGCTCCAACAATGTCCTGACGGACAATGG
 CCGCATAACAGCGGTCATTGACTGGAGCGAGGCGATGTTTCGGGGATTCCC
 AATACGAGGTCGCCAACATCTTCTTCTGGAGGCCGTGGTTGGCTTGTATGG
 AGCAGCAGACGCGCTACTTCGAGCGGAGGCATCCGGAGCTTGCAGGATCG
 CCGCGGCTCCGGGCGTATATGCTCCGCATTGGTCTTGACCAACTCTATCAG
 AGCTTGGTTGACGGCAATTTTCGATGATGCAGCTTGGGCGCAGGGTTCGATGC
 GACGCAATCGTCCGATCCGGAGCCGGACTGTCGGGCGTACACAAATCGC
 CCGCAGAAGCGCGGCCGTCTGGACCGATGGCTGTGTAGAAGTCGCGTCTG
 CGTTCGACCAGGCTGCGCGTTCTCGCGGCCATAGCAACCGACGTACGGCGT
 TGCGCCCTCGCCGGCAGCAAGAAGCCACGGAAGTCCGCCCGGAGCAGAAA
 ATGCCACGCTACTGCGGGTTTATATAGACGGTCCCCACGGGATGGGGAA
 AACCACCACCACGCAACTGCTGGTGGCCCTGGGTTCGCGCGACGATATCGT
 CTACGTACCCGAGCCGATGACTTACTGGCGGGTGCTGGGGGCTTCCGAGAC
 AATCGCGAACATCTACACCACACAACACCGCCTCGACCAGGGTGAGATAT
 CGGCCGGGGACGCGCGGTGGTAATGACAAGCGCCAGATAACAATGGGC
 ATGCCTTATGCCGTGACCGACGCGTTCTGGCTCCTCATATCGGGGGGGAG
 GCTGGGAGCTACATGCCCGCCCCCGGCCCTCACCTCATCTTCGACCGC
 CATCCATCGCCGCCCTCCTGTGCTACCCGGCCGCGCGGTACCTTATGGGC
 AGCATGACCCCCAGGCCGTGCTGGCGTTCGTGGCCCTCATCCCGCCGACC
 TTGCCCGGCACCAACATCGTGCTTGGGGCCCTTCCGGAGGACAGACACATC
 GACCGCCTGGCCAAACGCCAGCGCCCCGGCGAGCGGCTGGACCTGGCTAT
 GCTGGCTGCGATTTCGCCGCTTTACGGGCTACTTGCCAATACGGTGCGGTA
 TCTGCAGTGCGGCGGGTTCGTGGCGGGAGGACTGGGGACAGCTTTCGGGGA
 CGGCCGTGCCGCCCCAGGGTGCCGAGCCCCAGAGCAACGCGGGCCACGA
 CCCCATATCGGGGACACGTTATTTACCCTGTTTCGGGCCCCCGAGTTGCTG
 GCCCCAACGGCGACCTGTATAACGTGTTTGCCTGGGCCTTGACGTCTTGG
 CCCAACGCCTCCGTTCCATGCACGTCTTTATCCTGGATTACGACCAATCG

CCCGCCGGCTGCCGGGACGCCCTGCTGCAACTTACCTCCGGGATGGTCCAG
ACCCACGTCAACCACCCCGGCTCCATACCGACGATATGCGACCTGGCGCGC
ACGTTTGCCCGAGAAATGAAGCTTCGATAAGAATTCCGGATCCTTAGGCCA
TTAAGGCCGGCCGCCTCGGCCCACTTCGTGGGGTACCGAGCTCGAATTCAC
TGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCTGGCGTTACCCAAC
TTAATCGCCTTGCAGCACATCCCCCTTCGCCAGCTGGCGTAATAGCGAAG
AGGCCCGCACCGATCGCCCTTCCAACAGTTGCGTGGCCGAGGAGCAGGA
CTGACACGTGCTACGAGATTCGATTCCACCGCCGCCTTCTATGAAAGGTT
GGGCTTCGGAATCGTTTTCCGGGACGCCGGCTGGATGATCCTCCAGCGCGG
GGATCTCATGCTGGAGTTCTTCGCCACCCCAACTTGTTTATTGCAGCTTAT
AATGGTTACAAATAAAGCAATAGCATCACAAATTCACAAATAAAGCATT
TTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTAT
CATGTCTGTATACCGTCGACCTCTAGCTAGAGCTTGGCGTAATCATGGTCA
TAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATAC
GAGCCGGAAGCATAAAGTGTAAGCCTGGGGTGCCTAATGAGTGAGCTAA
CTCACATTAATTGCGTTGCGCTCACTGCCCGCTTCCAGTCGGGAAACCTG
TCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTT
GCGTATTGGGCGCTCTTCCGCTTCTCGCTCACTGACTCGCTGCGCTCGGTC
GTTCCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTT
ATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAAGGC
CAGCAAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCA
TAGGCTCCGCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGA
GGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGA
AGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGT
CCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCAATGCTCACGCTGTAG
GTATCTCAGTTCGGTGTAGGTCGTTTCGCTCCAAGCTGGGCTGTGTGCACGA
ACCCCCGTTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGA
GTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTA
ACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAG
TGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCT
CTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGC
AAACAAACCACCGCTGGTAGCGGTGGTTTTTTTTGTTTGAAGCAGCAGATT

ACGCGCAGAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGG
 GTCTGACGCTCAGTGGAACGAAAACCTCACGTTAAGGGATTTTGGTCATGAG
 ATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTA AAAATGAAGTTT
 TAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATG
 CTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTTCGTTCCATCCATA
 GTTGCCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCA
 TCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCA
 GATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGG
 TCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCT
 AGAGTAAGTAGTTCGCCAGTTAATAGTTTTCGCGCAACGTTGTTGCCATTGCT
 ACAGGCATCGTGGTGTACGCTCGTTCGTTTGGTATGGCTTCATTCAGCTCC
 GGTTCCCAACGATCAAGGCGAGTTACATGATCCCCATGTTGTGCAAAAAA
 GCGGTTAGTCTCCTTCGGTCTCCGATCGTTGTGAGAAGTAAGTTGGCCGCA
 GTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGC
 CATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCT
 GAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGG
 GATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAA
 ACGTTCTTCGGGGCGAAAACCTCTCAAGGATCTTACCGCTGTTGAGATCCAG
 TTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTC
 ACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAAATGCCGCAAAAAA
 GGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTCA
 ATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTT
 GAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCG
 AAAAGTGCCACCTGACGTCGACGGATCGGGAGATCTCCCGATCCCCTATG
 GTCGACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATCT
 GCTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTAGTGCGCGAGCAAAATTTA
 AGCTACAACAAGGCAAGGCTTGACCGACAATTGCATGAAGAATCTGCTTA
 GGGTTAGGCGTTTTGCGCTGCTTCGCGATGTACGGGCCAGATATACGCGTT
 GACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAG
 TTCATAGCCCATATATGGAGTTCGCGGTTACATAACTTACGGTAAATGGCC
 CGCCTGGCTGACCGCCCAACGACCCCCGCCATTGACGTCAATAATGACGT
 ATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGG

ACTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGC
CAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATT
ATGCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACG
TATTAGTCATCGCTATTACCATGGTGTATGCGGTTTTGGCAGTACATCAATG
GGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTG
ACGTCAATGGGAGTTTGTGGTGGCACAAAATCAACGGGACTTTCAAAAT
GTCGTAACAACCTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGG
TGGGAGGTCTATATAAGCAGAGCTCTCTGGCTAACTAGAGAACCCACTGCT
TACTGGCTTATCGAAATTAATACGACTCACTATAGGGAGACCCAAGCTGGC
TAG

