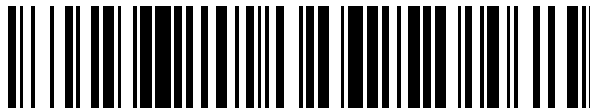


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 663 424**

51 Int. Cl.:

C12N 15/10 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **28.06.2013 PCT/US2013/048619**

87 Fecha y número de publicación internacional: **03.01.2014 WO14005042**

96 Fecha de presentación y número de la solicitud europea: **28.06.2013 E 13734638 (3)**

97 Fecha y número de publicación de la concesión europea: **01.11.2017 EP 2867361**

54 Título: **Genética combinatoria masivamente en paralelo**

30 Prioridad:

29.06.2012 US 201261666081 P

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

12.04.2018

73 Titular/es:

**MASSACHUSETTS INSTITUTE OF TECHNOLOGY
(100.0%)
77 Massachusetts Avenue
Cambridge, MA 02139, US**

72 Inventor/es:

**LU, TIMOTHY KUAN-TA y
CHENG, ALLEN**

74 Agente/Representante:

FÚSTER OLAGUIBEL, Gustavo Nicolás

ES 2 663 424 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

Genética combinatoria masivamente en paralelo

5 Interés del gobierno

Esta invención se realizó con apoyo del gobierno con la subvención n.º DP2 OD008435 otorgada por los National Institutes of Health. El gobierno tiene determinados derechos en esta invención.

10 Campo de la invención

La invención se refiere a métodos y composiciones para la generación rápida de combinaciones de orden alto de elementos genéticos y la identificación fácil de elementos genéticos.

15 Antecedentes de la invención

La biología se ha visto limitada continuamente por la incapacidad para generar perturbaciones celulares complejas. Estudios previos han cribado fenotipos deseados usando sólo estudios de sobreexpresión de un solo gen (Kitagawa *et al.* (2005) *DNA Research* 12:291-9; Soo *et al.* (2011) *Proc Natl Acad Sci USA* 108:1484-9) y estudios de desactivación de un solo gen, así como un conjunto limitado de estudios de desactivación doble (Butland *et al.* (2008) *Nature Methods* 5:789-95; Pan *et al.* (2004) *Molecular Cell* 16:487-96; Tong *et al.* (2004) *Science* 303:808-13). Sin embargo, estos métodos no pueden aumentarse a escala para generar combinaciones de perturbaciones de orden superior, impidiendo una consulta más poderosa y compleja de redes. Además, estos métodos no pueden aumentarse a escala fácilmente para generar más de unos pocos miles de combinaciones. Esto impide el cribado y estudio de millones de combinaciones en un modo agrupado de alto rendimiento, en lugar de depender de cribado basado en colonias o pocillos. Finalmente, estos métodos no están diseñados para permitir una caracterización rápida de combinaciones específicas.

Anderson J. C. *et al.*, *Journal of Biological Engineering*, vol. 4, n.º 1, 20 de enero de 2010, página 1, se refieren a BgLBriks: un criterio flexible para el ensamblaje de partes biológicas.

Sarrion-Perdigones A. *et al.*, *PLOS ONE*, vol. 6, n.º 7, 7 de julio de 2011, página e21622, se refieren a GoldenBraid: un sistema de clonación iterativo para el ensamblaje estandarizado de módulos genéticos reutilizables.

Merryman C. *et al.*, *Metabolic Engineering*, vol. 14, n.º 3, páginas 196-204, se refieren a métodos y aplicaciones para ensamblar constructos de ADN grandes.

Chang K. *et al.*, *Nature Methods*, vol. 3, n.º 9, 1 de septiembre de 2006, páginas 707-714 se refieren a lecciones de la naturaleza: colecciones de ARNhp a base de microARN.

Wingler L. *et al.*, *Proceedings of the National Academy of Sciences of the United States of America*, vol. 108, n.º 37, 13 de septiembre de 2011, páginas 15135-15140, se refieren a recombinación reiterativa para el ensamblaje *in vivo* de colecciones de rutas de múltiples genes.

Dymond J. S. *et al.*, *Nature*, vol. 477, n.º 7365, 2011, páginas 471-476, se refieren a brazos de cromosomas sintéticos que funcionan en levadura y generan diversidad fenotípica mediante diseño.

Sumario de la invención

La presente invención se define, entre otros, mediante los siguientes puntos:

1. Un constructo genético que comprende:

un elemento de ADN;

un primer elemento de extremos compatibles y un segundo elemento de extremos compatibles que flanquean al elemento de ADN, en el que los elementos de extremos compatibles primero y segundo son capaces de aparearse entre sí;

un elemento de código de barras;

un tercer elemento de extremos compatibles y un cuarto elemento de extremos compatibles que flanquean al elemento de código de barras, en el que los elementos de extremos compatibles tercero y cuarto son capaces de aparearse entre sí pero no son capaces de aparearse a los elementos de extremos compatibles primero o segundo;

y

un sitio de separación ubicado entre el cuarto elemento de extremos compatibles y el primer elemento de extremos compatibles, en el que el elemento de ADN, primer elemento de extremos compatibles y segundo elemento de extremos compatibles están en un lado del sitio de separación, y el elemento de código de barras, tercer elemento de extremos compatibles y cuarto elemento de extremos compatibles están en el otro lado del sitio de separación.

- 5 2. Un constructo genético que comprende:
- una pluralidad de elementos de ADN;
- 10 un primer elemento de extremos compatibles y un segundo elemento de extremos compatibles que flanquean a la pluralidad de elementos de ADN, en el que los elementos de extremos compatibles primero y segundo son capaces de aparearse entre sí;
- 15 una pluralidad de elementos de código de barras;
- un tercer elemento de extremos compatibles y un cuarto elemento de extremos compatibles que flanquean a la pluralidad de elementos de código de barras, en el que los elementos de extremos compatibles tercero y cuarto son capaces de aparearse entre sí pero no son capaces de aparearse a los elementos de extremos compatibles primero o segundo; y
- 20 un sitio de separación ubicado entre la pluralidad de elementos de ADN y la pluralidad de elementos de código de barras.

- 25 3. Un método para generar un constructo genético combinatorio, que comprende:
- proporcionar un vector que contiene un primer constructo genético según el punto 1 o punto 2;
- escindir el vector en el sitio de separación dentro del primer constructo genético, dando como resultado que el primer constructo genético se separe en segmentos primero y segundo;
- 30 proporcionar un segundo constructo genético según el punto 1 o punto 2; y aparear el segundo constructo genético al vector escindido, en el que el apareamiento se produce en elementos de extremos compatibles dentro de los constructos genéticos primero y segundo que son capaces de aparearse entre sí, y en el que tras el apareamiento, el segundo constructo genético se integra entre los segmentos primero y segundo del primer constructo genético, creando un constructo genético combinatorio.
- 35

4. El método del punto 3, en el que el método es iterativo.

- 40 5. Un constructo genético combinatorio producido mediante el método del punto 3 o punto 4.

6. Un método para la identificación de un elemento de ADN o una pluralidad de elementos de ADN, que comprende:
- proporcionar un constructo genético según el punto 1 o punto 2;
- 45 realizar un ensayo para determinar la secuencia de ADN del código de barras o la pluralidad de códigos de barras dentro del constructo genético y/o la secuencia de ADN del elemento de ADN o la pluralidad de elementos de ADN dentro del constructo genético; e
- 50 identificar el elemento de ADN o la pluralidad de elementos de ADN.

7. Una colección que comprende:
- dos o más constructos genéticos según el punto 1 o punto 2.
- 55 Se describen en el presente documento métodos y composiciones que permiten la generación rápida de combinaciones de orden alto de elementos genéticos y que proporcionan una base con código de barras para la caracterización rápida de la combinación específica de elementos genéticos codificados dentro de una única célula o en una población agrupada. Estos enfoques nuevos conllevan amplias implicaciones para múltiples campos, permitiendo un gran conjunto de investigaciones en donde la generación de combinaciones de elementos genéticos es informativa y productiva. Esta tecnología puede producir nuevas visiones de fenotipos complejos y propiedades de redes emergentes que previamente han sido inaccesibles al estudio.
- 60

Aspectos de la invención se refieren a un constructo genético que comprende: un elemento de ADN; un primer elemento de extremos compatibles y un segundo elemento de extremos compatibles que flanquean al elemento de ADN, en el que los elementos de extremos compatibles primero y segundo son capaces de aparearse entre sí; un elemento de código de barras; un tercer elemento de extremos compatibles y un cuarto elemento de extremos

65

compatibles que flanquean al elemento de código de barras, en el que los elementos de extremos compatibles tercero y cuarto son capaces de aparearse entre sí pero no son capaces de aparearse a los elementos de extremos compatibles primero o segundo; y un sitio de separación ubicado entre el cuarto elemento de extremos compatibles y el primer elemento de extremos compatibles, en el que el elemento de ADN, primer elemento de extremos compatibles y segundo elemento de extremos compatibles están en un lado del sitio de separación, y el elemento de código de barras, tercer elemento de extremos compatibles y cuarto elemento de extremos compatibles están en el otro lado del sitio de separación.

Aspectos adicionales de la invención se refieren a un constructo genético que comprende: una pluralidad de elementos de ADN; un primer elemento de extremos compatibles y un segundo elemento de extremos compatibles que flanquean a la pluralidad de elementos de ADN, en el que los elementos de extremos compatibles primero y segundo son capaces de aparearse entre sí; una pluralidad de elementos de código de barras; un tercer elemento de extremos compatibles y un cuarto elemento de extremos compatibles que flanquean a la pluralidad de elementos de código de barras, en el que los elementos de extremos compatibles tercero y cuarto son capaces de aparearse entre sí pero no son capaces de aparearse a los elementos de extremos compatibles primero o segundo; y un sitio de separación ubicado entre la pluralidad de elementos de ADN y la pluralidad de elementos de código de barras.

Aspectos adicionales de la invención se refieren a un método para generar un constructo genético combinatorio, que comprende: proporcionar un vector que contiene un primer constructo genético asociado con la invención; escindir el vector en el sitio de separación dentro del primer constructo genético, dando como resultado que el primer constructo genético se separe en segmentos primero y segundo; proporcionar un segundo constructo genético asociado con la invención; y aparear el segundo constructo genético al vector escindido, en el que el apareamiento se produce en elementos de extremos compatibles dentro de los constructos genéticos primero y segundo que son capaces de aparearse entre sí, y en el que tras el apareamiento, el segundo constructo genético se integra entre los segmentos primero y segundo del primer constructo genético, creando un constructo genético combinatorio. En algunas realizaciones, el método es iterativo.

Aspectos adicionales de la invención se refieren a métodos para la identificación de un elemento de ADN o una pluralidad de elementos de ADN, que comprenden: proporcionar un constructo genético asociado con la invención; realizar un ensayo para determinar la secuencia de ADN del código de barras o la pluralidad de códigos de barras dentro del constructo genético y/o la secuencia de ADN del elemento de ADN o la pluralidad de elementos de ADN dentro del constructo genético; e identificar el elemento de ADN o la pluralidad de elementos de ADN.

Aspectos adicionales de la invención se refieren a una colección que comprende: dos o más constructos genéticos asociados con la invención.

Aspectos adicionales de la descripción se refieren a métodos para generar un constructo genético combinatorio, que comprende: proporcionar un vector que comprende: un primer elemento de ADN, un primer elemento de código de barras y dos elementos de recombinación específica de sitio ubicados entre el primer elemento de ADN y el primer elemento de código de barras; proporcionar un primer inserto que comprende: un segundo elemento de ADN, un segundo elemento de código de barras y elementos de recombinación específica de sitio que flanquean a cada uno del segundo elemento de ADN y el segundo elemento de código de barras, de manera que dos elementos de recombinación específica de sitio están ubicados entre el segundo elemento de ADN y el segundo elemento de código de barras que no son compatibles con los elementos de recombinación específica de sitio dentro del vector, y dos elementos de recombinación específica de sitio están ubicados fuera del segundo elemento de ADN y el segundo elemento de código de barras que son compatibles con los elementos de recombinación específica de sitio dentro del vector; realizar recombinación específica de sitio entre el vector y el primer inserto, en el que la recombinación específica de sitio se produce entre los elementos de recombinación específica de sitio dentro del vector ubicados entre el primer elemento de ADN y el primer elemento de código de barras y los elementos de recombinación específica de sitio compatibles dentro del primer inserto ubicados fuera del segundo elemento de ADN y el segundo elemento de código de barras, y en el que tras la recombinación específica de sitio, el primer inserto está ubicado dentro del vector, y el vector contiene múltiples elementos de ADN y múltiples elementos de código de barras, con dos elementos de recombinación específica de sitio ubicados entre los múltiples elementos de ADN y los múltiples elementos de código de barras; proporcionar un segundo inserto que comprende: un tercer elemento de ADN, un tercer elemento de código de barras y elementos de recombinación específica de sitio que flanquean a cada uno del tercer elemento de ADN y el tercer elemento de código de barras, de manera que dos elementos de recombinación específica de sitio están ubicados entre el tercer elemento de ADN y el tercer elemento de código de barras que no son compatibles con los dos elementos de recombinación específica de sitio ubicados entre los múltiples elementos de ADN y los múltiples elementos de código de barras del vector, y dos elementos de recombinación específica de sitio están ubicados fuera del tercer elemento de ADN y el tercer elemento de código de barras que son compatibles con los dos elementos de recombinación específica de sitio ubicados entre los múltiples elementos de ADN y los múltiples elementos de código de barras del vector; realizar recombinación específica de sitio entre el vector y el segundo inserto, en el que la recombinación específica de sitio se produce entre los elementos de recombinación específica de sitio dentro del vector ubicados entre los múltiples elementos de ADN y los múltiples elementos de código de barras y los elementos de recombinación específica de sitio compatibles dentro del segundo inserto ubicados fuera del tercer elemento de ADN y el tercer elemento de código de barras, y en

el que tras la recombinación específica de sitio, el segundo inserto está ubicado dentro del vector, y el vector contiene múltiples elementos de ADN y múltiples elementos de código de barras, con dos elementos de recombinación específica de sitio ubicados entre los múltiples elementos de ADN y los múltiples elementos de código de barras; y repetir la recombinación específica de sitio un número n de veces, alternando entre recombinación específica de sitio entre el vector y el primer inserto y recombinación específica de sitio entre el vector y el segundo inserto, creando de ese modo un constructo genético combinatorio.

Aspectos adicionales de la descripción se refieren a un constructo genético combinatorio producido mediante tales métodos. Aspectos adicionales de la descripción se refieren a un método para la identificación de un elemento de ADN o una pluralidad de elementos de ADN dentro de un constructo genético combinatorio de este tipo, que comprende: proporcionar un constructo genético combinatorio; realizar un ensayo para determinar la secuencia de ADN de uno o más elementos de código de barras dentro del constructo genético combinatorio y/o la secuencia de ADN de uno o más elementos de ADN dentro del constructo genético combinatorio; e identificar el elemento de ADN o la pluralidad de elementos de ADN.

Aspectos adicionales de la descripción se refieren a métodos para generar un constructo genético combinatorio, que comprende: proporcionar un vector que comprende: un primer elemento de ADN, un primer elemento de código de barras y un sitio de reconocimiento para una primera enzima de restricción ubicado entre el primer elemento de ADN y el primer elemento de código de barras; proporcionar un inserto que comprende: un segundo elemento de ADN, un segundo elemento de código de barras, un sitio de reconocimiento para la primera enzima de restricción ubicado entre el segundo elemento de ADN y el segundo elemento de código de barras, y dos sitios de reconocimiento para una o más enzimas de restricción que son distintas de la primera enzima de restricción ubicados fuera del segundo elemento de ADN y segundo elemento de código de barras, de manera que la digestión por restricción en el sitio de reconocimiento dentro del vector y en los dos sitios de reconocimiento ubicados fuera del segundo elemento de ADN y segundo elemento de código de barras dentro del inserto genera extremos compatibles; digerir el vector e inserto con enzimas de restricción; aparear el inserto al vector, produciendo de ese modo un constructo genético combinatorio que contiene múltiples elementos de ADN y múltiples elementos de código de barras; y opcionalmente repetir el método un número n de veces.

Aspectos adicionales de la descripción se refieren a un constructo genético combinatorio producido mediante tales métodos. Aspectos adicionales de la descripción se refieren a métodos para la identificación de un elemento de ADN o una pluralidad de elementos de ADN dentro de un constructo genético combinatorio de este tipo, que comprende: proporcionar un constructo genético combinatorio; realizar un ensayo para determinar la secuencia de ADN de uno o más elementos de código de barras dentro del constructo genético combinatorio y/o la secuencia de ADN de uno o más elementos de ADN dentro del constructo genético combinatorio; e identificar el elemento de ADN o la pluralidad de elementos de ADN.

Estos y otros aspectos de la invención, así como diversas realizaciones de los mismos, resultarán más evidentes en referencia a los dibujos y la descripción detallada de la invención.