SEQ ID NO: 6

p3'TKPUR

GCTAGCTTATCGCATGGCCTCGTACCCCGGCCATCAACACGCGTCTG
CGTTCGACCAGGCTGCGCGTTCTCGCGGCCATAGCAACCGACGTACGGCGT
TGCGCCCTCGCCGGCAGCAAGAAGCCACGGAAGTCCGCCCCGGAGCAGAAA
ATGCCACGCTACTGCGGGTTTATATAGACGGTCCCCACGGGATGGGGAA
AACCACCACCACGCAACTGCTGGTGGCCCTGGGTTTCGCGCGACGATATCGT
CTACGTACCCGAGCCGATGACTTACTGGCGGGTGTGGGGGCTTCCGAGAC
AATCGCGAACATCTACACCACACAACACCGCCTCGACCAGGGTGAGATAT
CGGCCGGGGACGCGGCGGTGGTAATGACAAGCGCCCAGATAACAATGGGC
ATGCCTTATGCCGTGACCGACGCGTTCTGGCTCCTCATATCGGGGGGGAG
GCTGGGAGCTCACATGCCCCGCCCCGGCCCTCACCCTCATCTTCGACCGC
CATCCCATCGCCGCCCTCCTGTGCTACCCGGCCGCGCGGTACCTTATGGGC
AGCATGACCCCCAGGCCGTGCTGGCGTTCGTGGCCCTCATCCCGCCGACC
TTGCCCGGCACCAACATCGTGCTTGGGGCCCTTCCGGAGGACAGACACATC
GACCGCCTGGCCAAACGCCAGCGCCCCGGCGAGCGGCTGGACCTGGCTAT
GCTGGCTGCGATTTCGCCGCGTTTACGGGCTACTTGCCAATACGGTGCGGTA
TCTGCAGTGCGGCGGGTTCGTGGCGGGAGGACTGGGGACAGCTTTCGGGGA
CGGCCGTGCCGCCCCAGGGTGCCGAGCCCCAGAGCAACGCGGGGCCACGA
CCCCATATCGGGGACACGTTATTTACCCTGTTTCGGGCCCCCGAGTTGCTG
GCCCCAACGGCGACCTGTATAACGTGTTTGCCTGGGCCTTGACGTCTTGG
CCCAAACGCCTCCGTTCCATGCACGTCTTATCCTGGATTACGACCAATCG

CCCGCCGGCTGCCGGGACGCCCTGCTGCAACTTACCTCCGGGATGGTCCAG
 ACCCACGTCACCACCCCGGCTCCATAACCGACGATATGCGACCTGGCGCGC
 ACGTTTGCCCGAGAAATGAAGCTTACCATGACCGAGTACAAGCCCACGGT
 GCGCCTCGCCACCCGCGACGACGTCCCCAGGGCCGTACGCACCCTCGCCGC
 CGCGTTCGCCGACTACCCCGCCACGCGCCACACCGTCGATCCGGACCGCCA
 CATCGAGCGGGTCACCGAGCTGCAAGAACTCTTCCTCACGCGCGTCCGGGCT
 CGACATCGGCAAGGTGTGGGTGCGGGACGACGGCGCCGCGGTGGCGGTCT
 GGACCACGCCGAGAGCGTCGAAGCGGGGGCGGTGTTCCGGAGATCGGC
 CCGCGCATGGCCGAGTTGAGCGGTTCCCGGCTGGCCGCGCAGCAACAGAT
 GGAAGGCCTCCTGGCGCCGACCCGGCCCAAGGAGCCCGCGTGGTTCCTGG
 CCACCGTCGGCGTCTCGCCCGACCACCAGGGCAAGGGTCTGGGCAGCGCC
 GTCGTGCTCCCCGGAGTGGAGGCGGCCGAGCGCGCCGGGGTGCCCGCCTT
 CCTGGAGACCTCCGCGCCCCGCAACCTCCCCTTCTACGAGCGGCTCGGCTT
 CACCGTCACCGCCGACGTCGAGGTGCCCGAAGGACCGCGCACCTGGTGCA
 TGACCCGCAAGCCCGGTGCCTGACGCCCGCCCCACGACCCGCAGCGCCCG
 ACCGAAAGGAGCGCACGACCCCATGCATCGATAAGAATTCCGGATCCTTA
 GGCCATTAAGGCCGGCCGCCTCGGCCCACTTCGTGGGGTACCGAGCTCGA
 ATTCACTGGCCGTCGTTTTACAACGTGCTGACTGGGAAAACCCTGGCGTTA
 CCCAACTTAATCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATA
 GCGAAGAGGCCCGCACCGATCGCCCTTCCCAACAGTTGCGTGGCCGAGGA
 GCAGGACTGACACGTGCTACGAGATTCGATTCCACCGCCGCCTTCTATGA
 AAGGTTGGGCTTCGGAATCGTTTTCCGGGACGCCGGCTGGATGATCCTCCA
 GCGCGGGGATCTCATGCTGGAGTTCCTCGCCACCCCAACTTGTTTATTGC
 AGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTACAAATA
 AAGCATTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGT
 ATCTTATCATGTCTGTATACCGTCGACCTCTAGCTAGAGCTTGGCGTAATC
 ATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATCCACAC
 AACATACGAGCCGGAAGCATAAAGTGTAAGCCTGGGGTGCTAATGAGT
 GAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCAGTCGGG
 AAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAG
 GCGGTTTGCATATTGGGCGCTCTCCGCTTCCCTCGCTCACTGACTCGCTGCG
 CTCGGTCGTTCCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAAT

ACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAA
AAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTT
TTCCATAGGCTCCGCCCCCTGACGAGCATCACAAAATCGACGCTCAAGT
CAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCC
TGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATA
CCTGTCCGCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCAATGCTCACGC
TGTAGGTATCTCAGTTCGGTGTAGGTCGTTGCTCCAAGCTGGGCTGTGTG
CACGAACCCCCGTTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGT
CTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACT
GGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTT
GAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATTTGGTATCTG
CGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATC
CGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTTGTTTGCAAGCAGCA
GATTACGCGCAGAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTAC
GGGTCTGACGCTCAGTGGAACGAAAACCTCACGTTAAGGGATTTTGGTCAT
GAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTA AAAAATGAAG
TTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCA
ATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTGTTTCATCC
ATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTA
CCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCT
CCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAG
TGGTCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAA
GCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATT
GCTACAGGCATCGTGGTGTACGCTCGTCGTTTGGTATGGCTTCATTCAGC
TCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAA
AAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCC
GCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCA
TGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCAT
TCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATAC
GGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGA
AAACGTTCTTCGGGGCGAAAACCTCTCAAGGATCTTACCGCTGTTGAGATCC
AGTTCGATGTAACCCACTCGTGCAACCAACTGATCTTCAGCATCTTTTACTT

TCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAA
AAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTT
CAATATTATTGAAGCATTATCAGGGTTATTGTCTCATGAGCGGATACATA
TTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCGCGCACATTTCCC
CGAAAAGTGCCACCTGACGTCGACGGATCGGGAGATCTCCCGATCCCCTAT
GGTCGACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATC
TGCTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTAGTGCGCGAGCAAAATTT
AAGCTACAACAAGGCAAGGCTTGACCGACAATTGCATGAAGAATCTGCTT
AGGGTTAGGCGTTTTTGCCTGCTTCGCGATGTACGGGCCAGATATACGCGT
TGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTA
GTTTCATAGCCCATATATGGAGTTCGCGTTACATAACTTACGGTAAATGGC
CCGCCTGGCTGACCGCCCAACGACCCCGCCATTGACGTCAATAATGACG
TATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTG
GACTATTTACGGTAAACTGCCACTTGGCAGTACATCAAGTGTATCATATG
CCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCAT
TATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTAC
GTATTAGTCATCGCTATTACCATGGTGTATGCGGTTTTTGGCAGTACATCAAT
GGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCAT
GACGTCAATGGGAGTTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAA
TGTCGTAACAACCTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACG
GTGGGAGGTCTATATAAGCAGAGCTCTCTGGCTAACTAGAGAACCCACTG
CTTACTGGCTTATCGAAATTAATACGACTCACTATAGGGAGACCCAAGCTG

SEQ ID NO: 7

p3⁺TKZEO

CATGGCTTCGTACCCCGGCCATCAACACGCGTCTGCGTTCGACCAG
GCTGCGCGTTCTCGCGGCCATAGCAACCGACGTACGGCGTTGCGCCCTCGC
CGGCAGCAAGAAGCCACGGAAGTCCGCCCCGAGCAGAAAATGCCACGCT
ACTGCGGGTTTATATAGACGGTCCCCACGGGATGGGGAAAACCACCACCA
CGCAACTGCTGGTGGCCCTGGGTTTCGCGCGACGATATCGTCTACGTACCCG
AGCCGATGACTTACTGGCGGGTGTGGGGGCTTCCGAGACAATCGCGAAC
ATCTACACCACACAACCCGCCTCGACCAGGGTGAGATATCGGCCGGGGA
CGCGGCGGTGGTAATGACAAGCGCCAGATAACAATGGGCATGCCTTATG