Cada una de las limitaciones de la invención puede abarcar diversas realizaciones de la invención. Por tanto, se anticipa que cada una de las limitaciones de la invención que implican cualquier elemento o combinaciones de elementos puede incluirse en cada aspecto de la invención. Esta invención no está limitada en su aplicación a los detalles de construcción y la disposición de componentes expuestos en la siguiente descripción o ilustrados en los dibujos. La invención es capaz de otras realizaciones y de ponerse en práctica o llevarse a cabo de diversos modos.

Breve descripción de los dibujos

No se pretende que los dibujos adjuntos estén dibujados a escala. En los dibujos, cada componente idéntico o casi idéntico que se ilustra en diversas figuras se representa mediante el mismo número. Para propósitos de claridad, no todos los componentes pueden estar marcados en cada dibujo. En los dibujos:

La figura 1 presenta un esquema que representa realizaciones no limitativas de la invención. La figura 1A muestra la creación de un constructo con código de barras único para cada elemento de ADN único. La figura 1B muestra que constructos genéticos asociados con aspectos de la invención pueden dar lugar a un vector (por ejemplo, por medio de escisión con enzimas de restricción) y a un inserto homólogo (por ejemplo, por medio de PCR). Los extremos compatibles pueden aparearse y quedar ligados. La figura 1C muestra que constructos combinatorios asociados con aspectos de la invención permiten la identificación rápida de elementos de ADN por medio de secuenciación de los códigos de barras únicos (flecha), y la retención de sitios de separación para permitir inserciones adicionales.

La figura 2 demuestra combinaciones de un solo elemento único consigo mismo. La figura 2A representa la electroforesis en gel de un plásmido no cortado que muestra un aumento de aproximadamente ~1 kb con cada adición, correspondiente a la longitud de la inserción. La figura 2B representa un hidrolizado de restricción en dos sitios que flanquean a la región de constructo, mostrando longitudes crecientes.

La figura 3 representa la combinación y expresión funcional de constructos de mCherry y GFP. Se representan

imágenes fluorescentes usando filtros para mCherry (figura 3A), GFP (figura 3B) y superposición (figura 3C).

La figura 4 representa los resultados de secuenciación y la alineación de cuatro clones de una reacción combinatoria. En la parte inferior se muestran anotaciones de secuencias. Las secuencias sombreadas en oscuro representan porciones únicas del código de barras. Se recuperaron las cuatro posibles permutaciones de dos constructos con código de barras, junto con la conservación del sitio de separación XbaI a la derecha.

La figura 5 representa la tecnología de concepción de sitios de restricción asociada con aspectos de la invención. Las enzimas de restricción BamHI y BglII generan proyecciones compatibles que, cuando se ligan, forman una cicatriz no reconocida por ninguna enzima. El sitio de separación consiste en los sitios de restricción BglII y EcoRI. La digestión del vector y el inserto con diferentes enzimas de restricción permite la inserción del inserto con retención del sitio de separación para combinaciones adicionales.

La figura 6 representa la generación fácil de combinaciones de orden superior. La figura 6A muestra secuencias verificadas para combinaciones por parejas, 3x y 4x a partir de colonias individuales. Los sitios de restricción SpeI y AvrII flanquean a la región de código de barras. La figura 6B muestra la digestión por restricción de aislados de plásmido para un gen con código de barras solo (1x), y combinaciones 2x-4x. Cada plásmido se digirió para separar la región combinatoria variable (indicada por puntos) del resto constante del vector.

La figura 7 representa la evaluación cuantitativa de combinaciones de GFP y mCherry por medio de ensamblaje de sitios de restricción. Las muestras n.º 1-24 son colonias aisladas de la transformación de la colección combinatoria en células de *E. coli* DH5a. Las combinaciones de GFP-GFP se representan mediante las muestras n.º 1, 8, 14, 15, 16, 18, 21, 22 y 24; las combinaciones de mCherry-mCherry mediante las n.º 7, 10, 11, 17, 20; las combinaciones de GFP y mCherry mediante las n.º 2-6, 8, 12, 13, 19, 23.

La figura 8 representa la unión de códigos de barras. Los códigos de barras se colocan en cualquier extremo de los elementos de ADN y se fusionan entre sí. Esta estrategia se limita a combinaciones por parejas y requiere colecciones diferenciadas para cada trozo. Esta figura está adaptada de Merryman (2012) *Metabolic Engineering* 14:196-204.

La figura 9 representa la tecnología de clonación Gateway® que permite la transferencia de ADN sin escisión del ADN (de <http://wolfson.huji.ac.il/expression/gatewayman.pdf>).

La figura 10 representa un ejemplo no limitativo de la tecnología de clonación Multisite Gateway® Pro para el ensamblaje de múltiples insertos dentro de un constructo genético (del sitio [web:pfgrc.jcvi.org/index.php/gateway_clones/about_knockoutclones.html](http://web.pfgrc.jcvi.org/index.php/gateway_clones/about_knockoutclones.html)).

La figura 11 representa una realización no limitativa mediante la cual dos colecciones de vectores con secuencias de recombinación únicas pueden generar, por medio de PCR, dos colecciones de insertos con secuencias de recombinación que flanquean al elemento de ADN y el elemento de código de barras dentro de cada inserto. Las secuencias de recombinación *att* están representadas por B, P, L o R. BC = código de barras, ADN = elemento de ADN.

La figura 12 representa la generación iterativa de combinaciones de orden superior por medio de genética combinatoria mediada por recombinación usando una colección de vectores y dos colecciones de insertos. El proceso iterativo alterna entre las dos colecciones de insertos. Los códigos de barras están en proximidad inmediata entre sí, permitiendo la determinación de los códigos de barras y, por tanto, la identidad de los elementos de ADN en el sentido de 3'.

La figura 13 representa una realización no limitativa mediante la cual las dos secuencias de recombinación en el medio de un constructo genético son ortogonales a los extremos todas las veces para evitar la autorrecombinación.

La figura 14 representa una realización no limitativa de la invención mediante la cual los extremos reactivos (B1 + B2, L3 + L4) se añaden mediante PCR. Esta realización evita que el vector reaccione finalmente consigo mismo.

La figura 15 representa una realización no limitativa de genética combinatoria masivamente en paralelo usando un solo sitio de separación enzimática. En esta realización, una sola enzima, la enzima 1, escinde en el sitio de separación del vector, generando una proyección en 3' de 3 bases. El elemento de código de barras y elemento de ADN dentro del inserto están flanqueados por sitios de restricción para una enzima diferente, la enzima 2, que genera una proyección compatible con la enzima 1. La digestión y el ligamiento del vector y el inserto producen un vector que contiene el inserto, con un sitio de restricción para la enzima 1 mantenido en el vector.

La figura 16A representa una realización no limitativa de una estrategia de ensamblaje por genética combinatoria masivamente en paralelo. Los constructos de expresión de factor de transcripción (TF) tienen código de barras (BC), y cuatro sitios de restricción (1A, 1B, 2A, 2B) están situados tal como se muestra. Las parejas, 1A/1B y 2A/2B, son sitios de restricción únicos que generan proyecciones compatibles dentro de una pareja pero que son incompatibles

con la otra pareja. Los vectores con código de barras se agrupan y se digieren con las enzimas 1B + 2A. Se generan insertos a partir de los vectores mediante PCR y se digieren con 1A + 2B. La figura 16B representa la distribución de lecturas de secuenciación de alto rendimiento entre combinaciones en la colección por parejas. La figura 16C representa la expresión de constructos de genética combinatoria masivamente en paralelo. Las cuatro combinaciones por parejas de constructos de GFP y mCherry y constructos de GFP y mCherry individuales se indujeron con aTc a 250 ng/ml y se evaluaron mediante citometría de flujo.

La figura 17 representa un mapa de plásmido que muestra un constructo de expresión de un solo gen (GFP) con código de barras y flanqueado por sitios de restricción para ensamblaje de colección por genética combinatoria masivamente en paralelo.

La figura 18A representa una realización no limitativa de un método de genética combinatoria masivamente en paralelo. Se diluyen células NDM-1 de *E. coli* que contienen una colección de genética combinatoria masivamente en paralelo en cultivos con y sin aTc y se hacen crecer hasta la fase semilogarítmica. Entonces se diluye adicionalmente cada cultivo en cultivos con y sin antibiótico y con y sin aTc. Se recoge el ADN de cada condición en las fases de crecimiento logarítmica temprana y logarítmica tardía y se procesan para secuenciación de alto rendimiento. Las comparaciones de abundancias de combinación entre condiciones diferentes revelan genotipos que conducen a los fenotipos deseados. La figura 18B representa un mapa de calor de resumen de "aciertos". La agrupación jerárquica de puntuaciones S de combinación a través de los experimentos muestra perfiles de fenotipos diferentes. La sinergia de área marcada muestra combinaciones que se abandonan tras la adición de tanto ceftriaxona como aTc. Lo opuesto, el crecimiento, muestra combinaciones que están relativamente sobrerrepresentadas tras la adición de tanto ceftriaxona como aTc.

La figura 19 representa la agrupación jerárquica de combinaciones de genes y recíprocos. Los números 1-8 indican parejas de genes correspondientes a las de la figura 18B. R indica la disposición recíproca de la pareja de genes. Las parejas de genes que confieren el mismo fenotipo y sus recíprocos se agrupan juntos.

La figura 20A representa gráficos que muestran la sinergia para los fenotipos de sensibilidad a antibiótico de NDM-1. Fenotipos de sensibilidad a antibiótico de NDM-1. La figura 20B representa gráficos que muestran controles para fenotipos de sensibilidad a antibiótico de NDM-1. Las combinaciones muestran una potenciación minoritaria de ceftriaxona. El comienzo de la acción de eCFP + norR está retrasado en comparación con adiY + marA. La figura 20C representa gráficos que muestran la ventaja de crecimiento para fenotipos de sensibilidad a antibiótico de NDM-1. Las combinaciones muestran un efecto mínimo tras la inducción a lo largo de todo el transcurso del experimento.

Las figuras 21A y 21B representan gráficos que muestran la sensibilidad a antibiótico de genes individuales.

La figura 22 representa la agrupación jerárquica a través de los antibióticos ceftriaxona, imipenem, piperacilina-tazobactam y gentamicina. El tratamiento intenso de beta-lactamas se agrupa junto, por separado de gentamicina.

La figura 23A representa gráficos que muestran que las combinaciones potencian la letalidad de los antibióticos en 4-6 órdenes de magnitud cuando se inducen y suprimen el crecimiento durante más de 10 horas. La figura 23B representa gráficos que muestran que combinaciones letales aleatorias tales como *cadC* + *allR* muestran una amplificación significativamente menos potente de ceftriaxona y una supresión del crecimiento menos sostenida a lo largo del tiempo. La figura 23C representa gráficos que muestran que *torR* y *metR* muestran individualmente una amplificación significativamente más débil de ceftriaxona en comparación con la combinación *torR* + *metR*. La figura 23D representa gráficos que muestran que fagémidos que contienen *torR* + *metR* y *nhaR* + *meiR* que infectan células de NDM1 muestran una destrucción amplificada de cultivos con ceftriaxona e inducción.

La figura 24 representa gráficos de curvas de inducción fluorescentes.

La figura 25 representa la agrupación jerárquica de aciertos superiores a través de todas las condiciones de antibiótico-aTc. 6311 parejas de genes muestran al menos una puntuación S por encima de 2,4 o por debajo de -2,4, en donde 2,4 es tres veces la desviación estándar de la distribución en la población de la puntuación S. Se ha empleado la correlación euclídea con la unión promedio en el análisis de agrupación. Los aciertos para las betalactamas imipenem, ceftriaxona y piperacilina-tazobactam en presencia de aTc se agrupan juntos (corchete rojo), por separado de gentamicina.

La figura 26 representa un gráfico que muestra una colección combinatoria de tríos de complejidad inferior que combina 189 x 187 x 18 genes. De las 636174 combinaciones posibles, se recuperaron 582.433 combinaciones únicas (el 92 % de todas las combinaciones posibles) a partir de aproximadamente 30 millones de lecturas de secuenciación.

La figura 27 representa un gráfico que muestra una colección combinatoria de tríos de complejidad superior que combina 189 x 187 x 187 genes. De las 6.609.141 combinaciones de genes posibles, se recuperaron 4.042.316 combinaciones (61 %) a partir de aproximadamente 20 millones de lecturas de secuenciación.

La figura 28 representa las cepas de cribado genético generadas integrando múltiples copias en tándem de α Syn con un promotor inducible por galactosa (Gal). La expresión de colecciones combinatorias se controló mediante el sistema inducible Tet-ON. Los experimentos se diseñaron para recuperar combinaciones genéticas que contribuyen a la supervivencia celular con tratamiento con Gal y doxiciclina (Dox).

5 La figura 29 representa un constructo de expresión individual de un código de barras (BC) único y cuatro sitios de restricción. Los vectores con código de barras se agrupan y se digieren con las enzimas AvrII + NotI. Se generan insertos mediante digestión con SpeI + PspOMI. Una reacción de ligamiento en un solo recipiente produce una colección combinatoria por parejas.

10 La figura 30 representa un gráfico que muestra el análisis cuantitativo de la expresión génica de combinaciones por parejas de eGFP y mCherry construidos mediante estrategias de genética combinatoria masivamente en paralelo y P2A.

15 La figura 31 representa un cribado genético combinatorio masivamente en paralelo mediante programación transcripcional guiada por ARN.

La figura 32 representa un cribado genético combinatorio masivamente en paralelo de atenuaciones genéticas.

20 La figura 33 representa diseños de vectores lentivíricos. Se construyeron unidades transcripcionales en tándem que expresan precursor de miR junto con un gen de GFP dirigido por CMVp y/o una secuencia de sensor de miR (por ejemplo, cuatro repeticiones de las secuencias diana complementarias de miR) en la UTR en 3' de un gen de RFP dirigido por Ubcp, en una estructura principal de vector lentivírico. Ubcp, promotor de ubiquitina C humana; CMVp, promotor de citomegalovirus humano; RFP, proteína fluorescente roja; GFP, proteína fluorescente verde.

25 La figura 34 representa gráficos que muestran la expresión de precursor de miR en células HEK293T. Se usaron lentivirus que contenían unidades transcripcionales en tándem que expresaban tres precursores de miR diferentes (es decir, miR-124, miR-128 y miR132) junto con un gen de GFP dirigido por CMVp y/o una secuencia de sensor de miR en la UTR en 3' de un gen de RFP dirigido por Ubcp para infectar células HEK293T. Se determinó el porcentaje de células RFP⁺ con respecto a células GFP⁺ mediante citómetro de flujo tras separar con dispersión frontal y lateral. HEK293T; 4 días tras la infección.

30 La figura 35 representa gráficos que muestran la expresión de precursor de miR en células de fibroblastos dérmicos humanos primarios (HDF). Se usaron lentivirus que contenían unidades transcripcionales en tándem que expresaban dos precursores de miR diferentes (es decir, miR-128 y miR132) (SEQ ID NO: 1) junto con un gen de GFP dirigido por un promotor de CMV (CMVp) y/o una secuencia de sensor de miR en la UTR en 3' de un gen de RFP dirigido por Ubcp para infectar células HDF. Se determinó el porcentaje de células RFP⁺ con respecto a células GFP⁺ mediante citómetro de flujo tras separar con dispersión frontal y lateral. HEK293T; 7 días tras la infección.

35 La figura 36 representa gráficos que muestran la expresión de precursor de miR combinatorio en células HEK293T. Se usaron lentivirus que contenían unidades transcripcionales en tándem que expresaban precursores de miR individuales o combinatorios (es decir, miR-128, miR-132 o miR-128-132) junto con un gen de GFP dirigido por CMVp y/o una secuencia de sensor de miR en la UTR en 3' de un gen de RFP dirigido por Ubcp para infectar células HEK293T. Se determinó el porcentaje de células RFP⁺ con respecto a células GFP⁺ mediante citómetro de flujo tras separar con dispersión frontal y lateral. HEK293T; 4 días tras la infección.

40 La figura 37 representa gráficos que muestran la expresión de precursor de miR combinatorio en células HDF primarias. Se usaron lentivirus que contenían unidades transcripcionales en tándem que expresaban precursores de miR individuales o combinatorios (es decir, miR-128, miR-132 o miR-128-132) junto con un gen de GFP dirigido por CMVp y/o una secuencia de sensor de miR en la UTR en 3' de un gen de RFP dirigido por Ubcp para infectar células HDF. Se determinó el porcentaje de células RFP⁺ con respecto a células GFP⁺ mediante citómetro de flujo tras separar con dispersión frontal y lateral. HEK293T; 7 días tras la infección.

45 La figura 38 representa un mapa de un vector que contiene mir-128 y mir-132 en tándem.

55 **Descripción detallada**

Se describe en el presente documento una tecnología nueva, denominada genética combinatoria masivamente en paralelo, que supera los obstáculos que han limitado campos tales como la genética y la biología de sistemas. La invención se basa, al menos en parte, en el descubrimiento sorprendente de métodos y composiciones que permiten la generación rápida de combinaciones de orden alto de elementos genéticos y la identificación rápida de combinaciones que conducen a un fenotipo deseado. Esta tecnología permite nuevos métodos de investigación y es ampliamente aplicable. El enfoque descrito en el presente documento presenta ventajas significativas con respecto a métodos previos, incluyendo flexibilidad de la perturbación genética permitida, facilidad de uso y capacidad de aumento a escala rápido a combinaciones de orden alto.