CCGTGACCGACGCCGTTCTGGCTCCTCATATCGGGGGGGAGGCTGGGAGCT
CACATGCCCCGCCCCGGCCCTCACCTCATCTTCGACCGCCATCCCATCG
CCGCCCTCCTGTGCTACCCGGCCGCGCGGTACCTTATGGGCAGCATGACCC
CCCAGGCCGTGCTGGCGTTCGTGGCCCTCATCCCGCCGACCTTGCCCCGCA
CCAACATCGTGCTTGGGGCCCTTCCGGAGGACAGACACATCGACCGCCTG
GCCAAACGCCAGCGCCCCGGCGAGCGGCTGGACCTGGCTATGCTGGCTGC
GATTCGCCGCGTTTACGGGCTACTTGCCAATACGGTGCGGTATCTGCAGTG
CGGCGGGTCGTGGCGGGAGGACTGGGGACAGCTTTCGGGGACGGCCGTGC
CGCCCCAGGGTGCCGAGCCCCAGAGCAACGCGGGCCACGACCCCATATC
GGGGACACGTTATTTACCCTGTTTCGGGCCCCCGAGTTGCTGGCCCCAAC
GGCGACCTGTATAACGTGTTTGCCTGGGCCTTGGACGTCTTGGCCAAACGC
CTCCGTTCCATGCACGTCTTTATCCTGGATTACGACCAATCGCCCGCCGGCT
GCCGGGACGCCCTGCTGCAACTTACCTCCGGGATGGTCCAGACCCACGTCA
CCACCCCGGCTCCATACCGACGATATGCGACCTGGCGCGCACGTTTGCCC
GAGAGATGATCAGCGGAGCTAATGGCGTCATGGCCAAGTTGACCAGTGCC
GTTCCGGTGCTCACCGCGCGGACGTGCGCGGAGCGGTTCGAGTTCTGGACC
GACCGGCTCGGGTTCTCCCGGGACTTCGTGGAGGACGACTTCGCCGGTGTG
GTCCGGGACGACGTGACCCTGTTTCATCAGCGCGGTCCAGGACCAGGTGGT
GCCGGACAACACCCTGGCCTGGGTGTGGGTGCGCGGCCTGGACGAGCTGT
ACGCCGAGTGGTTCGGAGGTCGTGTCCACGAACTTCCGGGACGCCTCCGGG
CCGGCCATGACCGAGATCGGCGAGCAGCCGTGGGGGCGGGAGTTCGCCCT
GCGCGACCCGGCCGCAACTGCGTGCACCTTCGTGGCCGAGGAGCAGGACT
GACCGACGCCGACCAACACCGCCGTCCGACGGCGGCCACGGGTCCCAG
GGTCGACCTCGAGATCCTTAGGCCATTAAGGCCGGCCGCCTCGGCCCACTT
CGTGGGGTACCGAGCTCGAATTCCTGGCCGTGTTTTACAACGTCGTGAC
TGGGAAAACCCTGGCGTTACCCAACCTAATCGCCTTGCAGCACATCCCCCT
TTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCAA
CAGTTGCGTGCCGAGGAGCAGGACTGACACGTGCTACGAGATTTGATT
CCACCGCCGCCTTCTATGAAAGGTTGGGCTTCGGAATCGTTTTCCGGGACG
CCGGCTGGATGATCCTCCAGCGCGGGGATCTCATGCTGGAGTTCTTCGCC
ACCCCAACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCA
TCACAAATTCACAAATAAAGCATTTTTTCACTGCATTCTAGTTGTGGTTT

GTCCAAACTCATCAATGTATCTTATCATGTCTGTATACCGTCGACCTCTAGC
 TAGAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATC
 CGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAGCC
 TGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTG
 CCCGCTTTCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGC
 CAACGCGCGGGGAGAGGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCG
 CTCACTGACTCGCTGCGCTCGGTTCGTTTCGGCTGCGGCGAGCGGTATCAGCT
 CACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGG
 AAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAG
 GCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCTGACGAGCATCAC
 AAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAG
 ATACCAGGCGTTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACC
 CTGCCGCTTACCGGATACCTGTCCGCCTTCTCCCTTCGGGAAGCGTGGCG
 CTTTCTCAATGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTGCTTCGCT
 CCAAGCTGGGCTGTGTGCACGAACCCCCGTTACGCCGACCGCTGCGCCT
 TATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGC
 CACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGC
 GGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAG
 GACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAG
 AGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTT
 TTTTGTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAG
 ATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACCTCAC
 GTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCC
 TTTTAAATTA AAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAA
 CTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGA
 TCTGTCTATTTGTTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAAC
 TACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCG
 AGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCG
 GAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGT
 CTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTT
 TGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTACGCTCGTCGT
 TTGGTATGGCTTCATTCAGCTCCGGTTC CAACGATCAAGGCGAGTTACAT

GATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCG
 TTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCAC
 TGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGG
 TGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGGCACCAGTTG.
 CTCTTGCCCGGGGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTT
 AAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACCTCTCAAGGA
 TCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACT
 GATCTTCAGCATCTTTTACTTTACCAGCGTTTCTGGGTGAGCAAAAACAG
 GAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTG
 AATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTATCAGGGTTAT
 TGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATA
 GGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTCGACGGATCG
 GGAGATCTCCCGATCCCCTATGGTCGACTCTCAGTACAATCTGCTCTGATG
 CCGCATAGTTAAGCCAGTATCTGCTCCCTGCTTGTGTGTTGGAGGTCGCTG
 AGTAGTGCGGAGCAAAATTTAAGCTACAACAAGGCAAGGCTTGACCGAC
 AATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGCTTCGCGAT
 GTACGGGCCAGATATACGCGTTGACATTGATTATTGACTAGTTATTAATAG
 TAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTT
 ACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGC
 CCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACT
 TTCCATTGACGTCAATGGGTGGACTATTTACGGTAAACTGCCCACTTGGCA
 GTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGAC
 GGTAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTT
 CCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATG
 CGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTTGACTCACGGGGA
 TTTCCAAGTCTCCACCCATTGACGTCAATGGGAGTTTTGTTTTGGCACCAA
 AATCAACGGGACTTTCCAAAATGTCGTAACAACCTCCGCCCCATTGACGCAA
 ATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCTCTG
 GCTAACTAGAGAACCCACTGCTTACTGGCTTATCGAAATTAATACGACTCA
 CTATAGGGAGACCCAAGCTGGCTAGTGGATCCCCCGGGCTGCAGGAATTC
 GATATCAAGCTTATCG

REIVINDICACIONES

1. Un ácido nucleico que comprende los elementos contiguos siguientes dispuestos en la dirección 5' a 3':
 - a) un promotor;
 - b) al menos dos marcadores seleccionables;
- 5 c) un sitio de clonación para recepción de un segmento de ácido nucleico, comprendiendo dicho segmento una secuencia diana candidata de RNA regulador; y
 - d) una señal de poliadenilación,

estando dichos elementos dispuestos de tal manera que un transcrito dirigido por dicho promotor comprende dichos al menos dos marcadores seleccionables, dicha secuencia diana candidata de RNA regulador, y dicha señal de poliadenilación por este orden.
- 10 2. Un ácido nucleico de acuerdo con la reivindicación 1, en donde dicha secuencia diana candidata de RNA regulador es una secuencia diana de microRNA (miRNA) candidata.
3. Un ácido nucleico de acuerdo con la reivindicación 1 o la reivindicación 2, que comprende adicionalmente un codón de parada localizado entre dichos marcadores seleccionables y dicho sitio de clonación.
- 15 4. Un ácido nucleico de acuerdo con cualquiera de las reivindicaciones anteriores en el cual al menos dos de dichos marcadores seleccionables se proporcionan como un marco de lectura abierto que codifica un solo polipéptido que comprende al menos dos marcadores seleccionables.
5. Un ácido nucleico de acuerdo con la reivindicación 4, en el cual dichos marcadores seleccionables comprenden un marcador para selección positiva y un marcador para selección negativa.
- 20 6. Un ácido nucleico de acuerdo con cualquiera de las reivindicaciones anteriores, en el cual dicho sitio de clonación tiene insertado en él un segmento de ácido nucleico que comprende una región no traducida (UTR) 3' o una UTR 3' candidata.
7. Una biblioteca UTR 3', comprendiendo dicha biblioteca una pluralidad de ácidos nucleicos de acuerdo con la reivindicación 6.
- 25 8. Una biblioteca UTR 3' de acuerdo con la reivindicación 7, en la cual dichas secuencias diana de miRNA candidatas están constituidas por cDNA's.
9. Una célula que comprende un ácido nucleico de acuerdo con cualquiera de las reivindicaciones 1 a 6 o una biblioteca de acuerdo con cualquiera de las reivindicaciones 7 a 8.
- 30 10. Un método de construcción de una biblioteca UTR 3' que comprende proporcionar un ácido nucleico de acuerdo con cualquiera de las reivindicaciones 1 a 6, e insertar en dicho sitio de clonación un ácido nucleico que comprende una UTR 3' o una UTR 3' candidata.
11. Un método de construcción de una biblioteca UTR 5' que comprende proporcionar un ácido nucleico de acuerdo con cualquiera de las reivindicaciones 1 a 6, e insertar en dicho sitio de clonación un ácido nucleico que comprende una UTR 5' o una UTR 5' candidata.
- 35 12. Un vector que comprende un ácido nucleico de acuerdo con cualquiera de las reivindicaciones 1 a 6.
13. Un método para identificación de una secuencia diana de miRNA que comprende los pasos de
 - (a) introducir un ácido nucleico de acuerdo con cualquiera de las reivindicaciones 1 a 6, que comprende una secuencia diana de miRNA candidata en una célula hospedadora;
 - (b) seleccionar una o más células hospedadoras que expresan el primer marcador seleccionable de dicho ácido nucleico;
 - (c) introducir al menos un miRNA de interés en dicha o dichas células hospedadoras de (b), y
 - (d) ensayar la expresión del segundo marcador seleccionable de dicho ácido nucleico en las células de (c),

en donde, si las células de (c) no exhiben la expresión del segundo marcador seleccionable, entonces la secuencia diana de miRNA candidata se identifica como una secuencia diana de miRNA.
- 45 14. Un método para identificación de identificación de un miRNA activo contra una secuencia diana de miRNA que comprende los pasos de

(a) introducir un ácido nucleico de acuerdo con cualquiera de las reivindicaciones 1 a 6, que comprende dicha secuencia diana de miRNA en una célula hospedadora;

(b) seleccionar una o más células hospedadoras que expresan el primer marcador seleccionable de dicho ácido nucleico;

5 (c) introducir al menos un miRNA de interés en dicha o dichas células hospedadoras de (b), y

(d) ensayar la expresión del segundo marcador seleccionable de dicho ácido nucleico en las células de (c),

en donde, si las células de (c) no exhiben la expresión del segundo marcador seleccionable, entonces el miRNA de interés se identifica como un miRNA activo contra dicha secuencia diana de miRNA.

15. Un método para identificación de un inhibidor de un RNA regulador que comprende los pasos de

10 (a) introducir al menos un RNA regulador de interés en una célula hospedadora;

(b) introducir un ácido nucleico de acuerdo con cualquiera de las reivindicaciones 1 a 6, que comprende una secuencia diana candidata de RNA regulador en dicha célula hospedadora;

(c) seleccionar una o más células hospedadoras que no exhiben expresión del primer marcador seleccionable de dicho ácido nucleico;

15 (d) introducir en dichas células hospedadoras una sustancia o ácido nucleico de test;

(e) ensayar la expresión del segundo de dichos marcadores seleccionables en las células de (d);

en donde, si las células de (d) exhiben la expresión del segundo marcador seleccionable, entonces la sustancia o ácido nucleico de test se identifica como inhibidora de dicho RNA regulador.

16. Un método para identificación de una secuencia diana de RNA regulador que comprende los pasos de

20 (a) introducir un ácido nucleico de acuerdo con cualquiera de las reivindicaciones 1 a 6 que comprende una secuencia diana candidata de RNA regulador en una célula hospedadora;

(b) seleccionar una o más células hospedadoras que expresan el primer marcador seleccionable de dicho ácido nucleico;

(c) introducir al menos un RNA regulador de interés en dicha o dichas células hospedadoras de (b), y

25 (d) ensayar la expresión del segundo marcador seleccionable de dicho ácido nucleico en las células de (c),

en donde, si las células de (c) no exhiben la expresión del segundo marcador seleccionable, la secuencia diana candidata de RNA regulador se identifica como una secuencia diana de RNA regulador.

17. Un método de acuerdo con la reivindicación 16, en el cual dicho RNA regulador es un siRNA y en el cual dicha secuencia diana candidata de RNA regulador es una secuencia diana de siRNA candidata.