En el pasado, los estudios de perturbación sistemática a gran escala se han visto limitados a perturbaciones de baja complejidad. Por ejemplo, la colección ASKA es una colección de modelos de sobreexpresión de un solo gen para los ~4000 marcos de lectura abiertos ("ORF") en el genoma de *E. coli*. Los ORF se basan en plásmidos y se disponen en un gran número de placas. Los estudios han cribado esta colección de sobreexpresión individual para detectar fenotipos deseados, tales como resistencia a antibióticos, e identificaron clones a través de secuenciación individual (Soo *et al.* (2011) Proc Natl Acad Sci USA 108:1484-9). Pudo derivarse teóricamente una estrategia combinatoria por parejas a partir de este enfoque generando otra colección de ORF en un plásmido con un marcador de selección diferente, y cotransformando una cepa con un miembro de la colección ASKA y un miembro de la nueva colección. Sin embargo, un intento de generar los ~16 millones de combinaciones de ORF por parejas estaría impedido inmediatamente por limitaciones de escala. Esto es incluso más evidente cuando se generan combinaciones de orden superior; por ejemplo, la baja eficiencia de la transformación simultánea de múltiples vectores y el bajo número de marcadores de selección únicos impiden combinaciones de orden alto.

Impedimentos similares afectan a enfoques previos para desactivaciones combinatorias. Se han generado colecciones de desactivación de un solo gen para *E. coli* (por ejemplo, colección KEIO) y para *S. cerevisiae*. Las colecciones con código de barras permiten el cribado agrupado cuando se emparejan con secuenciación de alto rendimiento. Sin embargo, no ha habido ningún método eficaz de generación de colecciones de desactivación de genes doble. La matriz genética sintética (Tong *et al.* (2004) Science 303:808-13) emplea acoplamiento de levaduras para generar desactivaciones dobles en un formato basado en placa; se ha notificado un método similar en *E. coli* (Butland *et al.* (2008) Nature Methods 5:789-95). Sin embargo, debido a que el acoplamiento se produce cuando dos cepas únicas se disponen conjuntamente sobre una placa, el aumento a escala más allá de miles de combinaciones no es práctico. Además, cada experimento posterior requiere un proceso laborioso para volver a sembrar en placa las células.

Combinaciones de elementos genéticos han sido útiles en varias aplicaciones específicas, tales como factores de transcripción en diferenciación de células madre. En el pasado, el método típico ha sido cotransfectar un tipo de célula con muchos factores de transcripción individuales y observar la diferenciación; cada miembro de esta agrupación se retira entonces uno a uno hasta que se identifica el conjunto mínimo para inducir diferenciación (Son *et al.* (2011) Cell Stem Cell 9:205-18). Este método es laborioso, ya que cada combinación única requiere un experimento diferenciado, y tiene baja eficiencia, ya que requiere cotransfección con muchos elementos individuales. Estos problemas se agravan por el hecho de que se observan a menudo fenotipos de diferenciación semanas tras la transfección.

Un reciente trabajo ha ilustrado el concepto de "unión" de código de barras consecutiva (Merryman (2012) Metabolic Engineering 14:196-204; Roth *et al.* (publicación de patente estadounidense n.º US 2009/0098555)) en la que dos elementos de ADN con códigos de barras en el extremo 3' y el extremo 5' respectivamente se fusionan entre sí, colocando dos códigos de barras en proximidad (figura 8). Los métodos usados para lograr esto incluyen solapamiento por PCR y ensamblaje de Gibson. La proximidad permite que las combinaciones se rastreen por medio de secuenciación. Sin embargo, la estrategia de ese enfoque no puede aumentarse a escala más allá de combinaciones por parejas. Además, cada trozo con código de barras requiere la construcción de una colección diferenciada en lugar de utilizar la misma colección de insertos de manera iterativa.

El enfoque de genética combinatoria masivamente en paralelo descrito en el presente documento ofrece múltiples ventajas en relación con los métodos previos. Permite la generación rápida de conjuntos combinatorios de una variedad de elementos genéticos, tales como factores de transcripción. Además, esta tecnología permite el cribado agrupado de múltiples órdenes de combinación (por ejemplo, pueden agruparse combinaciones por parejas, por tríos y de n elementos y cribarse juntas simultáneamente), identificando las combinaciones mínimas necesarias para una aplicación dada.

Una ventaja adicional de la tecnología descrita en el presente documento con respecto a los métodos anteriores es su flexibilidad en el tipo de perturbación del ADN. Los métodos anteriores, tales como el método de matriz genética sintética, por ejemplo, son adecuados para una técnica, tal como para generar desactivaciones, pero no son adecuados para otros tipos de perturbación. En cambio, la tecnología descrita en el presente documento permite combinaciones sin precedentes de atenuaciones y sobreexpresión génica. Una ventaja adicional con respecto a los métodos anteriores es que el enfoque descrito en el presente documento puede aumentarse a escala más allá de combinaciones por parejas.

Aspectos del enfoque de genética combinatoria masivamente en paralelo se refieren a constructos genéticos que incluyen uno o más elementos de ADN y uno o más elementos de código de barras, en los que cada elemento de ADN específico está asociado con un elemento de código de barras único. Tal como se usa en el presente documento, asociación entre un elemento de ADN específico y un elemento de código de barras único significa que un elemento de ADN específico y un elemento de código de barras único están contenidos siempre dentro del mismo constructo genético. Por consiguiente, la presencia de un elemento de código de barras único dentro de un constructo genético indica que el elemento de ADN específico asociado también está presente dentro del mismo constructo genético.

Debe apreciarse que un elemento de ADN puede incluir cualquier trozo de ADN y puede tener cualquier función o secuencia. En algunas realizaciones, el elemento de ADN comprende un gen o fragmento de gen que puede codificar o no una proteína. Un elemento de ADN puede incluir tanto regiones codificantes como no codificantes. Por ejemplo, en algunas realizaciones un elemento de ADN puede incluir un marco de lectura abierto o fragmento del mismo, un sitio de unión al ribosoma, un promotor y/o un terminador. En algunas realizaciones, el elemento de ADN comprende un ADN no codificante, ADN monocatenario o un precursor de ARN tal como ARN no codificantes largos (ARNncl), microARN (miARN)/ARN de interferencia pequeño (ARNip) o ARN en horquilla corto. En algunas realizaciones, el elemento de ADN es un gen implicado en estudio de asociación de todo el genoma (GWAS) o un elemento de ADN de colecciones de ORF de todo el genoma. El elemento de ADN también puede incluir una muestra metagenómica o uno o más componentes de un circuito de biología sintético.

En algunas realizaciones, el/los elemento(s) de ADN codifica(n) un factor de transcripción (incluyendo factores de transcripción endógenos y artificiales), enzimas de modificación de histonas, un microARN (miR), una cinasa o fosfatasa, una enzima metabólica, una enzima epigenética, una diana de un fármaco aprobado por la FDA, un oncogén, un anticuerpo monoclonal y/o una proteína mutante.

La figura 1 presenta varios esquemas de ejemplos no limitativos de constructos genéticos asociados con la invención. En la figura 1A, un elemento de ADN, designado ADN-1, está flanqueado por un primer elemento de extremos compatibles, G2A y un segundo elemento de extremos compatibles, G2B, que son capaces de aparearse entre sí. El constructo genético también contiene un elemento de código de barras, designado como BC-1, que está flanqueado por un tercer elemento de extremos compatibles, G1A y un cuarto elemento de extremos compatibles, G1B, que son capaces de aparearse entre sí pero no son capaces de aparearse a G2A o G2B. El constructo genético también contiene un sitio de separación, de manera que el elemento de código de barras está en un lado del sitio de separación, mientras que el elemento de ADN está en el otro lado del sitio de separación. Aunque la figura 1 representa el elemento de código de barras como que está en el sentido de 5' o en 5' en relación con el elemento de ADN, esta disposición también puede revertirse.

Pueden crearse elementos de extremos compatibles de una variedad de modos familiares para un experto habitual en la técnica y pueden consistir en una variedad de secuencias diferentes. Tal como se usa en el presente documento, elementos de extremos compatibles se refieren a regiones de ADN que son capaces de ligarse o aparearse entre sí. En varias realizaciones no limitativas, los elementos de extremos compatibles pueden estar compuestos por sitios de restricción con proyecciones compatibles, secuencias de ensamblaje de Gibson o elementos funcionales de cualquier otro método de ensamblaje de ADN, incluyendo recombinasas, meganucleasas, nucleasas de dedos de zinc/efector TAL, ADNzimas/ribozimas de escisión en trans o integrasas.

En algunas realizaciones, se usa ensamblaje de Gibson para generar proyecciones compatibles. El ensamblaje de Gibson se refiere a una técnica de unión de extremos de ADN isotérmica mediante la cual múltiples fragmentos de ADN pueden unirse en una única reacción. Este método se describe adicionalmente en Gibson *et al.* (2009) *Nature Methods* 6:343-5.

En otras realizaciones, se usa digestión de sitios de restricción para generar extremos compatibles, tal como se representa en la figura 5. Usando este método, dos enzimas de restricción únicas generan proyecciones compatibles. Cuando se ligan estas proyecciones, se crea una cicatriz que ya no reconoce ninguna enzima. Debe apreciarse que puede usarse cualquier enzima de restricción que genere proyecciones compatibles. En algunas realizaciones no limitativas, se usan partes biológicas convencionales tales como BioBricks® (The BioBricks Foundation) o BglBricks (Anderson *et al.* (2010) *Journal of Biological Engineering* 4:1), y enzimas asociadas con tales partes biológicas convencionales. El uso de partes biológicas convencionales tales como BioBricks® o BglBricks se consideraría de rutina para un experto habitual en la técnica. Debe apreciarse que aunque pueden usarse enzimas de restricción (tales como enzimas de restricción de tipo I, II o III), pueden usarse también otras moléculas de escisión del ADN. Por ejemplo, pueden usarse ribozimas dirigidas para la escisión de sitios diana específicos. También pueden utilizarse meganucleasas para minimizar la posibilidad de interferencia con los elementos de ADN insertados. También pueden usarse nucleasas TALE o ZF para seleccionar como diana sitios de ADN largos para minimizar la probabilidad de escisión interna dentro de los elementos de ADN insertados. Además, puede usarse clonación TOPO® para lograr digestiones y ligamientos por restricción.

El sitio de separación dentro del constructo genético representa la región que permite la linealización del constructo. Debe apreciarse que el sitio de separación puede corresponder a cualquier medio de escisión de ADN. En algunas realizaciones, el sitio de separación es un sitio de reconocimiento de enzimas de restricción.

Aspectos adicionales de la invención se refieren a constructos combinatorios, y a métodos para producir constructos combinatorios. Tal como se usa en el presente documento, un "constructo combinatorio" se refiere a un constructo genético que contiene una pluralidad de elementos de ADN. Tal como se usa en el presente documento, una pluralidad de elementos de ADN se refiere a más de un elemento de ADN. Tal como se muestra en la figura 1B, la generación de un constructo combinatorio puede implicar la linealización de un vector que contiene un primer constructo genético asociado con la invención, escindiendo el vector en el sitio de separación dentro del constructo genético. Un segundo constructo genético asociado con la invención se representa en la figura 1B como un inserto.

Tal como se usa en el presente documento, un "inserto" se refiere a un constructo genético que está destinado a insertarse dentro de un vector escindido. En algunas realizaciones, el inserto se purifica de un vector, tal como mediante PCR o digestión por restricción. El inserto puede ligarse al vector escindido a través del apareamiento de los elementos de extremos compatibles terminales dentro del inserto y sus componentes compatibles dentro del vector linealizado.

La figura 1C representa un constructo combinatorio tras la combinación que contiene una pluralidad de elementos de ADN y una pluralidad de elementos de código de barras correspondientes. En el ejemplo no limitativo representado en la figura 1C, el constructo genético contiene dos elementos de ADN diferentes, denominados ADN-1 y ADN-2 y dos elementos de código de barras correspondientes, denominados BC-1 y BC-2. El constructo combinatorio contiene además un sitio de separación, ubicado entre la pluralidad de elementos de código de barras y la pluralidad de elementos de ADN.

El sitio de separación puede ser un solo sitio de reconocimiento de enzimas de restricción. La figura 15 muestra una realización no limitativa de un vector que comprende un elemento de ADN y un elemento de código de barras separados por un solo sitio de enzimas de restricción, tal como un sitio BgIII. Un ejemplo no limitativo de un inserto correspondiente, mostrado en la figura 15, contiene un elemento de ADN y un elemento de código de barras. En algunas realizaciones, un sitio de restricción, tal como un sitio BgIII, está ubicado entre el elemento de ADN y el elemento de código de barras del inserto, mientras que dos sitios de restricción, tales como sitios AlwNI, están ubicados fuera del elemento de ADN y el elemento de código de barras. BgIII y AlwNI generan extremos compatibles cuando escinden ADN. Por consiguiente, la digestión del vector con BgIII y el inserto con AlwNI permite el ligamiento del inserto dentro del vector, produciendo un vector que contiene dos elementos de ADN y dos elementos de código de barras separados por un sitio de restricción, tal como un sitio BgIII.

Debe apreciarse que puede usarse una variedad de combinaciones de enzimas diferentes que producen extremos compatibles cuando escinden ADN, conjuntamente con este aspecto de la invención. En algunas realizaciones, dentro del inserto, los dos sitios de restricción ubicados fuera del elemento de ADN y el elemento de código de barras los reconoce la misma enzima de restricción, que produce extremos compatibles con la enzima de restricción que digiere el vector. En otras realizaciones, dentro del inserto, los dos sitios de restricción ubicados fuera del elemento de ADN y el elemento de código de barras los reconocen dos enzimas de restricción diferentes, cada una de las cuales produce extremos compatibles con la enzima de restricción que digiere el vector.

Aspectos adicionales de la invención se refieren al uso de recombinación específica de sitio para generar constructos genéticos combinatorios. En tales realizaciones, constructos genéticos asociados con la invención no requieren un sitio de separación y no requieren escisión por enzimas de restricción. Un ejemplo no limitativo de un método de clonación basado en recombinación es la tecnología de clonación Gateway® (Life Technologies, Carlsbad, CA), que sería familiar para un experto habitual en la técnica.

La figura 9 proporciona un ejemplo de la metodología de clonación Gateway®. Elementos de ADN están flanqueados por elementos de recombinación específica de sitio, lo que permite la recombinación en presencia de mezclas de reacción enzimáticas Clonase™. Dentro del sistema Gateway, los elementos de recombinación específica de sitio se denominan sitios de unión o "sitios *att* (*attachment*)", e incluyen *attB1*, *attB2*, *attP1*, *attP2*, *attL1*, *attL2*, *attR1* y *attR2*. *attB1* reacciona con *attP1*, *attB2* reacciona con *attP2*, *attL1* reacciona con *attR1* y *attL2* reacciona con *attR2*. Debe apreciarse que cualquier método de clonación basado en recombinación puede ser compatible con aspectos de la invención.

La figura 10 representa un ejemplo no limitativo de la tecnología de clonación Multisite Gateway® Pro para ensamblar múltiples insertos dentro de un constructo genético (de http://pfgrc.jcvi.org/index.php/gateway_clones/about_knockoutclones.html).

La figura 11 representa una realización no limitativa de un enfoque de genética combinatoria basada en recombinación. En este ejemplo, se usan dos colecciones de vectores para generar dos colecciones de insertos. Cada vector contiene un elemento de ADN, un elemento de código de barras y dos elementos de recombinación específica de sitio. Debe apreciarse que pueden usarse diversos métodos para generar colecciones de insertos. En algunas realizaciones, se genera una colección de insertos, tal como la colección de insertos 1 o la colección de insertos 2 representada en la figura 11 a partir de colecciones de vectores tales como la colección de vectores 1 y la colección de vectores 2 representadas en la figura 11, tal como mediante PCR. En algunas realizaciones, se usa PCR para añadir una pareja flanqueante de secuencias de recombinación ortogonales a las colecciones de insertos.

La figura 12 representa una realización no limitativa mediante la cual se usa genética combinatoria basada en recombinación para generar un constructo genético combinatorio. En este ejemplo, se muestra un vector de la colección de vectores 1 que contiene un elemento de ADN, un elemento de código de barras y dos elementos de recombinación específica de sitio ubicados entre el elemento de ADN y el elemento de código de barras. El vector se recombina con un inserto de la colección 2 que contiene un elemento de ADN y un elemento de código de barras, cada uno de los cuales está flanqueado por elementos de recombinación específica de sitio. Dentro del inserto de la colección 2, están ubicados dos elementos de recombinación específica de sitio entre el elemento de ADN y el

elemento de código de barras, y están ubicadas dos secuencias de recombinación específica de sitio fuera del elemento de ADN y elemento de código de barras. Los elementos de recombinación específica de sitio que están ubicados fuera del elemento de ADN y el elemento de código de barras son compatibles con los dos elementos de recombinación específica de sitio en el vector ubicados entre el elemento de ADN y el elemento de código de barras dentro del vector.