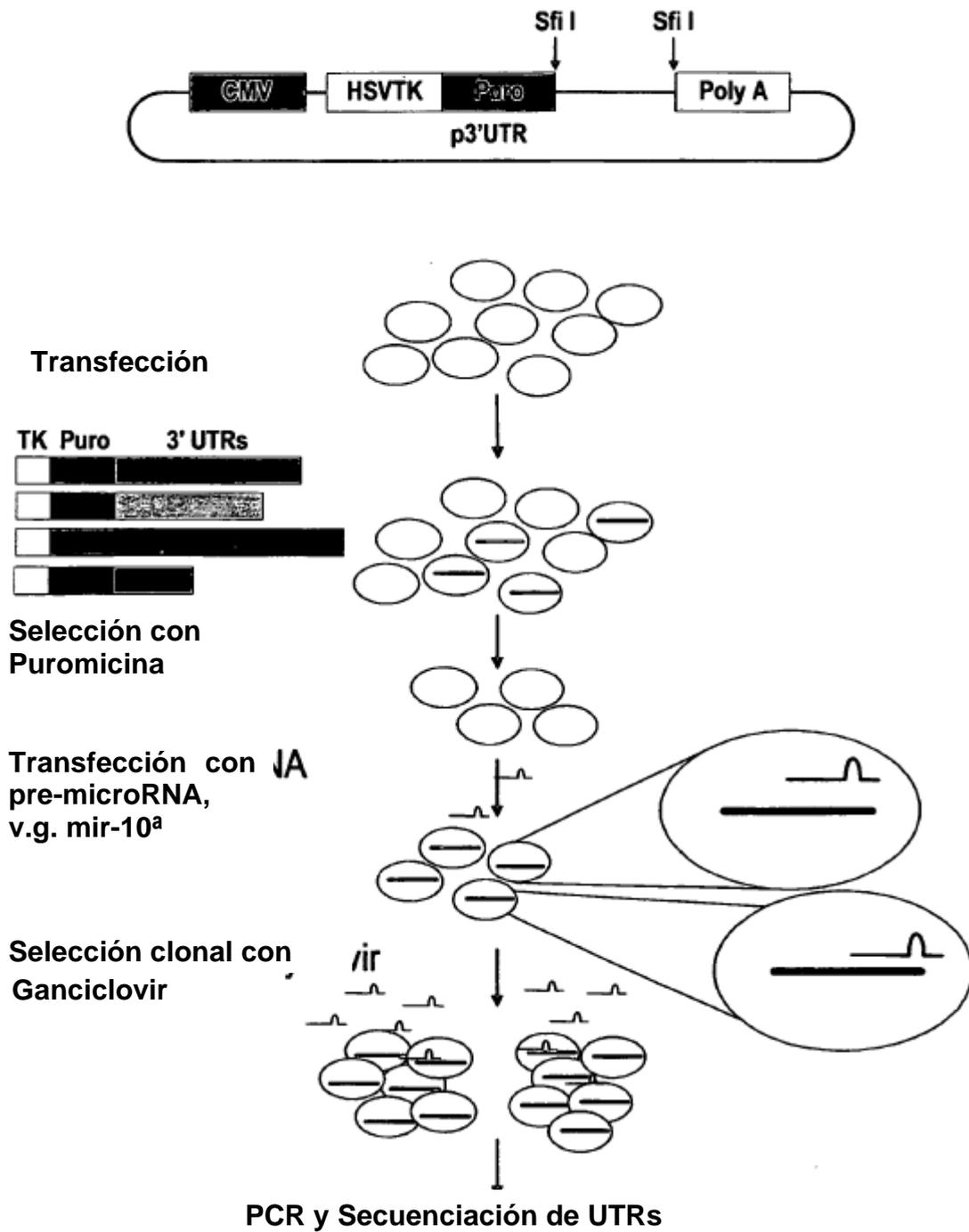


FIG. 1

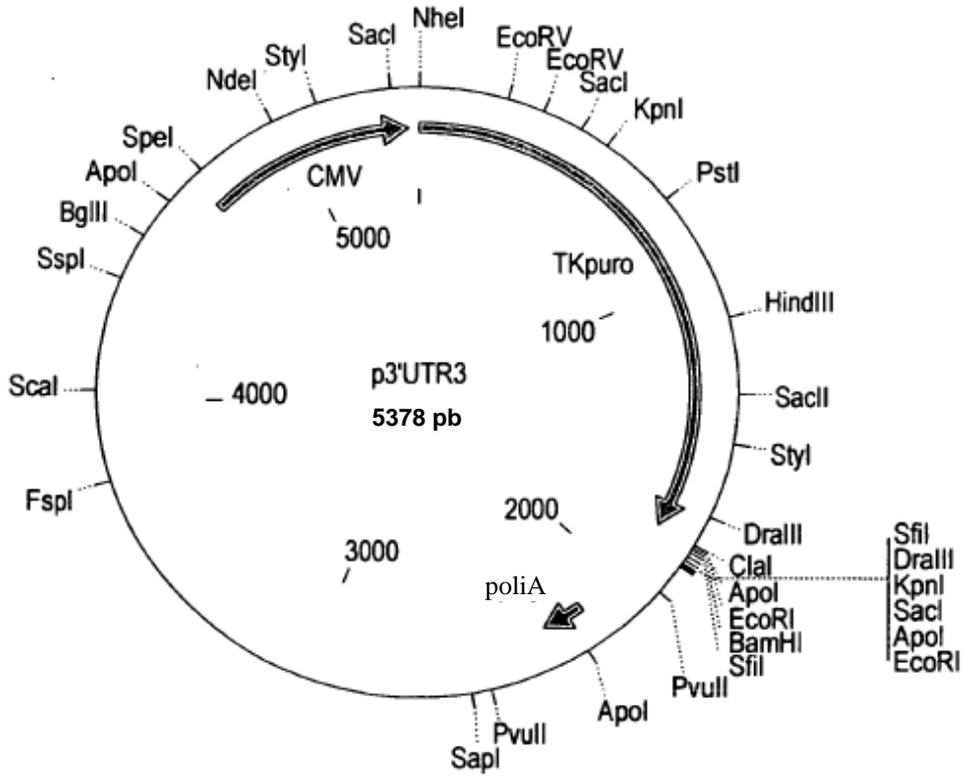


FIG. 2

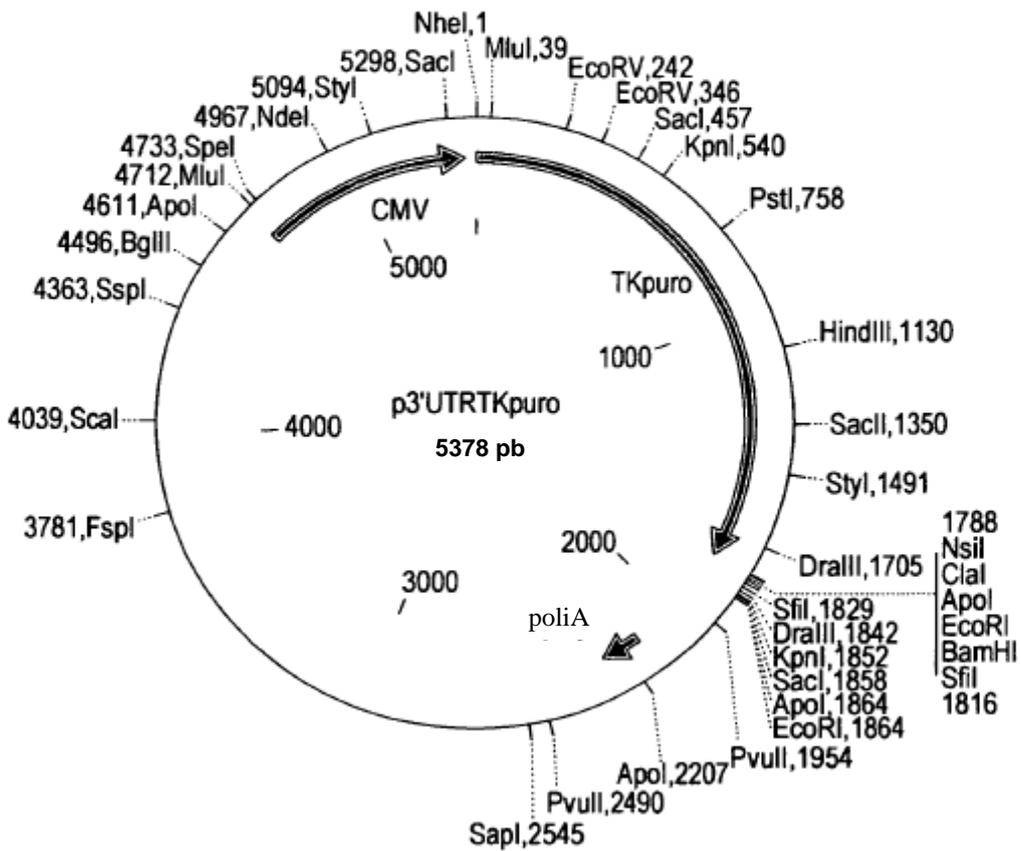


FIG. 3