Tras la recombinación en sitios compatibles, el inserto de la colección de insertos 2 está contenido dentro del vector, produciendo de ese modo un vector que contiene dos elementos de ADN y dos elementos de código de barras, con dos sitios de recombinación específica de sitio ubicados entre los elementos de ADN y los elementos de código de barras.

Tal como se representa en la figura 12, el vector puede recombinarse entonces con un inserto de la colección de insertos 1, que contiene un elemento de ADN y un elemento de código de barras, cada uno de los cuales está flanqueado por elementos de recombinación específica de sitio. Dentro del inserto de la colección 1, dos elementos de recombinación específica de sitio están ubicados entre el elemento de ADN y el elemento de código de barras, y dos secuencias de recombinación específica de sitio están ubicadas fuera del elemento de ADN y el elemento de código de barras. Los elementos de recombinación específica de sitio que están ubicados fuera del elemento de ADN y el elemento de código de barras son compatibles con los dos elementos de recombinación específica de sitio en el vector ubicados entre el elemento de ADN y el elemento de código de barras dentro del vector.

Tras la recombinación, el inserto de la colección de insertos 1 está contenido dentro del vector, produciendo de ese modo un vector que contiene tres elementos de ADN y tres elementos de código de barras, con dos sitios de recombinación específica de sitio ubicados entre los elementos de ADN y los elementos de código de barras.

Las figuras 13 y 14 representan cómo puede diseñarse un vector e inserto para que se recombinen entre sí al tiempo que se evita la recombinación vector-vector. En algunas realizaciones, se añaden elementos de recombinación específica de sitio a los constructos genéticos mediante PCR.

Los métodos descritos en el presente documento para generar constructos combinatorios pueden ser iterativos. Por ejemplo, el constructo combinatorio representado en la figura 1C generado a través de un acontecimiento de combinación, puede escindirse de nuevo en el sitio de separación, y uno o más insertos adicionales pueden ligarse en el constructo combinatorio, al tiempo que se mantiene un sitio de separación para inserciones adicionales. De manera similar, el vector en la figura 12 puede recombinarse alternativamente con insertos de la colección 1 y con insertos de la colección 2, generando un constructo genético combinatorio. Significativamente, a lo largo de todo el proceso iterativo, a medida que el número de elementos de ADN dentro del constructo genético sigue aumentando, los códigos de barras únicos asociados con cada elemento de ADN se mantienen dentro del mismo constructo genético que sus elementos de ADN asociados. En algunas realizaciones, el proceso de combinación se repite al menos 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 veces o más de 20 veces. En algunas realizaciones, el proceso se repite un número n de veces, en donde n puede ser 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, o un número mayor de 20.

Debe apreciarse que los constructos combinatorios pueden contener cualquier número de elementos de ADN y elementos de código de barras asociados. En algunas realizaciones un constructo combinatorio contiene 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100 o más de 100 elementos de ADN y elementos de código de barras asociados.

Aspectos adicionales de la invención se refieren a métodos para identificar uno o más elementos de ADN dentro de un constructo genético. Tras un acontecimiento de combinación, un código de barras único que está asociado con un elemento de ADN específico permanece dentro del mismo constructo genético que el elemento de ADN específico. Por consiguiente, la identificación de un elemento de código de barras o una pluralidad de elementos de código de barras permite la identificación del elemento de ADN asociado o la pluralidad de elementos de ADN dentro del mismo constructo genético. En algunas realizaciones, la secuencia de un elemento de código de barras y/o un elemento de ADN se determina mediante secuenciación o mediante análisis de micromatriz. Debe apreciarse que cualquier medio de determinación de la secuencia de ADN es compatible con la identificación de uno o más elementos de código de barras y elementos de ADN correspondientes. Significativamente, en un constructo combinatorio, tal como se representa en la figura 1C, la pluralidad de elementos de código de barras están dentro de una proximidad estrecha entre sí permitiendo la identificación rápida de múltiples elementos de código de barras, y por consiguiente múltiples elementos de ADN, simultáneamente a través de métodos tales como secuenciación de ADN.

Aspectos adicionales de la invención se refieren a colecciones que comprenden dos o más constructos genéticos tal como se describe en el presente documento que son compatibles con métodos para genética combinatoria masivamente en paralelo. Tal como se usa en el presente documento, una colección de constructos genéticos se refiere a una colección de dos o más constructos genéticos. En algunas realizaciones, se genera una colección de

constructos genéticos en la que cada elemento de ADN único está en un plásmido. Esta colección de plásmidos puede agruparse para formar una colección de vectores. Una colección de insertos puede generarse, por ejemplo, realizando PCR sobre la colección de vectores. En un primer acontecimiento de combinación, todos los vectores pueden emparejarse con todos los insertos, generando un conjunto combinatorio completo de combinaciones por parejas. Reacciones adicionales entre esta colección por parejas y una colección de insertos pueden conducir a una colección de tríos, de cuartetos o de más de cuatro elementos que surgen de una sola colección de vectores. La figura 6 demuestra la integración eficaz de combinaciones por parejas, por tríos y por cuartetos de un elemento de código de barras.

Algunos aspectos de la descripción se refieren a la realización de cribados usando una colección de constructos combinatorios. Por ejemplo, puede someterse a prueba una agrupación de células que contienen la colección combinatoria para determinar su resistencia a un factor estresante tal como quimioterapia. Para determinar qué combinación de elementos de ADN es la más eficaz para conferir un fenotipo particular, tal como resistencia a quimioterapia, pueden aislarse células que sobreviven a la quimioterapia y puede determinarse la secuencia del código de barras o de la pluralidad de códigos de barras, permitiendo la identificación rápida de un elemento de ADN o de una pluralidad de elementos de ADN que es eficaz para conferir un fenotipo deseado.

Debe apreciarse que puesto que la etapa combinatoria se realiza *in vitro*, esta tecnología puede aumentarse a escala hasta cualquier organismo que pueda recibir ADN. En algunas realizaciones, el organismo es una bacteria y los constructos se portan sobre plásmidos o fagos. En otras realizaciones, el organismo es una levadura y los constructos se portan sobre plásmidos o vectores lanzadera. En otras realizaciones, tales como en células de mamífero o roedor, los constructos genéticos descritos en el presente documento pueden portarse sobre plásmidos o suministrarse mediante virus tales como lentivirus o adenovirus.

Los métodos y las composiciones descritos en el presente documento son ampliamente aplicables a cualquier estudio que pueda beneficiarse de la generación de conjuntos combinatorios de elementos genéticos. Por ejemplo, este enfoque podría conducir a la identificación de dianas farmacológicas nuevas dilucidadas por perturbación de redes, que podría definir rutas enzimáticas más sutiles que conducen a una enfermedad, o permitir el descubrimiento de fármacos de mediadores biológicos o químicos nuevos (incluyendo combinaciones de mediadores químicos y/o biológicos) para tratar una enfermedad. Adicionalmente, las tecnologías descritas en el presente documento podrían aplicarse al descubrimiento de combinaciones de dianas farmacológicas existentes para el tratamiento y/o la prevención de enfermedades, y podrían conducir a tratamientos de combinación nuevos usando productos terapéuticos aprobados por la FDA.

Varios ejemplos no limitativos de modos en los que puede perturbarse la expresión génica según aspectos de la invención incluyen: sobreexpresión fuerte, sobreexpresión ajustable (por medio de promotores inducibles ajustables), atenuación fuerte (por medio de ARN en horquilla corto (ARNhc)) u otros constructos de ARN antisentido) y atenuación ajustable (por medio de ARNhc u otros constructos de ARN antisentido y promotores inducibles ajustables).

Varios ejemplos no limitativos de fenotipos de interés que pueden cribarse o seleccionarse según aspectos de la invención incluyen, en bacterias y hongos (tales como levaduras): sensibilidad o resistencia a antibióticos, persistencia, virulencia y modificación metabólica por ingeniería genética; y en células de mamífero: reducción de un estado patológico, producción de un estado patológico, enfermedades multifactoriales complejas, envejecimiento y enfermedades relacionadas con la edad, neurodegeneración, resistencia a la quimioterapia, modulación de rutas (por ejemplo, respuesta al estrés), resistencia a la infección, diferenciación de células madre, transdiferenciación de tipos de células y potenciación de fármacos aprobados por la FDA.

La capacidad para ajustar la sobreexpresión o atenuación por medio de promotores inducibles ajustables permite un conjunto de perturbaciones mucho más matizadas que las previamente posibles. Por ejemplo, pueden combinarse múltiples niveles de sobreexpresión para muchas proteínas y cribarse para detectar un fenotipo óptimo, dado que la dinámica de la red puede perturbarse de la mejor manera a través de la optimización de los niveles de expresión. Además, empleando promotores inducibles independientemente para diferentes elementos de ADN, pueden ejecutarse múltiples experimentos simultáneos para sondaje del espacio de expresión o espacio temporal.

Esta tecnología permite una multitud de aplicaciones biomédicas directas. El aprovechamiento de vectores de terapia génica podría dar como resultado el suministro de conjuntos combinatorios de constructos de atenuación y sobreexpresión *in vitro* e *in vivo*. Las aplicaciones biomédicas no limitativas de esta tecnología incluyen el tratamiento de enfermedades multifactoriales con fenotipos complejos y aplicaciones de ingeniería de tejidos mediante las cuales se modifican células y luego se implantan en humanos. Los enfoques descritos en el presente documento podrían aplicarse a productos terapéuticos de anticuerpos combinatorios que implican combinaciones de anticuerpos monoclonales y/o policlonales. De manera similar, podrían optimizarse vacunas combinatorias a través de la selección combinatoria de múltiples epítomos.

Además, la tecnología permite descubrimientos fundamentales con amplio potencial. Un ejemplo es el mapeo global de interacciones de proteínas *in vivo*. Midiendo los efectos de combinaciones por parejas o de orden superior de

sobreexpresión y atenuación de genes sobre las tasas de crecimiento y otros fenotipos, puede obtenerse un mapeo exhaustivo de interacciones de proteínas. A partir de estos datos, pueden dilucidarse núcleos de redes y rutas de interacción, conduciendo a una concepción mucho más matizada de la red proteínica intracelular y al descubrimiento de perturbaciones de redes nuevas para fenotipos deseados.

Otra aplicación fundamental de los métodos, constructos y colecciones descritos en el presente documento es para el descubrimiento de determinantes de enfermedades multifactoriales. Aunque se han notificado perfiles de expresión génica para células en estados patológicos y genes asociados en enfermedades, los determinantes genéticos específicos de afecciones complejas tales como diabetes, obesidad y envejecimiento se desconocen todavía. La genética combinatoria masivamente en paralelo puede producir células con estados patológicos, produciendo nuevos modelos de enfermedad e identificando agentes terapéuticos nuevos.

Una ventaja considerable de esta tecnología con respecto a enfoques previos es la capacidad para readaptar una colección construida para la investigación de prácticamente cualquier fenotipo. Por ejemplo, puede usarse una colección lentivírica combinatoria agrupada de todos los marcos de lectura abiertos (ORF) conocidos en el genoma humano para cribar fenotipos que oscilan entre diferenciación de células madre hasta inhibición de metástasis de cáncer. Esto reduce drásticamente el esfuerzo marginal necesario para investigar cuestiones adicionales.

La invención abarca cualquier tipo de célula en la que pueda introducirse ADN, incluyendo células procariotas y eucariotas. En algunas realizaciones la célula es una célula bacteriana, tal como *Escherichia* spp., *Streptomyces* spp., *Zyomonas* spp., *Acetobacter* spp., *Citrobacter* spp., *Synechocystis* spp., *Rhizobium* spp., *Clostridium* spp., *Corynebacterium* spp., *Streptococcus* spp., *Xanthomonas* spp., *Lactobacillus* spp., *Lactococcus* spp., *Bacillus* spp., *Alcaligenes* spp., *Pseudomonas* spp., *Aeromonas* spp., *Azotobacter* spp., *Comamonas* spp., *Mycobacterium* spp., *Rhodococcus* spp., *Gluconobacter* spp., *Ralstonia* spp., *Acidithiobacillus* spp., *Microtholus* spp., *Geobacter* spp., *Geobacillus* spp., *Arthrobacter* spp., *Flavobacterium* spp., *Serratia* spp., *Saccharopolyspora* spp., *Thermus* spp., *Stenotrophomonas* spp., *Chromobacterium* spp., *Sinorhizobium* spp., *Saccharopolyspora* spp., *Agrobacterium* spp. y *Pantoea* spp. La célula bacteriana puede ser una célula Gram-negativa tal como una célula de *Escherichia coli* (*E. coli*), o una célula grampositiva tal como una especie de *Bacillus*.

En otras realizaciones, la célula es una célula fúngica tal como una célula de levadura, por ejemplo, *Saccharomyces* spp., *Schizosaccharomyces* spp., *Pichia* spp., *Paffia* spp., *Kluyveromyces* spp., *Candida* spp., *Talaromyces* spp., *Brettanomyces* spp., *Pachysolen* spp., *Debaryomyces* spp., *Yarrowia* spp. y cepas de levaduras poliploides industriales. Preferiblemente, la cepa de levadura es una cepa de *S. cerevisiae*. Otros ejemplos de hongos incluyen *Aspergillus* spp., *Penicillium* spp., *Fusarium* spp., *Rhizopus* spp., *Acremonium* spp., *Neurospora* spp., *Sordaria* spp., *Magnaporthe* spp., *Allomyces* spp., *Ustilago* spp., *Botrytis* spp. y *Trichoderma* spp.

En otras realizaciones, la célula es una célula de algas, una célula vegetal, una célula de insecto, una célula de roedor o una célula de mamífero, incluyendo una célula humana (por ejemplo, una célula de riñón embrionario humano (por ejemplo, célula HEK293T), un fibroblasto dérmico humano).

En algunas realizaciones, uno o más de los genes asociados con la invención se expresan en un vector de expresión recombinante. Tal como se usa en el presente documento, un "vector" puede ser cualquiera de varios ácidos nucleicos en los que puede insertarse una secuencia o secuencias deseadas mediante restricción y ligamiento o mediante recombinación para su transporte entre diferentes entornos genéticos o para su expresión en una célula huésped. Los vectores se componen normalmente de ADN, aunque también están disponibles vectores de ARN. Los vectores incluyen, pero no se limitan a: plásmidos, fósidos, fagémidos, genomas de virus y cromosomas artificiales.

Un vector de clonación es uno que es capaz de replicarse de manera autónoma o de integrarse en el genoma en una célula huésped, y que se caracteriza además por uno o más sitios de restricción de endonucleasas en los que el vector puede cortarse de un modo determinable y en los que puede ligarse una secuencia de ADN deseada o sitios de recombinación en los que un inserto con extremos compatibles puede integrarse de manera que el nuevo vector recombinante conserva su capacidad para replicarse en la célula huésped. En el caso de plásmidos, la replicación de la secuencia deseada puede producirse muchas veces a medida que aumenta el número de copias del plásmido dentro de la célula huésped tal como una bacteria huésped o tan sólo una sola vez por huésped antes de que el huésped se reproduzca por mitosis. En el caso de un fago, la replicación puede producirse de manera activa durante una fase lítica o de manera pasiva durante una fase lisogénica.

Un vector de expresión es uno en el que puede insertarse una secuencia de ADN deseada mediante restricción y ligamiento o recombinación de manera que está operativamente unida a secuencias reguladoras y puede expresarse como un transcrito de ARN. Los vectores pueden contener además una o más secuencias marcadoras adecuadas para su uso en la identificación de células que se han transformado o transfectado o no con el vector. Los marcadores incluyen, por ejemplo, genes que codifican proteínas que aumentan o disminuyen o bien la resistencia o bien la sensibilidad a antibióticos u otros compuestos, genes que codifican enzimas cuyas actividades son detectables mediante ensayos convencionales conocidos en la técnica (por ejemplo, β -galactosidasa, luciferasa o fosfatasa alcalina), y genes que afectan visiblemente al fenotipo de placas, colonias, huéspedes o células

transformadas o transfectadas (por ejemplo, proteína fluorescente verde). Vectores preferidos son los capaces de replicación autónoma y expresión de los productos génicos estructurales presentes en los segmentos de ADN a los que están operativamente unidos.