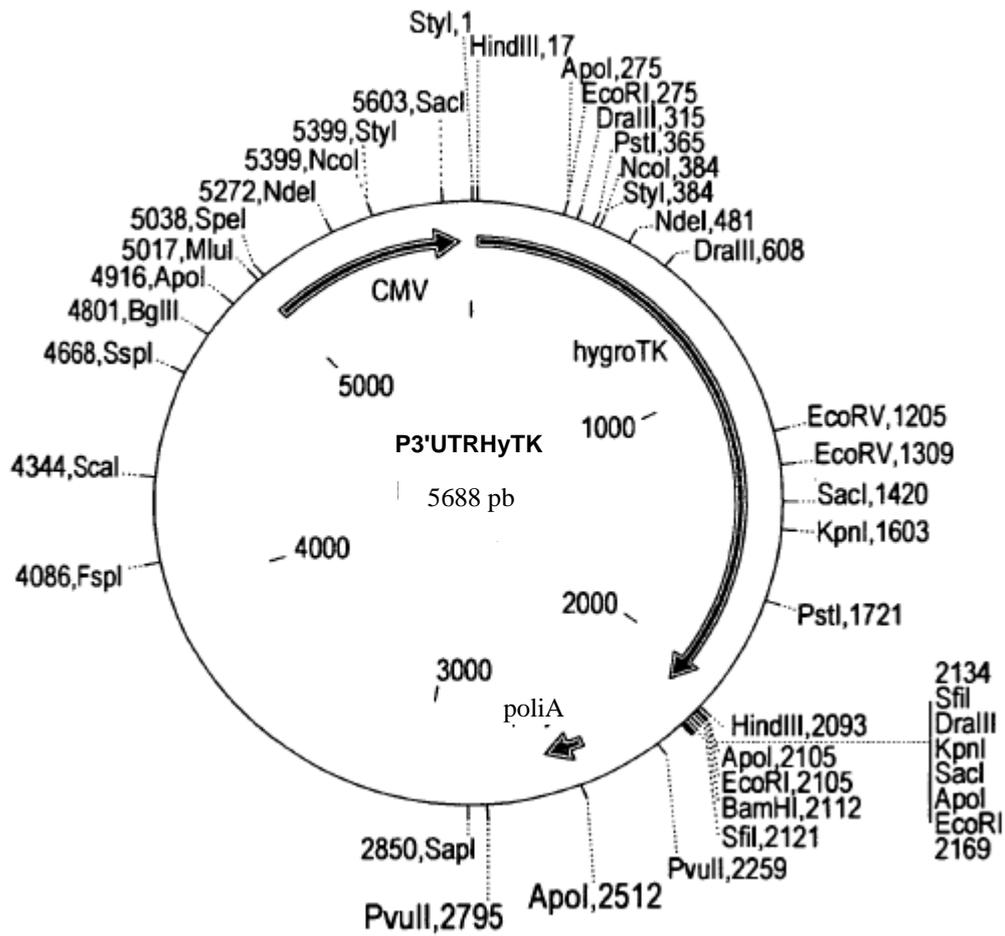


FIG. 4

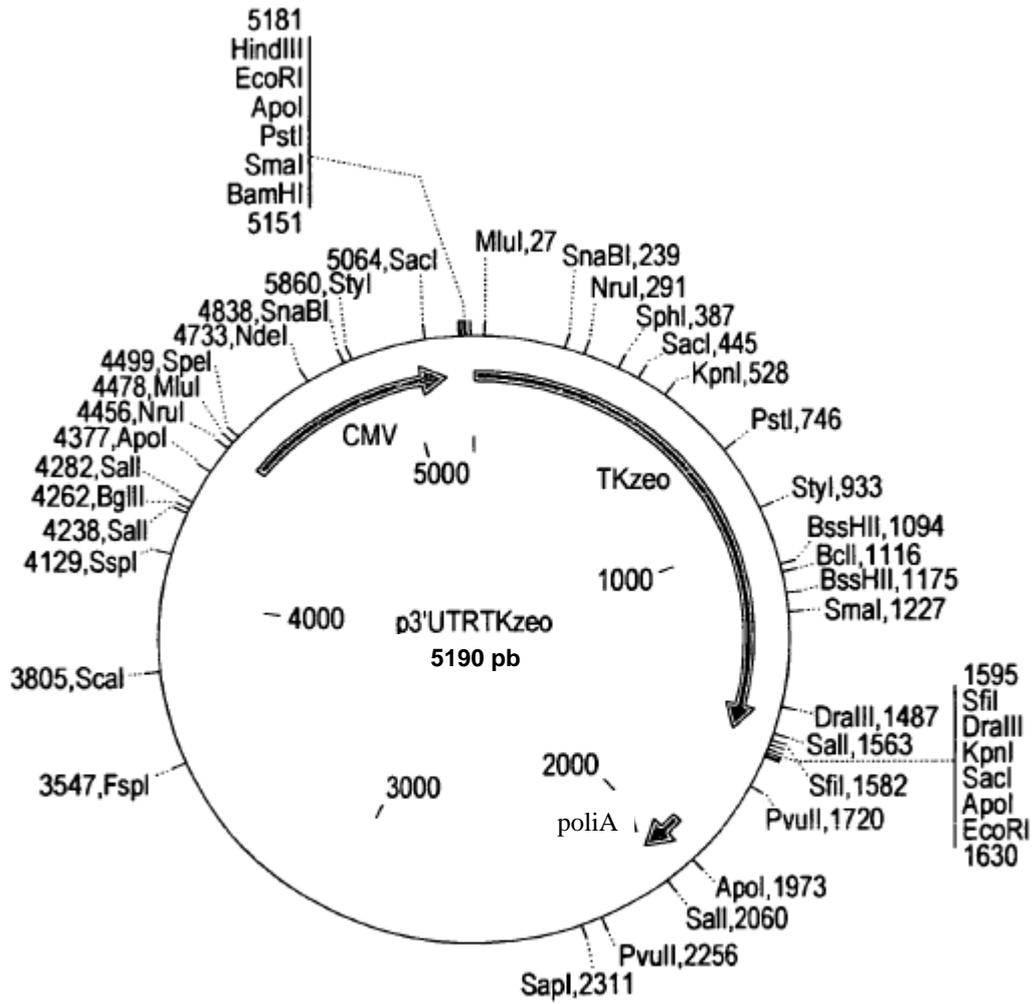


FIG. 5

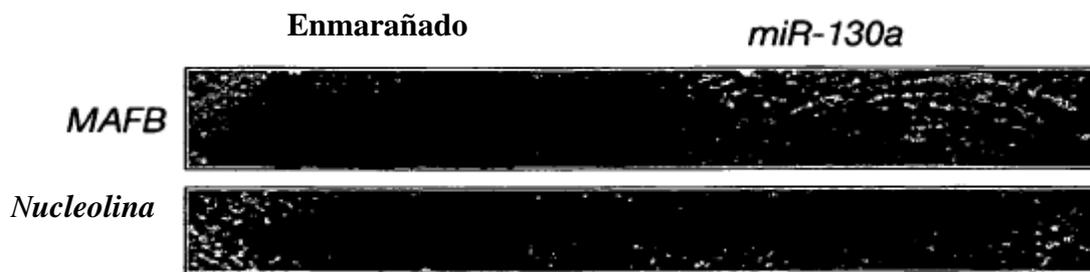
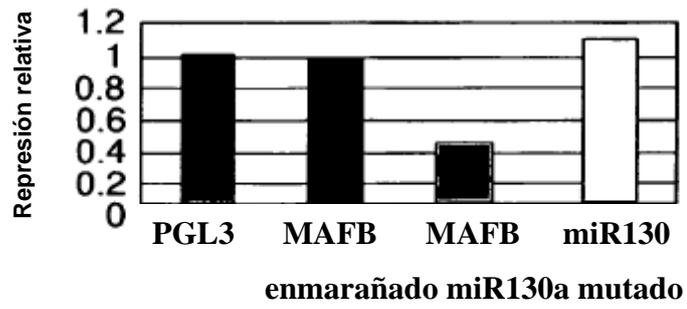


FIG. 6

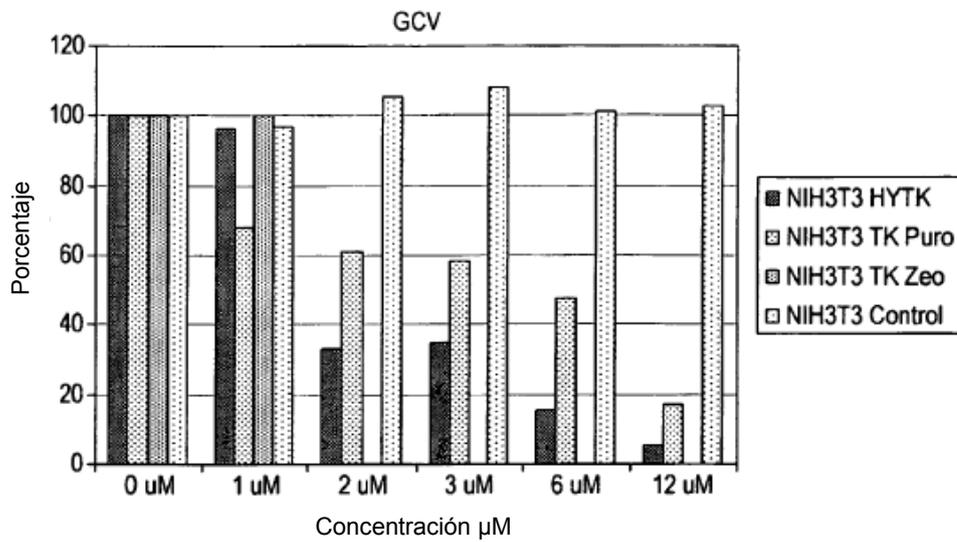


FIG. 7

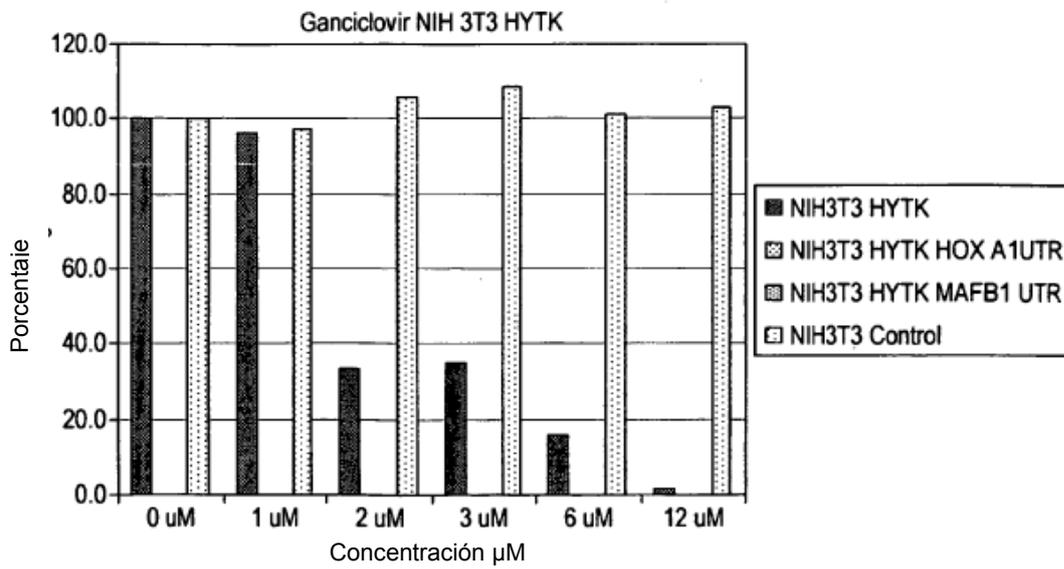


FIG. 8

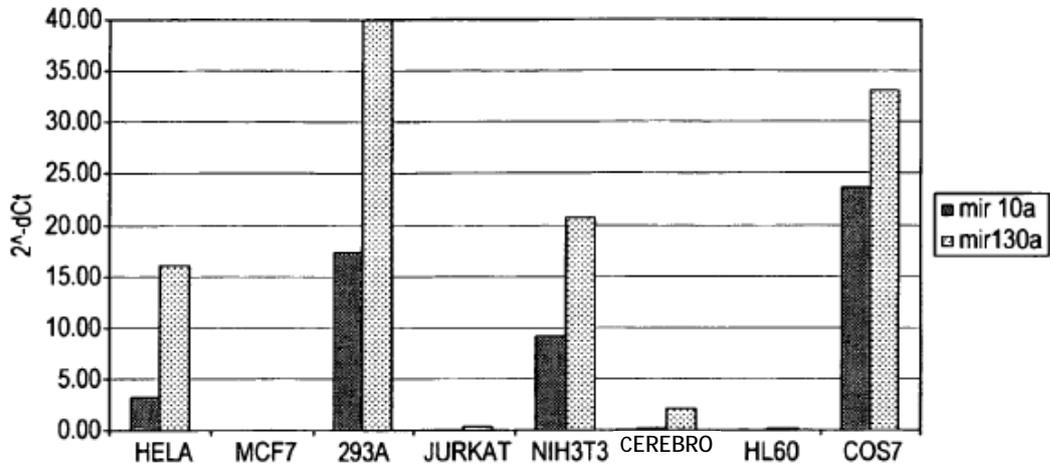


FIG. 9

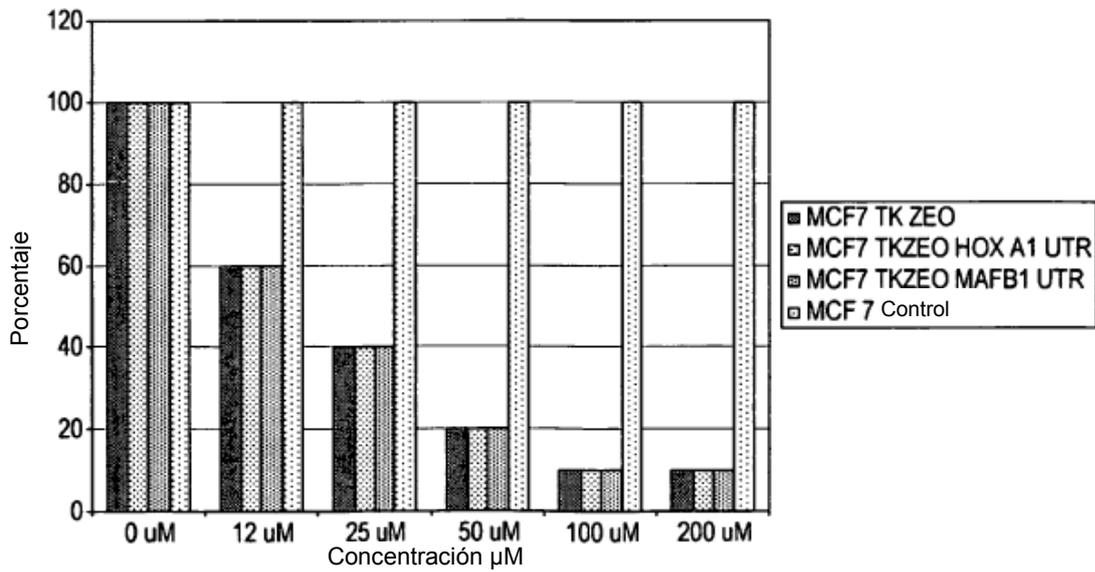


FIG. 10

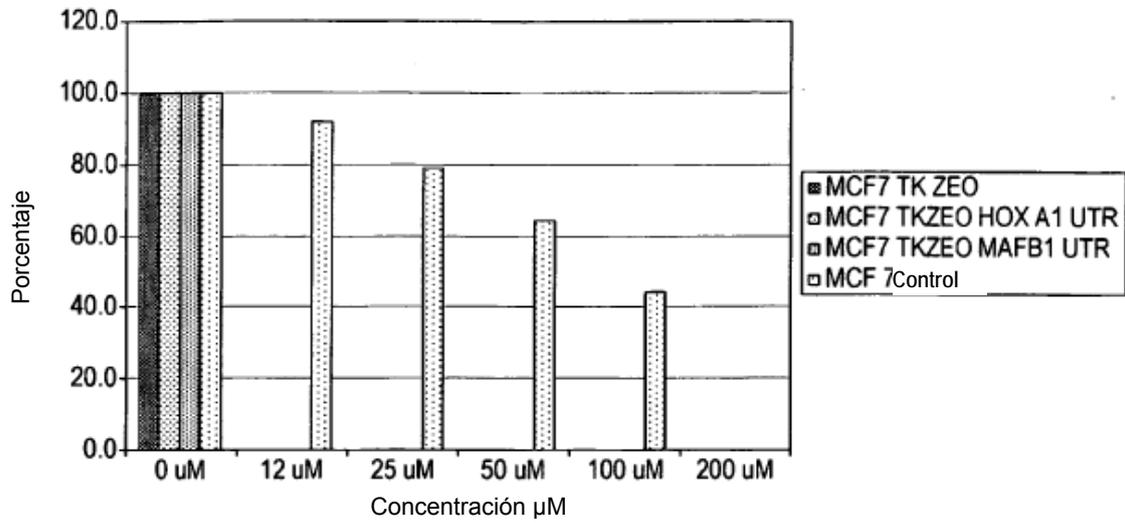


FIG. 11