5 Tal como se usa en el presente documento, una secuencia codificante y secuencias reguladoras se dice que están “operativamente” unidas cuando están covalentemente unidas de un modo tal de colocar la expresión o transcripción de la secuencia codificante bajo la influencia o el control de las secuencias reguladoras. Si se desea que las secuencias codificantes se traduzcan en una proteína funcional, se dice que dos secuencias de ADN están operativamente unidas si la inducción de un promotor en las secuencias reguladoras en 5' da como resultado la transcripción de la secuencia codificante y si la naturaleza de la unión entre las dos secuencias de ADN (1) no da como resultado la introducción de una mutación de desplazamiento de marco, (2) no interfiere con la capacidad de la región promotora para dirigir la transcripción de las secuencias codificantes o (3) no interfiere con la capacidad del transcrito de ARN correspondiente para traducirse en una proteína. Por tanto, una región promotora estaría operativamente unida a una secuencia codificante si la región promotora fuese capaz de efectuar la transcripción de esa secuencia de ADN de manera que el transcrito resultante pueda traducirse en la proteína o el polipéptido deseado.

20 Cuando la molécula de ácido nucleico se expresa en una célula, puede usarse una variedad de secuencias de control de la transcripción (por ejemplo, secuencias de promotor/potenciador) para dirigir su expresión. El promotor puede ser un promotor nativo, es decir, el promotor del gen en su contexto endógeno, que proporciona regulación normal de expresión del gen. En algunas realizaciones el promotor puede ser constitutivo, es decir, el promotor no está regulado, permitiendo la transcripción continua de su gen asociado. También puede usarse una variedad de promotores condicionales, tales como promotores controlados por la presencia o ausencia de una molécula. En algunas realizaciones, el promotor es un promotor de ubiquitina C humana (Ubcp). En algunas realizaciones, el promotor es un promotor de citomegalovirus humano (CMVp).

30 La naturaleza precisa de las secuencias reguladoras necesarias para la expresión génica puede variar entre especies o tipos de célula, pero en general incluirán, según se necesite, secuencias no transcritas en 5' y no traducidas en 5' implicadas en el inicio de la transcripción y traducción respectivamente, tal como una caja TATA, secuencia de ocupación de extremos, secuencia CAAT, y similares. En particular, tales secuencias reguladoras no transcritas en 5' incluirán una región promotora que incluye una secuencia de promotor para el control transcripcional del gen operativamente unido. Las secuencias reguladoras también pueden incluir secuencias de potenciador o secuencias activadoras en el sentido de 5' según se desee. Los vectores de la invención pueden incluir opcionalmente secuencias señal o líder en 5'. La elección y el diseño de un vector apropiado están dentro de la capacidad y el criterio de un experto habitual en la técnica.

40 Los vectores de expresión que contienen todos los elementos necesarios para la expresión están disponibles comercialmente y los conocen los expertos en la técnica. Véanse, por ejemplo, Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*, cuarta edición, Cold Spring Harbor Laboratory Press, 2012. Las células se modifican genéticamente mediante la introducción en las células de ADN heterólogo (ARN). Ese ADN heterólogo (ARN) se coloca bajo el control operable de elementos transcripcionales para permitir la expresión del ADN heterólogo en la célula huésped.

45 Puede introducirse una molécula de ácido nucleico asociada con la invención dentro de una célula o células usando métodos y técnicas que son convencionales en la técnica. Por ejemplo, pueden introducirse moléculas de ácido nucleico mediante protocolos convencionales tales como transformación incluyendo transformación química y electroporación, transducción, bombardeo de partículas, etc. La expresión de la molécula de ácido nucleico también puede lograrse integrando la molécula de ácido nucleico dentro del genoma.

50 En algunas realizaciones uno o más genes asociados con la invención se expresan de manera recombinante en una célula bacteriana. Las células bacterianas según la invención pueden cultivarse en medios de cualquier tipo (ricos o mínimos) y cualquier composición. Tal como entendería un experto habitual en la técnica, una variedad de tipos de medios puede ser compatibles con aspectos de la invención. El medio seleccionado puede complementarse con diversos componentes adicionales. Algunos ejemplos no limitativos de componentes de complemento incluyen glucosa, antibióticos, β -D-1-tiogalactopiranosido de isopropilo (IPTG) para la inducción génica, complemento de oligoelementos de la ATCC y glicolato. De manera similar, otros aspectos del medio y las condiciones de crecimiento de las células de la invención pueden optimizarse a través de experimentación de rutina. Por ejemplo, el pH y la temperatura son ejemplos no limitativos de factores que pueden optimizarse. En algunas realizaciones la concentración y la cantidad de un componente de complemento pueden optimizarse.

60 Varios aspectos de la invención se refieren al uso de genética combinatoria masivamente en paralelo para identificar perturbaciones diversas en fenotipos de resistencia a antibióticos usando cepas víricas o bacterianas resistentes a fármacos (por ejemplo, resistentes a múltiples fármacos) tales como, por ejemplo, cepas de Nueva Delhi metalo-beta-lactamasa 1 (NDM-1) de *E. coli*. En particular, los métodos y constructos en el presente documento pueden usarse para identificar, por ejemplo, combinaciones de factores de transcripción que podrían potenciar los antibióticos existentes para el tratamiento. Por tanto, en algunas realizaciones, se proporcionarían en el presente

documento colecciones de combinaciones por parejas, por tríos y de n elementos (por ejemplo, de 4, 5, 6, 7, 8, 9 ó 10 elementos) de factores de transcripción (TF), que pueden analizarse a través de, por ejemplo, secuenciación de siguiente generación.

5 En algunas realizaciones, las colecciones combinatorias de factores de transcripción pueden usarse para identificar combinaciones (por ejemplo, dos o más, tales como 2, 3, 4, 5, 6, 7, 8, 9 ó 10) de TF que presentan efectos sinérgicos con respecto a los fenotipos de resistencia a antibióticos para un solo antibiótico o a lo largo de una gama de antibióticos (por ejemplo, más de un antibiótico, o dos, tres, cuatro o más antibióticos). En algunas realizaciones, se proporcionan en el presente documento constructos que expresan *mtlR* y *uidR*, o *qseB* y *bolA*. En algunas realizaciones, se proporcionan en el presente documento células (por ejemplo, células bacterianas) que contienen constructos que expresan *mtlR* y *uidR*, o *qseB* y *bolA*.

15 En algunas realizaciones, se proporcionan en el presente documento constructos que expresan *rstA* y *rob*, *rstA* y *mirA*, *rscB* y *mirA*, o *feaR* y *hcaR*. En algunas realizaciones, se proporcionan en el presente documento células (por ejemplo, células bacterianas) que contienen constructos que expresan *rstA* y *rob*, *rstA* y *mirA*, *rscB* y *mirA*, o *feaR* y *hcaR*.

20 En algunas realizaciones, las colecciones combinatorias de factores de transcripción pueden usarse para identificar combinaciones de TF que presentan efectos letales que amplifican la actividad bactericida de los antibióticos. En algunas realizaciones, se proporcionan en el presente documento constructos que expresan *torR* y *metR*, *nhaR* y *meiR*, *allR* y *metJ*, *malL* y *yfeT*, *cadC* y *allR*, *torR*, *metR*, o *torR*. En algunas realizaciones, se proporcionan en el presente documento células (por ejemplo, células bacterianas) que contienen constructos que expresan *torR* y *metR*, *nhaR* y *meiR*, *allR* y *metJ*, *malL* y *yfeT*, *cadC* y *allR*, *torR* o *metR*.

25 Los antibióticos para su uso según la invención incluyen, sin limitación, aminoglucósidos, amikacina, gentamicina, kanamicina, neomicina, netilmicina, tobramicina, paromomicina, espectinomicina, ansamicinas, geldanamicina, herbimicina, rifaximina, estreptomina, carbacefem, loracarbef, carbapenemos, ertapenemo, doripenemo, imipenemo/cilastatina, meropenemo, cefalosporinas (primera generación), cefadroxilo, cefazolina, cefalotina, cefalexina, cefalosporinas (segunda generación), cefaclor, cefamandol, cefoxitina, cefprozilo, cefuroxima, cefalosporinas (tercera generación), cefixima, cefdinir, cefditoreno, cefoperazona, cefotaxima, cefpodoxima, ceftazidima, ceftibuteno, ceftizoxima, ceftriaxona, cefalosporinas (cuarta generación), cefepima, cefalosporinas (quinta generación), ceftarolina fosamilo, ceftobiprol, glucopéptidos, teicoplanina, vancomicina, telavancina, lincosamidas, clindamicina, lincomicina, lipopéptido, daptomicina, macrólidos, azitromicina, claritromicina, diritromicina, eritromicina, roxitromicina, troleandomicina, telitromicina, espiramicina, monobactamas, aztreonam, nitrofuranos, furazolidona, nitrofurantoina, oxazolidinonas, linezolidina, posizolida, radezolida, torezolida, penicilinas, amoxicilina, ampicilina, azlocilina, carbenicilina, cloxacilina, dicloxacilina, flucloxacilina, mezlocilina, metilicilina, nafcilina, oxacilina, bencilpenicilina, fenoximetilpenicilina, piperacilina, bencilpenicilina, temocilina, ticarcilina, combinaciones de penicilina, amoxicilina/clavulanato, ampicilina/sulbactam, piperacilina/tazobactam, ticarcilina/clavulanato, polipéptidos, bacitracina, colistina, polimixina B, quinolonas, ciprofloxacino, enoxacino, gatifloxacino, levofloxacino, lomefloxacino, moxifloxacino, ácido nalidíxico, norfloxacino, ofloxacino, trovafloxacino, grepafloxacino, esparfloxacino, temafloxacino, sulfonamidas, mafenida, sulfacetamida, sulfadiazina, sulfadiazina de plata, sulfadimetoxina, sulfametizol, sulfametoxazol, sulfanilimida (arcaico), sulfasalazina, sulfisoxazol, trimetoprim-sulfametoxazol (cotrimoxazol) (tmp-smx), sulfonamidocrisoidina (arcaico), tetraciclinas, demeclociclina, doxiciclina, minociclina, oxtetraciclina, tetraciclina, clofazimina, dapsona, capreomicina, cicloserina, etambutol, etionamida, isoniazida, pirazinamida, rifampicina (rifampina en EE. UU.), rifabutina, rifapentina, estreptomina, arsfenamida, cloramfenicol, fosfomicina, ácido fusídico, metronidazol, mupirocina, platensimicina, quinupristina/dalfopristina, tiamfenicol, tigeciclina, tinidazol y trimetoprima.

50 Otros patógenos resistentes a antibióticos contemplados en el presente documento incluyen, sin limitación, *Staphylococcus aureus*, *Streptococcus* y *Enterococcus*, *Pseudomonas aeruginosa*, *Clostridium difficile*, *Salmonella*, *Escherichia coli*, *Acinetobacter baumannii* y *Mycobacterium tuberculosis*.

55 Diversos aspectos de la invención se refieren al uso de genética combinatoria masivamente en paralelo para dilucidar los mecanismos patológicos que subyacen a enfermedades humanas complejas. Por ejemplo, los métodos de la invención pueden usarse para investigar la regulación de la esperanza de vida y trastornos neurodegenerativos en modelos de levadura. A través de la exploración de interacciones genéticas de orden alto, la genética combinatoria masivamente en paralelo puede proporcionar visiones de estrategias terapéuticas nuevas y descubrimiento de fármacos en enfermedades relacionadas con la edad. Se han usado de manera extensa modelos de levadura para estudiar varios trastornos neurodegenerativos humanos caracterizados por agregación y plegamiento erróneo de proteínas (resumido en la tabla 1).

Tabla 1. Trastornos neurológicos humanos modelados en levadura

Enfermedad	Proteína
Enfermedad de Alzheimer (AD)	amiloide-β (Aβ), APP

Enfermedad de Parkinson (PD)	alfa-sinucleína (α Syn)
Enfermedad de Huntington (HD)	Huntingtina
Prión	PrP
Esclerosis lateral amiotrófica (ALS)	SOD-1
Ataxia de Friedreich (FRDA)	Frataxina

- Entre los modelos de levaduras disponibles, el aumento de amiloide- β ($A\beta$) y alfa-sinucleína (α Syn), que se cree ampliamente que impulsan la manifestación de la enfermedad de Alzheimer (AD) y la enfermedad de Parkinson (PD), respectivamente, inducen profundamente muerte celular de una manera dependiente de la concentración. Lo más importante, varios factores genéticos no sólo suprimen la toxicidad de $A\beta$ o α Syn en levaduras, sino que también rescatan rasgos distintivos patológicos clave en un organismo modelo superior de AD y PD. Por tanto, estos dos modelos pueden usarse para explorar factores genéticos combinatorios que pueden ser beneficiosos para la terapia clínica.
- En algunas realizaciones de la invención, se generan cepas de levadura para cribado genético integrando múltiples copias en tándem de un gen de interés tal como, por ejemplo, alfa-sinucleína, o amiloide- β , con un promotor inducible (por ejemplo, promotor inducible por galactosa (Gal)). En algunas realizaciones, las cepas de levadura se modifican por ingeniería genética para contener el transactivador de tetraciclina inverso (rtTA), que es parte del sistema de expresión Tet-ON. Esta característica permite la expresión de una colección combinatoria generada usando la cepa de levadura modificada por ingeniería genética para que se controle mediante un sistema inducible Tet-ON (por ejemplo, la transcripción se activa de manera reversible en presencia del antibiótico tetraciclina o uno de sus derivados (por ejemplo, doxiciclina). En algunas realizaciones, las cepas de levadura modificadas por ingeniería genética se usan en combinación con genética combinatoria masivamente en paralelo para recuperar combinaciones genéticas que contribuyen a la supervivencia celular con tratamiento con Gal y doxiciclina (Dox).
- Aspectos de la invención también proporcionan colecciones de sobreexpresión (por ejemplo, colecciones combinatorias) de reguladores transcripcionales tales como, por ejemplo, factores de transcripción y enzimas de modificación de histonas. En algunas realizaciones, las colecciones pueden contener también genes de levaduras implicados en rutas de ubiquitina/proteasoma, autofagia y chaperona. Cada gen en una colección puede estar flanqueado por sitios de corte de enzimas de restricción universales para, por ejemplo, aumentar a escala de manera sistemática el ensamblaje de la colección y realizar secuenciación de alto rendimiento para combinaciones genéticas usando genética combinatoria masivamente en paralelo.
- En algunas realizaciones, un constructo de expresión de la invención contiene un código de barras (BC) único y cuatro sitios de restricción situados tal como se muestra en la figura 29. Los vectores con código de barras se agrupan y se digieren con enzimas (por ejemplo, AvrII y NotI). Pueden generarse insertos mediante digestión con diferentes enzimas (por ejemplo, SpeI y PspOMI). En algunas realizaciones, puede usarse una reacción de ligamiento “en un solo recipiente”, o individual, para producir una colección combinatoria por parejas.
- En otros aspectos de la invención, puede usarse genética combinatoria masivamente en paralelo para generar una colección combinatoria de microARN. Por tanto, en algunas realizaciones, un elemento de ADN de la invención puede codificar un microARN (miR). En el presente documento, un microARN se refiere a un ARN no codificante que regula la expresión génica a través de apareamiento de bases con secuencias canónicas presentes en la región no traducida en 3' (UTR en 3') de ARN mensajero (ARNm) diana (Berezikov, E. *et al.* Nat Rev Genet 12, 846-860 (2011)). Los miR se transcriben en primer lugar como precursores de miR que se pliegan sobre sí mismos para formar estructuras en horquilla y se procesan mediante los complejos Drosha y Dicer/RISC para generar su forma madura. Los miR maduros humanos presentan normalmente complementariedad parcial con sus dianas de ARNm. Las secuencias de seis a ocho nucleótidos de longitud en la región en 5' de miR se cree que son un determinante importante de especificidad de diana. Como tal, un solo miR puede tener múltiples dianas de ARNm, mientras que múltiples miR pueden seleccionar como diana un solo ARNm. El ARNm seleccionado como diana se degradará o se impedirá que se traduzca. Mediante los mecanismos anteriores, los miR regulan la expresión de miles de genes y están implicados en la mayoría de los procesos biológicos (Ambros, V. Nature 431:350-355 (2004); Gangarajiu, V.K. *et al.* Nat Rev Mol Cell Bio 10:116-125 (2009); Inui, M. *et al.* Nat Rev Mol Cell Bio 11:222-263 (2010)). Además, se ha notificado la expresión aberrante de miR en numerosos estados patológicos (Esteller M., *et al.*, Nat Rev Genet 12:861-874 (2011); Eacker, S.M. *et al.* Nat Rev Neurosci 10:837-841 (2009); Kong, Y.W. *et al.* Lancet Oncol 13:e249-258 (2012). La restauración de la expresión de diversos miR tiene efectos beneficiosos en determinados estados patológicos, y pueden desarrollarse como productos terapéuticos (Esteller M., *et al.* (2011); Kong, Y.W. *et al.* (2012)).
- En otros aspectos de la invención, puede usarse genética combinatoria masivamente en paralelo para generar vectores (por ejemplo, vectores víricos tales como vectores lentivíricos) que comprenden unidades transcripcionales en tándem que expresan precursor de miR, opcionalmente junto con un gen marcador/indicador (por ejemplo, GFP o RFP). Una unidad transcripcional en el presente documento se refiere a una secuencia de nucleótidos que codifica un miR. Un vector puede comprender dos o más de las mismas unidades transcripcionales (por ejemplo, codificando cada unidad el mismo miR, por ejemplo, miR-124), y/o un vector puede comprender dos o más unidades

transcripcionales diferentes (por ejemplo, una o más unidades que codifican miR-124, una o más unidades que codifican miR-128 y/o una o más unidades que codifican miR-132). Por tanto, una unidad de transcripción en tándem puede contener secuencia(s) de nucleótidos que codifica(n) miR-124, miR-128, miR-132 o cualquier combinación de los mismos. En algunas realizaciones, un vector puede contener dos, tres, cuatro, cinco, seis, siete, ocho, nueve, diez, veinte, treinta, cuarenta, cincuenta (y números enteros entre estos números) o más de las mismas y/o diferentes unidades transcripcionales.