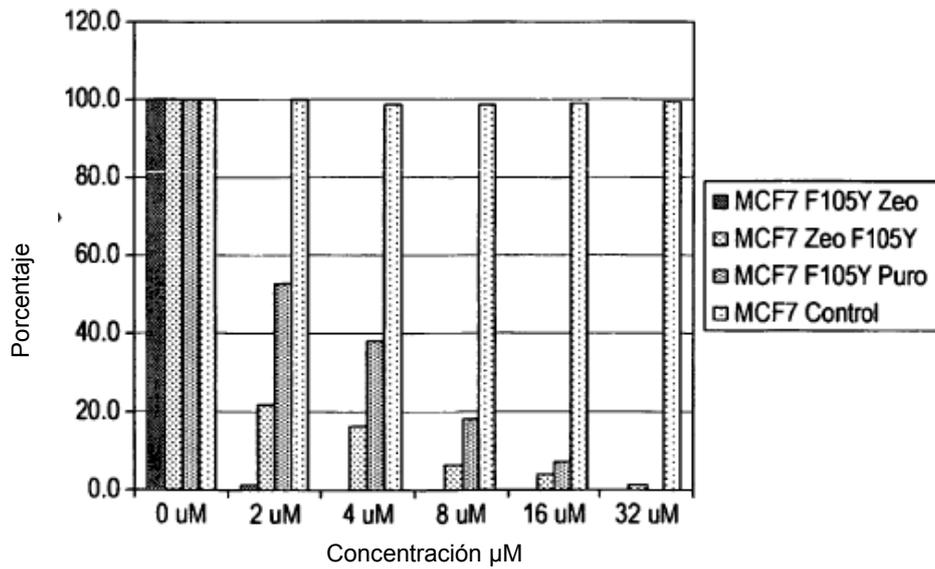


FIG. 12

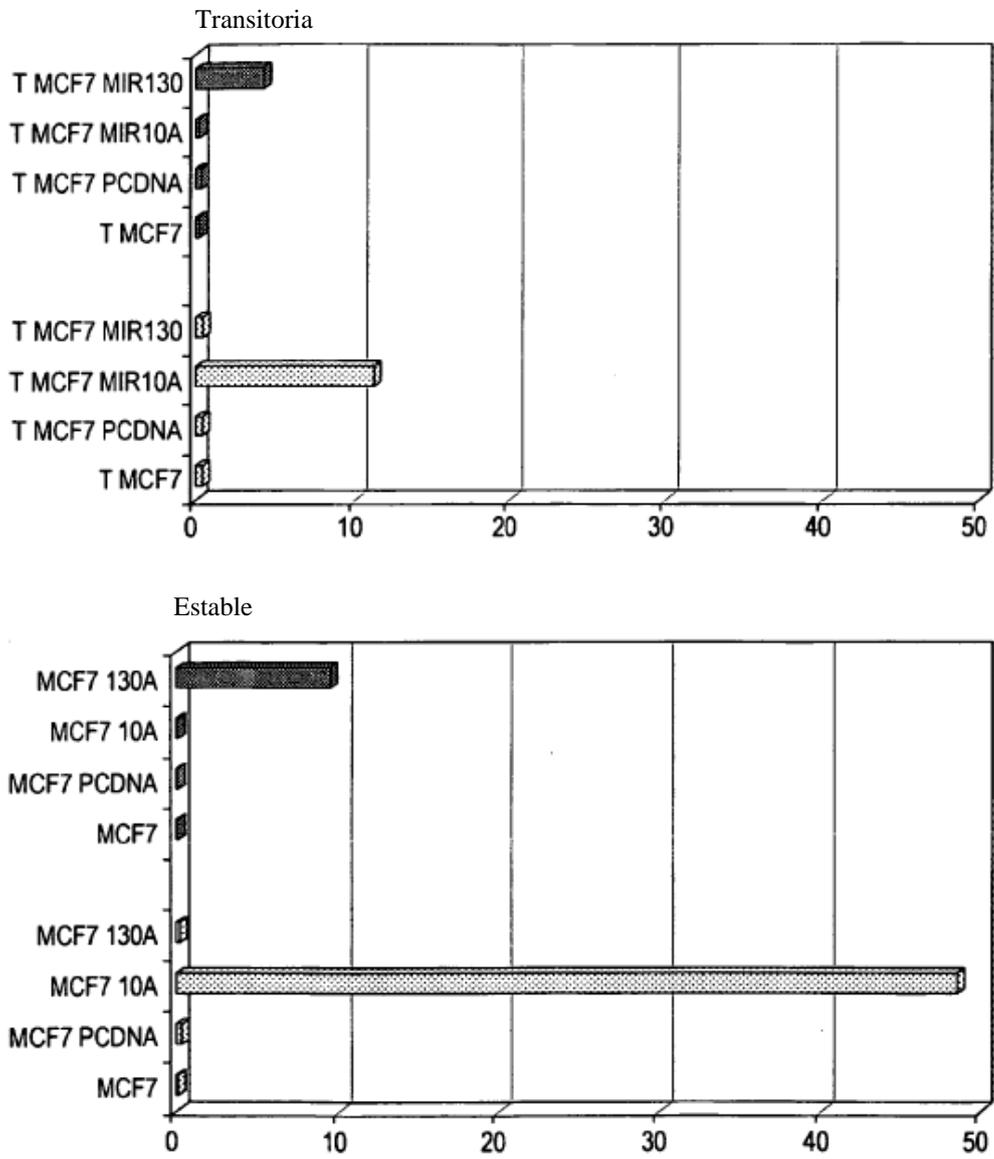


FIG. 13

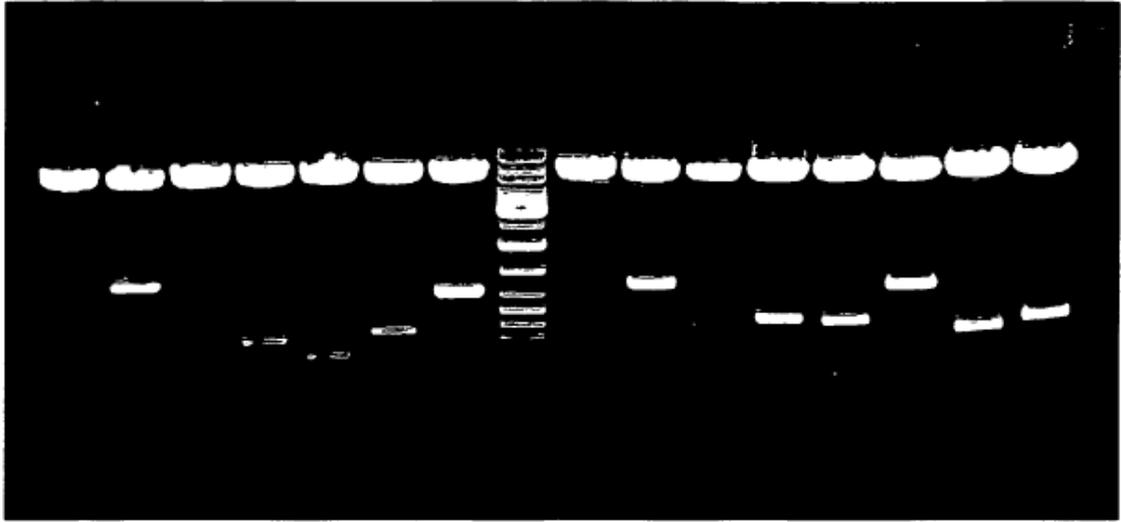


FIG. 14

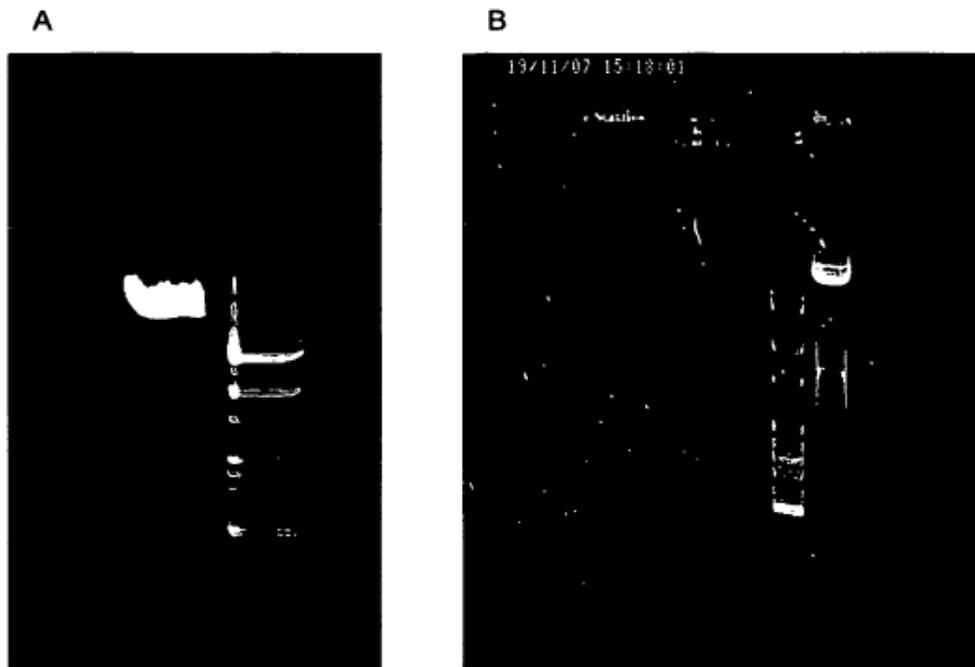


FIG. 15

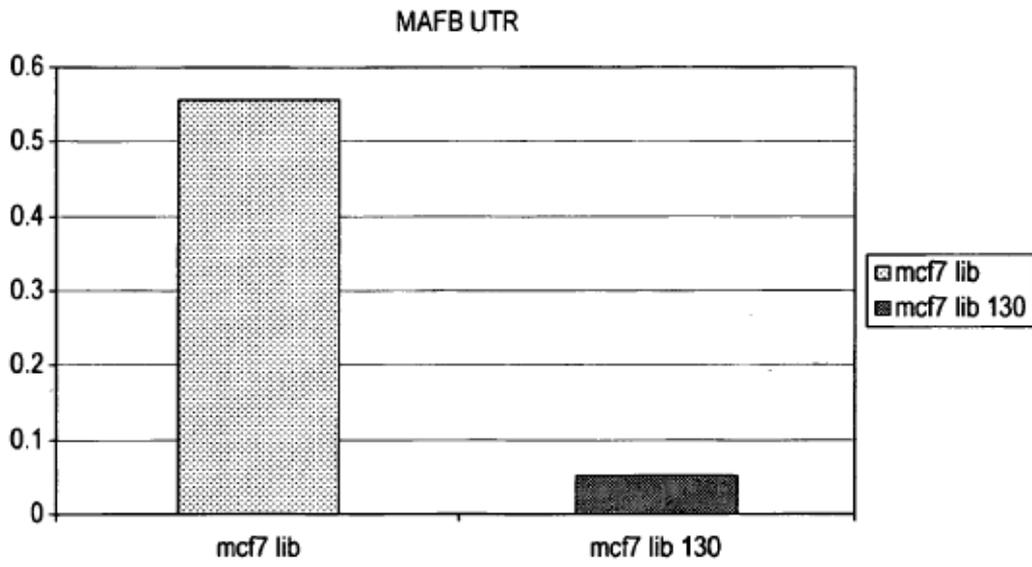


FIG. 16