En algunas realizaciones, el precursor de miR de la invención comprende/es la secuencia expuesta como SEQ ID NO: 6, SEQ ID NO: 7 o SEQ ID NO: 8. En algunas realizaciones, el sensor de miR de la invención comprende/es la secuencia expuesta como SEQ ID NO: 9, SEQ ID NO: 10 o SEQ ID NO: 11. En algunas realizaciones, un vector comprende un precursor de miR en tándem que comprende/es la secuencia expuesta como SEQ ID NO: 1.

En algunas realizaciones, la invención proporciona un constructo genético que comprende al menos una secuencia de nucleótidos (por ejemplo, una secuencia de nucleótidos o una pluralidad de secuencias de nucleótidos) que codifica un microARN; un primer elemento de extremos compatibles y un segundo elemento de extremos compatibles que flanquean al elemento de ADN, en el que los elementos de extremos compatibles primero y segundo son capaces de aparearse entre sí; un elemento de código de barras; un tercer elemento de extremos compatibles y un cuarto elemento de extremos compatibles que flanquean al elemento de código de barras, en el que los elementos de extremos compatibles tercero y cuarto son capaces de aparearse entre sí pero no son capaces de aparearse a los elementos de extremos compatibles primero o segundo; y un sitio de separación ubicado entre el cuarto elemento de extremos compatibles y el primer elemento de extremos compatibles, en el que el elemento de ADN, primer elemento de extremos compatibles y segundo elemento de extremos compatibles están en un lado del sitio de separación, y el elemento de código de barras, tercer elemento de extremos compatibles y cuarto elemento de extremos compatibles están en el otro lado del sitio de separación.

También se proporcionan en el presente documento ensayos y constructos para someter a prueba la eficiencia y/o eficacia de vectores que comprenden unidades transcripcionales en tándem que expresan precursor de miR. Por ejemplo, en algunas realizaciones, un vector (por ejemplo, vector lentivírico, vector adenovírico, vector vírico adenoasociado, vector retrovírico) puede comprender unidades transcripcionales en tándem que expresan precursor de miR, opcionalmente junto con un primer gen marcador/indicador (por ejemplo, GFP o RFP) y una secuencia de sensor de miR en la UTR en 3' de un segundo gen marcador/indicador diferente del primero. Una secuencia de sensor de miR en el presente documento se refiere a una secuencia de nucleótidos que es complementaria a un miR (es decir, una secuencia diana complementaria a miR). Una secuencia de sensor de miR contiene al menos una secuencia diana complementaria, y en algunos casos, puede contener repeticiones de la secuencia diana complementaria. Por ejemplo, en algunas realizaciones, una secuencia de sensor de miR puede contener dos, tres, cuatro, cinco, seis, siete, ocho, nueve, diez o más secuencias repetitivas complementarias a la secuencia de nucleótidos de un miR.

Esta invención no está limitada en su aplicación a los detalles de construcción y la disposición de componentes expuestos en la siguiente descripción o ilustrados en los dibujos. La invención es capaz de otras realizaciones y de ponerse en práctica o de llevarse a cabo de diversos modos. Además, la fraseología y terminología usadas en el presente documento es con el propósito de descripción y no debe considerarse como limitativa. El uso de "que incluye", "que comprende" o "que tiene", "que contiene", "que implica", y variaciones de los mismos en el presente documento, pretende abarcar los puntos enumerados después de eso y equivalentes de los mismos así como puntos adicionales.

La presente invención se ilustra adicionalmente mediante los siguientes ejemplos, que de ningún modo deben interpretarse como limitativos adicionalmente.

Ejemplos

Ejemplo 1: Desarrollo de constructos genéticos con código de barras

Se desarrollaron constructos genéticos con código de barras y se manipularon para generar combinaciones de elementos con código de barras. En la figura 1A se muestra un diagrama esquemático general de un ejemplo no limitativo de un constructo genético. Un elemento de ADN (por ejemplo, ADN-1) se asocia con un código de barras único (por ejemplo, BC-1). Flanqueando al elemento de ADN hay extremos compatibles, en este caso, G2A y G2B, que pueden aparearse entre sí, como lo pueden hacer G1A y G1B, mostrado en la figura 1A flanqueando a BC-1. Sin embargo, G1A y G2B no son compatibles. Entre medias de G1B y G2A hay un sitio de separación que permite la linealización del constructo, tal como un sitio de enzima de restricción o cualquier otro método de escisión de ADN. La imagen especular de esta disposición también es funcional, es decir, que tiene un código de barras y extremos 3' compatibles con el elemento de ADN.

La figura 1B representa un acontecimiento de combinación que se produce entre un vector único y un inserto único. El vector se escinde en su sitio de separación. El inserto puede prepararse a partir de su vector, por ejemplo a través de PCR o digestión de restricción. Los extremos terminales del inserto se aparean a sus parejas compatibles en el

vector linealizado y se ligan.

La figura 1C representa el constructo tras la combinación, con el que pueden transformarse células a través de métodos conocidos por los expertos habituales en la técnica. Los códigos de barras BC-1 y BC-2 están ahora en estrecha proximidad y pueden leerse fácilmente mediante secuenciación, revelando la identidad de los elementos de ADN. Además, el sitio de separación se conserva, permitiendo la construcción fácil de combinaciones de orden alto.

En una realización no limitativa de un esquema de generación de muestra, se prepara una colección de constructos como en la figura 1A en plásmidos para cada elemento de ADN único a través de robótica de alto rendimiento. Esta colección de plásmidos se agrupa (generando la colección de vectores), y se realiza PCR en la agrupación (generando la colección de insertos). En una reacción de un solo recipiente, todos los vectores se emparejan con todos los insertos, generando el conjunto combinatorio completo de combinaciones por parejas. Esta colección por parejas se hace reaccionar luego adicionalmente con la misma colección de insertos para generar una colección por tríos, una colección por cuartetos, etc., que surgen de una colección de vectores individuales.

La proximidad de los códigos de barras únicos en la colección combinatoria permite una identificación rápida de los elementos de ADN presentes dentro de un solo constructo combinatorio a través de secuenciación, micromatrices, u otros métodos de determinación de la secuencia de ADN. En un ejemplo, una agrupación de células que contienen la colección combinatoria se somete a selección para un fenotipo específico (tal como resistencia a quimioterapia); el aislamiento y la secuenciación posteriores de las células supervivientes dilucida qué combinación de elementos de ADN produjo el fenotipo deseado. En otro ejemplo, las células agrupadas se someten a secuenciación de siguiente generación de alto rendimiento, produciendo la prevalencia de combinaciones de ADN específicas dentro de la agrupación; estos datos pueden usarse entonces para generar una red de interacciones de proteínas.

Ejemplo 2: Concepción de ensamblaje de Gibson

En algunas realizaciones no limitativas de la invención, se usa el método de ensamblaje de Gibson para generar elementos de extremos compatibles. La reacción de Gibson emplea exonucleasa de T5 para digerir la hebra en 5', permitiendo que se apareen secuencias complementarias, que se rellenen huecos con ADN polimerasa y que se ligan constructos mediante Taq ligasa (Gibson *et al.* (2009) Nature Methods 6:343-5). Tal como se representa en la figura 1, G1A y G1B son secuencias idénticas; G2A y G2B son asimismo idénticas. El sitio de separación es un sitio de reconocimiento de enzimas de restricción.

En algunas realizaciones, se usa un solo sitio de restricción para generar el vector linealizado para la reacción, evitando así la interferencia con elementos de ADN. Como tal, esta metodología es especialmente útil para consultar combinaciones de elementos de todo el genoma tales como ORF y ADN no codificante.

Ejemplo 3: Demostración de genética combinatoria masivamente en paralelo en *E. coli*

Se aplicaron satisfactoriamente métodos para genética combinatoria masivamente en paralelo en *E. coli* a dos proteínas fluorescentes (GFP y mCherry). El elemento de ADN total para cada una consistió en un promotor inducible, la proteína ORF y un terminador. A estos elementos se les añadió un código de barras como BC1-GFP y BC2-mCherry, respectivamente.

En la primera demostración, se insertó BC1-GFP en sí mismo dos veces en sucesión para generar BC1-BC1-GFP-GFP y BC1-BC1-BC1-GFP-GFP-GFP. La figura 2A muestra la electroforesis en gel del propio plásmido no cortado, añadiendo cada integración sucesiva de una unidad de GFP con código de barras otro ~1 kb al tamaño del plásmido. El plásmido también se sometió a digestión por restricción en puntos en el plásmido separando la región combinatoria cambiante de la región estática en el plásmido. En la figura 2B, se dibujan recuadros que rodean el inserto digerido, mostrando un aumento de ~1 kb por inserción sucesiva, mientras que la región estática sin marcar sigue teniendo el mismo tamaño.

A continuación, se creó un conjunto combinatorio usando constructos tanto para GFP como para mCherry. La microscopía de fluorescencia mostró la combinación satisfactoria de elementos de ADN y la expresión funcional de proteínas (figura 3). La figura 3 representa la combinación y la expresión funcional de constructos de mCherry y GFP. Las imágenes fluorescentes se representan usando filtros para mCherry (figura 3A), GFP (figura 3B) y superposición (figura 3C). También se identificaron clones que muestran sólo expresión de GFP o sólo expresión de mCherry.

A través de secuenciación, también se confirmaron los espacios combinatorios completos de cuatro combinaciones con sitios de separación y códigos de barras intactos, indicando la robustez del método (figura 4). Por tanto, esto demuestra la capacidad de los métodos desarrollados en el presente documento para i) generar conjuntos combinatorios de elementos de ADN con códigos de barras únicos, ii) permitir la identificación de combinaciones a través de la secuenciación de regiones con código de barras, y iii) generar combinaciones de orden superior usando el mismo constructo con código de barras para cada etapa de combinación.

Ejemplo 4: Concepción del sitio de restricción

En algunas realizaciones no limitativas de la tecnología, puede aplicarse la metodología de sitios de restricción para generar elementos de extremos compatibles (figura 5). De manera similar a la pieza estandarizada BioBrick®, dos enzimas de restricción únicas generan proyecciones compatibles que, cuando se ligan entre sí, forman una cicatriz que no se reconoce por ninguna enzima de restricción y que no puede escindirse adicionalmente (Anderson *et al.* (2010) *Journal of Biological Engineering* 4:1). La cicatriz más corta asociada con este enfoque, en relación con el método de ensamblaje de Gibson, permite que se lea un número mayor de elementos de código de barras dentro de una lectura de secuenciación de siguiente generación convencional, permitiendo así combinaciones de orden superior. Aunque el uso de enzimas de restricción adicionales plantea la posibilidad de interferencia con los elementos de ADN insertados, las enzimas de restricción específicas que van a usarse pueden modificarse para cualquier conjunto de elementos de ADN, de manera que las enzimas de restricción necesarias para este método interfieren con el menor número de elementos. Además, puede realizarse mutagénesis en los elementos de ADN para eliminar sitios de restricción de sus secuencias.

Usando la estrategia de ensamblaje de sitios de restricción, se sometió a prueba la facilidad de generación de combinaciones de orden superior. Usando una colección de cinco elementos de ADN con código de barras, se repitió la secuencia de combinación tres veces para generar combinaciones por parejas, por tríos y por cuartetos. La secuenciación de colonias individuales mostró la retención de códigos de barras y cicatrices de un modo fácilmente legible (figura 6A). Entonces se digirieron plásmidos aislados de colonias para separar la región combinatoria variable, que consiste en los códigos de barras y sus elementos de ADN respectivos, del resto constante del vector (figura 6B).

Ejemplo 5: Generación de colecciones

A continuación, para someter a prueba la expresión génica, se construyó una colección de vectores con código de barras de GFP y mCherry bajo el control de promotores pLtetO inducibles por aTc con un terminador de la fusión T1+T7 en el sentido de 3' de cada ORF. Cada vector contenía el origen ColE1 de alto número de copias, tetR bajo el control de un promotor constitutivo, y cat que confiere resistencia a cloramfenicol.

Se generó una colección de insertos a través de PCR, creando un conjunto combinatorio por parejas como en la figura 5 con el que entonces se transformaron células DH5a. Se aislaron veinticuatro colonias y se hicieron crecer en concentraciones crecientes de aTc junto con controles positivos y negativos (figura 7). Significativamente, los datos muestran varias características: i) una distribución uniforme de las tres posibles combinaciones que no logran significación a través de pruebas de chi-cuadrado (24 muestras en total; 9 de GFP-GFP; 5 de mCherry-mCherry; 10 de GFP y mCherry); ii) expresión casi independiente de cada combinación diferenciada de promotor-ORF, tal como se determina por la comparación entre combinaciones de GFP-mCherry tales como N95-2 y combinaciones de mCherry-GFP tales como N95-3; y iii) un alto grado de control de inducción y represión, dando razones de activación/desactivación de 150-300x.

Ejemplo 6: Investigación de resistencia antibacteriana en *E. coli*

Se investigaron la resistencia antibacteriana y métodos para derrotar la resistencia en *E. coli*. Añadiendo códigos de barras a 173 factores de transcripción documentados en *E. coli*, se generaron combinaciones por parejas y por tríos, y se determinaron las combinaciones que disminuyen o aumentan más eficazmente la resistencia a antibióticos. El análisis estadístico de las lecturas de secuenciación identificó interacciones genéticas epistáticas no lineales, tales como sinergia y antagonismo. Estos resultados se validaron luego en pruebas de sensibilidad a antibióticos. Para demostrar toda la potencia de esta tecnología, se usaron entonces las combinaciones por parejas más satisfactorias para generar combinaciones por tríos y se evaluó de nuevo la sensibilidad. Pueden investigarse múltiples mecanismos de resistencia usando los métodos proporcionados en el presente documento, incluyendo beta-lactamasas de amplio espectro, mutantes mediante bomba de eflujo y metalo-beta-lactamasas, en respuesta a una gama de antibióticos.

La genética combinatoria masivamente en paralelo usa una estrategia de clonación iterativa comenzando con una colección de elementos de ADN con código de barras (figura 16A). Hidrolizados de restricción de colecciones agrupadas de vectores e insertos, seguido por una reacción de ligación de un solo recipiente, crean una colección de combinaciones por parejas con los códigos de barras en estrecha proximidad entre sí. Como resultado, pueden determinarse las identidades de los componentes en un constructo particular mediante una secuencia corta leyendo los códigos de barras en orden. La secuenciación de alto rendimiento puede censar la distribución de los miembros de la colección dentro de una población agrupada e identificar cambios en la población en condiciones experimentales diferentes. Alternativamente, puede realizarse secuenciación de Sanger en colonias aisladas seleccionadas a través de ensayos de siembra en placa o clasificación de células. De manera notable, los métodos proporcionados en el presente documento son métodos iterativos que pueden usar la colección combinatoria recién producida y la misma agrupación de insertos para generar combinaciones de orden superior en un tiempo lineal logarítmico.

La figura 16A muestra una idea general del método de ensamblaje. Se añaden códigos de barras (BC) a constructos de expresión de factor de transcripción (TF) y se sitúan cuatro sitios de restricción (1A, 1B, 2A, 2B) tal como se muestra. Las parejas, 1A/1B y 2A/2B, son sitios de restricción únicos que generan proyecciones compatibles dentro de la pareja pero son incompatibles con la otra pareja. Los vectores con código de barras se agrupan y se digieren con las enzimas 1B + 2A. Se generan insertos a partir de los vectores mediante PCR y se digieren con 1A + 2B. Una reacción de ligación de un solo recipiente produce una colección combinatoria por parejas, que puede digerirse y ligarse adicionalmente con la misma agrupación de insertos para producir combinaciones de orden superior.

En este estudio, se colocaron 173 factores de transcripción (TF) de *E. coli* y 18 proteínas fluorescentes bajo el control del promotor pLtetO inducible por aTc (figura 17). Se añadieron códigos de barras de manera única a estos constructos de expresión y se combinaron para dar todas las posibles 35.343 combinaciones por parejas (se rechazaron algunos marcos de lectura abiertos (ORF) como vectores o insertos debido a la escisión por enzimas de restricción). Con esta colección combinatoria por parejas se transformó una cepa MG1655 de *E. coli* y se secuenció en la plataforma Illumina HiSeq™, mostrando una distribución de combinaciones uniforme a través de la población y una recuperación de 34.554 combinaciones, o el 98 % de todas las combinaciones posibles (figura 16B). Para evaluar los niveles de expresión del constructo y la influencia del orden sobre los niveles de expresión, se midió la fluorescencia de las cuatro combinaciones por parejas de GFP y mCherry así como de GFP y mCherry individuales (figura 24). Las cuatro combinaciones por parejas de GFP y mCherry, y GFP y mCherry individuales se indujeron con aTc a 250 ng/ml y se evaluaron mediante citometría de flujo. GFP-mCherry y mCherry-GFP mostraron niveles comparables de fluorescencia, indicando una baja influencia de la secuencia sobre el nivel de expresión. GFP-GFP y mCherry-mCherry mostraron niveles de expresión de aproximadamente el doble de GFP y mCherry, respectivamente, en comparación con combinaciones heterogéneas y constructos individuales (figura 16C).

Se descubrieron combinaciones de factores de transcripción (TF) en la colección combinatoria, lo que condujo a fenotipos de viabilidad para la cepa de Nueva Delhi metalo-beta-lactamasa 1 (NDM-1) de *E. coli*. La enzima NDM-1 recién descubierta hidroliza un amplio espectro de antibióticos beta-lactámicos, incluyendo las carbapenemas (Mochon, A. B. *et al.* (2011) *Journal of clinical microbiology* 49:1667-1670). Esta resistencia a antibióticos usados frecuentemente en tratamiento produce infecciones resistentes que pueden tratarse sólo con un puñado de fármacos de último recurso. Este estudio, por tanto, buscó identificar principalmente combinaciones de factores de transcripción que pudieran potenciar antibióticos existentes para el tratamiento.

Con la colección combinatoria se transformó la cepa de NDM-1 de *E. coli* que se produjo mediante la conjugación de una cepa MG1655 de *E. coli* con un aislado clínico de NDM-1 de *Klebsiella pneumoniae*. Se verificó la resistencia de amplio espectro de la cepa de NDM-1 de *E. coli* a través de una gama de antibióticos, y se encontró que su perfil de resistencia se asemeja al del aislado original de *Klebsiella pneumoniae* (tabla 2). Para identificar las combinaciones que potenciaban la destrucción con antibióticos, se sometió la cepa de NDM-1 transformada a condiciones con y sin inducción de aTc y tratamiento con ceftriaxona (figura 18A). La ceftriaxona es una cefalosporina de tercera generación con actividad de amplio espectro y se usa clínicamente para tratar la neumonía, la meningitis bacteriana y la gonorrea. Se recogieron poblaciones en cada condición en fases de crecimiento especificadas para observar las distribuciones de cambio de la población a lo largo del tiempo. Estas muestras se multiplexaron para la secuenciación de alto rendimiento en la plataforma Illumina HiSeq.

Tabla 2. Perfil de resistencia a antibióticos de NDM-1 de *E. coli*, MIC de NDM-1 (µg/ml)

Antibiótico	Sometido a prueba	Notificado previamente ^b
Amoxicilina	>512	
Ceftriaxona	>128	>32
Colistina	<0,125	0,25
Gentamicina	>256	>10
Imipenem	>32	8
Piperacilina-tazobactam	>256 piperacilina, 32 tazobactam	>128

Para determinar la influencia de cada combinación sobre la viabilidad de NDM-1, se comparó la abundancia de cada combinación en una condición experimental particular con su abundancia en otros experimentos. Por ejemplo, las combinaciones que se sobrerrepresentaron especialmente en presencia tanto de aTc como de ceftriaxona, en comparación con aTc solo y ceftriaxona sola, sugirieron que la combinación confería una ventaja de crecimiento relativa a la célula. Asimismo, las combinaciones que estaban infrarrepresentadas tanto en aTc como en ceftriaxona confirieron sinergia bactericida con ceftriaxona. Se seleccionaron las combinaciones que condujeron a fenotipos de destrucción sinérgica, crecimiento neutro y crecimiento ventajoso. La agrupación jerárquica en las combinaciones y sus puntuaciones en cada experimento mostraron la agrupación por clase de fenotipo (figura 18B). Para evaluar la influencia del orden génico en el constructo de expresión, también se realizó agrupación jerárquica en estas combinaciones y sus configuraciones recíprocas. Las combinaciones del mismo fenotipo se agruparon entre sí independientemente del orden (figura 19).

Se clonó cada pareja de genes y se validó su actividad en ensayos de sensibilidad a antibióticos. Las combinaciones

identificadas como sinérgicas mostraron una amplificación significativa de destrucción de tres a cuatro órdenes de magnitud en comparación con el antibiótico solo o aTc solo (figura 20A), y una amplificación significativamente mayor que los controles (figura 20B). En contraposición, las combinaciones identificadas como ventajosas mostraron una amplificación mínima de destrucción por antibiótico solo (figura 20C). Aunque las combinaciones ventajosas no confirieron supervivencia adicional más allá del antibiótico solo, la expresión inocua de las parejas de genes condujo a altas abundancias en los experimentos de colección en comparación con combinaciones sinérgicas y de control.

Al permitir el cribado de alto rendimiento de colecciones combinatorias, la genética combinatoria masivamente en paralelo permite el descubrimiento de sinergia inesperada entre proteínas aparentemente no relacionadas. *mtlR* y *uidR* en combinación sinérgica regulan el metabolismo de manitol y β -glucósido, respectivamente. No se ha notificado que cada gen individualmente medie la resistencia a antibióticos y mostró un efecto mínimo en ensayos de sensibilidad a antibióticos (figura 21A). De manera similar, *qseB* y *bolA* están implicados en rutas asociadas con virulencia y respuesta al estrés, respectivamente, y mostraron una actividad significativamente menor como genes individuales en comparación con combinaciones por parejas (figura 21B).

Para estudiar si la genética combinatoria masivamente en paralelo puede identificar combinaciones a través de una gama de antibióticos, se expuso la colección combinatoria a los antibióticos beta-lactámicos imipenem y piperacilina-tazobactam, y al antibiótico aminoglucósido, gentamicina. Junto con ceftriaxona, cada antibiótico se usa ampliamente en supuestos clínicos y se degrada por la cepa de NDM-1. Resulta interesante que los antibióticos beta-lactámicos se agruparan entre sí por separado de gentamicina, posiblemente debido a sus mecanismos de acción distintos (figuras 22, 25). La genética combinatoria masivamente en paralelo es capaz por tanto de distinguir matices de tratamientos experimentales.

Para determinar si la genética combinatoria masivamente en paralelo puede identificar combinaciones letales amplificando la actividad bactericida de los antibióticos, se seleccionaron combinaciones con abundancia reducida en presencia de aTc y aTc más ceftriaxona. Se descubrieron combinaciones que mostraron una destrucción mayor de más de 5 órdenes de magnitud de la población bacteriana en comparación con ceftriaxona sola (figura 23A). Debido a que muchos genes son tóxicos, se construyeron combinaciones aleatorias de genes tóxicos para evaluar si la genética combinatoria masivamente en paralelo puede identificar combinaciones especialmente letales. Las combinaciones identificadas mostraron un promedio de destrucción mayor de un orden de magnitud en el punto de menor densidad celular (figura 23B). Además, las combinaciones mostraron destrucción sostenida más prolongada a lo largo de un periodo de 10 horas.

Este estudio también evaluó si el gran efecto de destrucción resultaba de la sinergia entre sus dos genes constituyentes. Aunque *torR* y *metR* son tóxicos cuando se expresan individualmente, la suma de sus actividades no recapituló completamente la actividad de la combinación, lo que sugiere una sinergia entre sus efectos (figura 23C). La genética combinatoria masivamente en paralelo puede aislar por tanto perturbaciones combinatorias lo que conduce a fuertes fenotipos que no son evidentes previamente a partir de los constituyentes individuales de la combinación.

Las combinaciones de genes tóxicos podrían ser adyuvantes eficaces para el tratamiento con antibióticos si se administra, por ejemplo, mediante fagos. Se empaquetaron constructos de combinación en fagémidos y se infectó una cepa EMG2 de *E. coli* conjugada con un aislado clínico de NDM-1 de *Klebsiella pneumoniae*. La inducción de genes conjuntamente con ceftriaxona condujo a una mayor destrucción de uno a dos órdenes de magnitud en comparación con ceftriaxona solo, mientras que un fagémido de control no mostró mejoría (figura 23D).

Para demostrar la fácil capacidad de aumento a escala del método de genética combinatoria masivamente en paralelo a combinaciones de orden superior, se construyeron colecciones de combinación por tríos a partir de la colección de TF por parejas. Se produjo una colección de complejidad menor combinando 189 x 187 x 18 genes (figura 26). De las 636.174 combinaciones posibles, se recuperaron 516.798 combinaciones únicas (el 81 % de todas las combinaciones posibles) a partir de aproximadamente 17 millones de lecturas de secuenciación. También se produjo una colección de complejidad mayor combinando 189 x 187 x 187 genes (figura 27). De las 6.609.141 combinaciones de genes posibles, se recuperaron 4.042.316 combinaciones (el 61 %) de aproximadamente 20 millones de lecturas de secuenciación. Probablemente puede lograrse una cobertura mayor de la colección a través de electroporaciones adicionales y dedicando un mayor número de lecturas a la muestra.

A través de la genética combinatoria masivamente en paralelo, se identificaron de manera fiable combinaciones de genes que confieren fuertes fenotipos en resistencia a antibióticos. La secuenciación de alto rendimiento aplicada a la estrategia de ensamblaje de colecciones permitió censar poblaciones combinatorias agrupadas en una escala sin precedentes. Los métodos proporcionados en el presente documento son flexibles y pueden construir combinaciones de orden alto de cualquier elemento de ADN de elección, incluyendo constructos de atenuación y expresión génica, componentes de circuitos sintéticos y secuencias genómicas. La genética combinatoria masivamente en paralelo podría usarse para generar conjuntos de datos de interactoma complejos, para cribar inductores combinatorios de los fenotipos deseados tales como de diferenciación, y para caracterizar circuitos con una escala y una eficacia sin precedentes. La genética combinatoria masivamente en paralelo que usan enzimas de restricción también puede adaptarse al ensamblaje isotérmico de Gibson o recombinasas en situaciones en las que

la colección es incompatible con sitios de restricción. El sondaje de fenotipos complejos y propiedades de red emergentes contribuirá enormemente a la comprensión de los sistemas biológicos y a sus aplicaciones.

Métodos

5 *Cepas bacterianas y condiciones de cultivo.* Las células DH5a de *E. coli* usadas para clonar se obtuvieron de New England Biolabs. Las células ElectroTen-Blue de *E. coli* usadas para la electroporación de colecciones se obtuvieron de Agilent. Las células MG1655 de *E. coli* conjugadas con un aislado clínico de *K. pneumoniae* que produce NDM-1 se usaron para los ensayos de sensibilidad a antibióticos y persistencia. Los cultivos se hicieron crecer en caldo
10 Luria-Bertoni (LB) con estas concentraciones químicas a menos que se indique otra cosa: aTc a 100 ng/μl, carbenicilina (Carb) a 50 μg/ml, cloramfenicol (Cm) a 30 μg/ml, ceftriaxona a 192 μg/ml.

15 *Compuestos químicos.* La ADN ligasa de T4, la exonucleasa de T5 y las enzimas de restricción se obtuvieron de New England Biolabs. Las reacciones en cadena de la polimerasa se realizaron con HiFi™ HotStart de Kapa Biosystems. Los oligonucleótidos se adquirieron de Integrated DNA Technologies. Los antibióticos se adquirieron de Sigma y Enzo Life Sciences. Todos los demás compuestos químicos se adquirieron de Sigma-Aldrich.

20 *Construcción de colecciones.* Para construir la colección de factores de transcripción (TF) individuales con código de barras, se usó el ensamblaje isotérmico de Gibson para ligar un código de barras único de 6 pares de bases con el promotor P_LtetO, el marco de lectura abierto (ORF) de TF, y la estructura principal del vector. Cada código de barras tenía una longitud de 6 pares de bases y estaba separado de cada otro código de barras en una distancia genética de al menos 2 bases. Las secuencias de ORF de factores de transcripción se obtuvieron de la colección de clones(-) ASKA de NBRP-*E. coli* en NIG. Todos los constructos de expresión de TF estaban seguidos por el terminador BBa_B1006 del Registry of Standard Biological Parts. Cada TF con código de barras contenía sitios de restricción en
25 esta configuración: constructo de expresión Spel-código de barras-AvrII-PspOMI-TF-NotI. Los insertos se generaron a partir de cada vector mediante PCR con dos cebadores comunes.

30 Para construir la colección combinatoria por parejas, se midieron todos los vectores e insertos para determinar la concentración en NanoDrop 2000 (Thermo Scientific) y se agruparon en cantidades equimolares para formar una agrupación de vectores y una agrupación de insertos. La agrupación de vectores se digirió con AvrII y PspOMI, y la agrupación de insertos se digirió con Spel y NotI. Ambos hidrolizados se purificaron entonces a través de extracción con fenol/cloroformo y precipitación con etanol. Las agrupaciones de hidrolizados se ligaron con ADN ligasas de T4, se purificaron con fenol/cloroformo y precipitación con etanol, y se sometieron a electroporación en células ElectroTen-Blue según el protocolo del fabricante. Se obtuvo un total de 3.000.000 de transformantes y se hicieron
35 crecer hasta fase semilogarítmica en 100 ml de LB + 30 μg/ml de cloramfenicol. Se aislaron los plásmidos a través de Midi Prep (Qiagen) y se sometieron a electroporación en NDM-1 de *E. coli*.

40 *Tratamientos con antibiótico.* Se realizaron ensayos con antibióticos a 37 °C en una placa de fondo transparente de 96 pocillos en un lector de microplacas VersaMax™ y un agitador (Molecular Devices). Todos los pocillos de cultivo contenían 200 μl. Se diluyeron las reservas congeladas de NDM-1 con colección en LB con o sin aTc 100 ng/μl. Cuando estos cultivos alcanzaron una DO₆₀₀ = 0,6, se diluyeron 1:100 en pocillos con o sin antibiótico y con o sin aTc 100 ng/μl. Las concentraciones de antibiótico usadas fueron: baja concentración de ceftriaxona: 64 μg/ml; alta concentración de ceftriaxona: 256 μg/ml; baja concentración de imipenem: 32 μg/ml; alta concentración de imipenem: 96 μg/ml; baja concentración de piperacilina-tazobactam: piperacilina 64 μg/ml, tazobactam 8 μg/ml; alta
45 concentración de piperacilina-tazobactam: piperacilina 256 μg/ml, tazobactam 32 μg/ml; baja concentración de gentamicina: 32 μg/ml; alta concentración de gentamicina: 256 μg/ml. Estos pocillos se hicieron crecer hasta DO = 0,3 ó 0,9. Todos los ADN de los pocillos se recogieron mediante lisis alcalina y precipitación con etanol. Cada muestra de miniprep se sometió a ensayo para determinar la concentración usando los kits SYBR® Fast qPCR (Kapa Biosystems).

50 *Secuenciación de alto rendimiento.* Se preparó cada muestra para secuenciación por Illumina HiSeq™ añadiendo un código de barras de indexación y secuencias de anclaje Illumina a través de la reacción en cadena de la polimerasa (PCR). Para impedir el sesgo de PCR que distorsionaría la distribución de población, se terminaron las reacciones de PCR durante la fase exponencial. Los productos de PCR se purificaron con perlas AMPure XP (Agencourt®), la concentración se cuantificó a través de qPCR, y se agrupó en cantidades equimolares. Entonces se secuenciaron las muestras multiplexadas usando los cebadores indicados a continuación.

Cebadores para amplificar poblaciones de CombiGEM para secuenciación.

60 Cebador directo:

AATGATACGGCGACCACCGAGATCTACACCGCTGGCAAGTGTAGC (SEQ ID NO: 2)

65 Cebador inverso con código de barras:

CAAGCAGAAGACGGCATAACGAGATNNNNNNGGGAGGGCCCGTTG (SEQ ID NO: 3)

Cebador de secuenciación de Illumina:

5 CCACGAGGATTCGAAAAGGTGAACCGACCCGGTCGATGCACTAGT (SEQ ID NO: 4)

Cebador de indexación de Illumina: CCTAGGAGCAAGTACGAACAACGGGCCCTCCC (SEQ ID NO: 5)

10 *Análisis de población.* Se procesaron las lecturas sin procesar para cada pareja de genes en cada experimento a partir de los datos de secuenciación. Para garantizar una transformación logarítmica válida posterior, se añadió una pseudolectura de 1 para cada combinación de experimento de pareja de genes. Se normalizaron las lecturas para cada combinación para las lecturas totales en cada experimento y para los controles con fluoróforo (consistiendo las combinaciones por parejas sólo en eCFP (proteína fluorescente cian potenciada), GFP (proteína fluorescente verde), y mCherry) en cada experimento. Para corregir el efecto del lote de aTc usando el modelo multiplicativo, las lecturas normalizadas para cada experimento de pareja de genes se dividieron entre la mediana de las lecturas de esa pareja de genes de todos los experimentos con o sin aTc. Finalmente, las lecturas normalizadas y corregidas por lotes se transformaron logarítmicamente, produciendo la abundancia comparable para cada combinación de fármaco y pareja de genes. La diferencia entre las abundancias de una pareja de genes en dos experimentos corresponde a la razón logarítmica de sus lecturas normalizadas y corregidas por lotes. El cálculo de las puntuaciones de interacción se basa en el modelo aditivo entre la abundancia comparable para una pareja de genes en una condición farmacológica particular y la media de abundancias para un conjunto de referencias que consiste en todos los

cribados de fármacos. La fórmula para la puntuación S de interacción sinérgica se define como $S_{ij}^k = \frac{\alpha_{ij}^k - \bar{\alpha}_{ij}}{\sigma_{ij}}$, donde

$$\bar{\alpha}_{ij} = \frac{1}{n_{ij}} \sum_k \alpha_{ij}^k$$

25 i, j es una combinación de los genes i y j , k abarca un conjunto de experimentos, α_{ij}^k es la media de abundancias α_{ij}^k de n_{ij} experimentos, y σ_{ij}^k es la desviación estándar global con una unión mínima que representa la varianza sistemática.

Análisis de agrupamiento. Se realizó agrupamiento jerárquico a través de un subconjunto de pares de genes y experimentos con correlación de Pearson y unión promedio.

30 *Ensayos de sensibilidad a antibióticos.* Se construyeron combinaciones específicas identificadas a través de análisis en la estructura principal del vector, verificadas a través de secuenciación de Sanger, y con ellas se transformaron NDM-1. Se hicieron crecer reservas congeladas durante la noche en LB + Carb + Cm, entonces se diluyeron 1:100 en 2 ml de LB + Cm y se hicieron crecer a 37 °C durante 1 hora. Se añadieron aTc y ceftriaxona cuando fue apropiado, y los cultivos volvieron a colocarse a 37 °C. Para obtener recuentos de unidades formadoras de colonias (ufc) en puntos de tiempo indicados, se recogieron 100 μ l de cultivos relevantes, se lavaron en solución salina tamponada con fosfato (PBS), y se resuspendieron en 100 μ l de PBS. Se realizaron diluciones en serie con PBS, y se sembraron en placa 10 μ l de cada dilución en placas de agar LB. Se incubaron las placas de agar LB a 37 °C durante la noche antes del recuento.

40 *Citometría de flujo.* Se usaron cultivos que contenían niveles mínimos y máximos de fluorescencia para calibrar el isotiocianato de fluoresceína y voltajes de filtro PE-TexasRed® en un dispositivo de toma de muestras de alto rendimiento LSRFortessa de BD para medir los niveles de expresión de GFP y mCherry, respectivamente. GFP se excitó con un láser de 488 nm y mCherry se excitó con un láser de 561 nm. No fue necesaria compensación de voltaje para el isotiocianato de fluoresceína y PE-TexasRed para ningún experimento.

Ejemplo 7A: Demostración de genética combinatoria masivamente en paralelo en células de levadura

Se añaden códigos de barras a cada uno de GFP y mCherry, se colocan bajo el control de un promotor que responde a GAL, y se flanquean por sitios de restricción para la metodología de sitios de restricción. Estos se colocan en un vector lanzadera pRS que contiene un origen pMB1 de *E. coli*, un casete de resistencia a ampicilina, un origen centromérico CEN6 de levadura, y el gen URA3 para selección auxotrófica. Se generan insertos a partir de vectores a través de PCR. Se crea una colección combinatoria agrupada con la que se transforma *E. coli*, a partir de la cual se aísla el plásmido y se introduce en células de levadura. Se hace crecer la población agrupada que contiene todas las combinaciones de indicadores en presencia de galactosa y se analiza a través de citometría de flujo en verde y rojo, mostrando tres poblaciones que representan las tres combinaciones por parejas únicas de GFP y mCherry.

Ejemplo 7B: Demostración de genética combinatoria masivamente en paralelo en células de levadura

60 La genética combinatoria masivamente en paralelo puede usarse para dilucidar los mecanismos patológicos que subyacen a enfermedades complejas en humanos. Por ejemplo, los métodos proporcionados en el presente

documento pueden usarse para investigar la regulación de la vida y los trastornos neurodegenerativos en modelos de levaduras. A través de la exploración de interacciones genéticas de orden alto, pueden usarse los métodos de la invención para proporcionar visiones sobre nuevas estrategias terapéuticas y de descubrimiento de fármacos en enfermedades relacionadas con la edad.

5 Se usaron dos modelos de levaduras, uno para la enfermedad de Alzheimer (producción aumentada de amiloide- β , A β) y el otro para la enfermedad de Parkinson (alfa-sinucleína aumentada, α Syn), para explorar factores genéticos combinatorios que pueden ser beneficiosos para la terapia clínica.

10 Se integró la matriz de expresión repetitiva de A β o α Syn en genoma de levadura (el contexto de W303) y se controló mediante la inducción de galactosa. La figura 28 representa una cepa de levadura generada integrando múltiples copias en tándem de α Syn con un promotor inducible de galactosa (Gal). Con el fin de minimizar el tamaño de la estructura principal del plásmido para la construcción de colecciones, se integró el transactivador de tetraciclina inverso (rtTA), parte del sistema de expresión Tet-ON (Gossen, M. *et al.*, Science 268(51218):1766-9 (1995)), en ambas cepas de levadura. Se sometió a prueba la inducción de Tet-ON usando diferentes proteínas fluorescentes (tal como eGFP y mCherry), y se detectó de manera sistemática la activación genética robusta (aumento de aproximadamente 50 veces) con tratamiento con doxiciclina (Dox). Además, ni el rtTA integrado ni el tratamiento con Dox afecta al crecimiento celular o a la toxicidad amiloide inducible. De manera similar, se integró satisfactoriamente el rtTA en la cepa de tipo natural (el contexto de BY4741) para estudiar la longevidad.

20 *Construcción de colección combinatoria con código de barras y ensayo de combinación por parejas*

25 Se construyó una colección de sobreexpresión de reguladores de transcripción que incluyen 196 factores de transcripción y 29 enzimas de modificación de histonas. La toxicidad amiloide se debe generalmente a la disfunción del control de calidad de las proteínas, por tanto también se recogieron genes de levaduras implicadas en rutas de ubiquitina/proteasoma, autofagos y chaperonas (otros 224 genes). Con 54 genes conocidos como supresores de A β (o α Syn) y moduladores de la vida de levaduras, todas las secuencias génicas se agruparon entre sí y se identificaron dos pares de enzimas de restricción compatibles para el ensamblaje BioBrick[®]. Tal como se muestra en la figura 29, cada gen en la colección y su código de barras único están flanqueados por sitios de corte de enzimas de restricción universales de modo que, por ejemplo, puede ampliarse a escala sistemáticamente el ensamblaje de colecciones y realizarse secuenciación de alto rendimiento para combinaciones genéticas, lo que es característico de la genética combinatoria masivamente en paralelo.

35 Para validar la estrategia de genética combinatoria masivamente en paralelo en levaduras, se estableció un ensayo con tres genes fluorescentes (EGFP, mCherry y EBFP2). Se clonó cada gen en el plásmido de la estructura principal de la colección con un código de barras asignado por el método isotérmico de Gibson. Para construir combinaciones por parejas, se agruparon entre sí plásmidos con código de barras individuales por igual y entonces se digirieron con AvrII y NotI como vectores de ensamblaje (figura 29). Se usaron Spel y PspOMI para crear insertos del mismo agrupamiento de ADN. A través del ligamiento de ADN, se recuperaron las nueve posibles combinaciones por parejas secuenciando códigos de barras de colonias aleatorias.

45 Entonces se eligieron combinaciones por parejas de EGFP y mCherry para examinar la expresión génica en diferente orden en el plásmido. En cada caso, ambos genes fluorescentes muestran niveles de expresión similares y son comparables al constructo de un solo gen (inducción de ~50 veces) (figura 30). Se comparó la genética combinatoria masivamente en paralelo con el método de P2A para la expresión policistrónica (Fang *et al.*, Nature Biotechnology 23:584-590 (2005)) para expresar simultáneamente dos genes fluorescentes en el mismo plásmido. A diferencia del perfil de expresión de los clones de genética combinatoria masivamente en paralelo, el perfil de expresión de los clones de P2A no puede predecirse intercambiando el orden de los genes. Por tanto, la estrategia de genética combinatoria masivamente en paralelo es poderosa y fiable para construir una colección combinatoria de orden alto.

Direccionamiento génico múltiple basado en CRISPR

55 Además de la genética combinatoria masivamente en paralelo, se usó una estrategia de CRISPR (repeticiones palindrómicas cortas agrupadas y regularmente interespaciadas) para programar la expresión génica combinatoria en modelos de levadura. Generalmente, CRISPR lleva a cabo rotura bicatenaria de ADN dirigida en cuatro etapas secuenciales. En primer lugar, se transcriben dos ARN no codificantes, la matriz de pre-crARN y tracrARN, a partir del locus de CRISPR. En segundo lugar, tracrARN hibrida con las regiones de repetición de pre-crARN y media el procesamiento de pre-crARN para dar crARN maduros que contienen secuencias espaciadoras individuales. En tercer lugar, el complejo crARN:tracrARN maduro dirige Cas9 al ADN diana a través de un emparejamiento de bases de Watson-Crick entre el espaciador en el crARN y el protoespaciador en el ADN diana junto al motivo adyacente al protoespaciador (PAM), un requisito adicional para el reconocimiento de diana. Finalmente, Cas9 media la escisión de ADN diana para crear una rotura bicatenaria dentro del protoespaciador.

65 En el presente documento, se usan los complejos sintéticos de endonucleasa Cas9 con ARN de guía (gARN)

diseñados para determinar digestión específica de sitio. Mediante el uso de un mutante de nucleasa Cas9 (dCas9) que conserva la capacidad de unión a ADN, se modificó dCas9 mediante ingeniería genética para que funcionara como activador o represor de la transcripción (dCas9-VP16), que se regula mediante el sistema inducible Tet-ON, y se integra en la cepa α Syn (también el ensayo de longevidad) (figura 31). A la especificidad de unión de secuencia de gARN sólo contribuye su nucleótido 16, lo que sugiere que un solo gARN puede mediar el direccionamiento génico múltiple. En paralelo, también se construyó una colección de gARN combinatoria para ensayos de envejecimiento y relacionados con la edad. Los gARN son diferentes sólo en 20 nucleótidos, por tanto pueden usarse como códigos de barras únicos para recuperar combinaciones genéticas mediante secuenciación directa.

También se integró Cas9 de tipo natural en cepas de modelo de levadura. Con recopilaciones de gARN específicas, se introducen codones de parada tempranos mediante edición de genoma mediada por gARN. Por tanto, se realiza el cribado combinatorio de orden alto de atenuación genética (figura 32).

Ejemplo 8A: Demostración de genética combinatoria masivamente en paralelo en células de mamífero

También se demuestran combinaciones de GFP-mCherry en células de mamífero. Se añaden códigos de barras a cada uno de GFP y mCherry, se colocan bajo el control de un promotor que responde a tetraciclina (TRE) y se flanquean por sitios de restricción para la metodología de sitios de restricción. Estos se colocan en un vector que consiste en un origen pUC y un casete de resistencia a neomicina. Se generan insertos a partir de vectores a través de PCR. Se crea una colección combinatoria agrupada con la que se transforma *E. coli*, a partir de la cual se aísla el plásmido y se introduce en células HeLa produciendo rTA. La población agrupada que contiene todas las combinaciones de indicadores se hace crecer en presencia de doxiciclina y se analiza a través de citometría de flujo en verde y rojo, mostrando tres poblaciones que representan las tres combinaciones por parejas únicas de GFP y mCherry.

Ejemplo 8B: Demostración de genética combinatoria masivamente en paralelo en células de mamífero usando micro ARN

Para expresar precursor de miR (microARN) y medir su actividad en células humanas, se construyó un vector lentivírico que consistía en unidades de transcripción en tándem que expresaban precursor de miR junto con un gen de GFP y una secuencia de sensor de miR (por ejemplo, cuatro repeticiones de las secuencias diana complementarias de miR) en la UTR en 3' de un gen de RFP (figura 33). También se construyeron vectores de expresión de control o bien sin precursor de miR o bien sin secuencia de sensor de miR para comparación (figura 33). Se usaron lentivirus generados para cada vector para infectar células de riñón embrionario humano (HEK293T) y fibroblastos dérmicos humanos (HDF) primarios. Se realizó citometría de flujo para medir poblaciones de células positivas para fluorescencia por GFP y RFP, y se determinó el porcentaje de células RFP⁺ con respecto a células GFP⁺. Los resultados revelaron que la mayoría de las células que expresaban GFP-precursor de miR no presentaban fluorescencia por RFP cuando se incluyó la secuencia de sensor (figuras 34 y 35), lo que indica que el precursor de miR individual puede expresarse con el vector lentivírico en células humanas. Un solo vector diseñado para permitir la expresión de miR combinatoria garantizaría su razón de expresión uniforme en todas las células seleccionadas como diana. A continuación se construyó el vector lentivírico para expresar dos precursores de miR dispuestos en tándem (por ejemplo, miR-128-132) junto con un gen de GFP. El vector lentivírico también puede construirse con elementos de código de barras, tal como se describe en otra parte en el presente documento. Los precursores de miR combinatorios expresados mostraron actividades comparables ya que se expresaban independientemente (figuras 36 y 37). Estos resultados demuestran que el vector lentivírico puede usarse para accionar eficazmente la expresión combinatoria de miR en células humanas. Los vectores lentivíricos usados en este estudio se representan en las figuras 38 y 39.

Método

Construcción de vectores lentivíricos. Se amplificaron por PCR las secuencias de precursor de miR (tabla 3) y sensor de miR (tabla 4) a partir de fragmentos génicos sintetizados y se clonaron en una estructura principal de vector lentivírico.

Tabla 3. Lista de secuencias de precursor de miR usadas.

Precursor de miR	Secuencia
------------------	-----------

miR-124	AGGTGGGAGTACTGCTCAGAGCTACAACCTCTAGGAGTAG GGACTCCAAGCCTAGAGCTCCAAGAGAGGGTGAAGGGCA GGGAGAAAATTATAGTAATAGTTGCAATGAGTCACTTGCT TCTAGATCAAGATCAGAGACTCTGCTCTCCGTGTTACAG CGGACCTTGATTTAATGTCATACAATTAAGGCACGCGGTG AATGCCAAGAGCGGAGCCTACAGCTGCACTTGAAGGACA TCCGAGAGAAGTTAGGAAGGGTGGGGAGAAACAATTCTA GAATGAACCCATCCTGTGCGACAC (SEQ ID NO:6)
miR-128	ATACTGTGAAGTACACTGCATATAAGGAGTGTGGTATAGT ATAAAGAACTTTCTGCAGGTAGTAATTATAGTGAAGATT TTAGGTTTACAAAGCCCTAGCTGTTTTCTGTGTAGCTTTTA TTATTCTTATGACTCTTGACAAGTTTGTAGCTTCACCATAT ACATTTAATATTTTGCAATAATTGGCCTTGTTCCCTGAGCTG TTGGATTTCGGGGCCGTAGCACTGTCTGAGAGGTTTACATT TCTCACAGTGAACCGGTCTCTTTTTTCAGCTGCTTCCTGGCT TCTTTTTACTCAGGTTTCCACTGCTTTTTTTGCTTTTTTTAAT GCTGTATGAAGGTGTTAACATTTGTTTATATTTTTTCATTAA TTGTAATACCTTTAAATCATGCATCATACTCAGAAATAGG GATTAGAATTTAAGTGACATCTTTGGCC (SEQ ID NO:7)
miR-132	CTAGCCCCGCAGACACTAGCGCCACCCCCGCCGCCCGCG GTGCTGACGTCAGCCTGCAAGCCCCGCCCGCGTCTCCA GGGCAACCGTGGCTTTCGATTGTTACTGTGGGAACCGGAG GTAACAGTCTACAGCCATGGTCCGCCCGCAGCACGCCAC GCTCCCCACCACTCCCGAGTTCTGCCAGCCTGGGTTTGGG CAGATACAGAGCAAGAGGAGGCGGGG (SEQ ID NO:8)

Tabla 4. Lista de secuencias de sensor de miR usadas.

Sensores de miRr	Secuencia
miR-124	GGCACAGATAATAACCTGCAAAAAGGCATTCACCGCGTGC CTTAGGCATTCACCGCGTGCCTTAGGCATTCACCGCGTGCC TTAGGCATTCACCGCGTGCCTTAAATGCAGGCGGGCCAGAT ATAC (SEQ ID NO:9)
miR-128	GGCACAGATAATAACCTGCAAAAAGAGACCGGTTCACTGT GAAAAGAGACCGGTTCACTGTGAAAAGAGACCGGTTTAC TGTGAAAAGAGACCGGTTCACTGTGAAATGCAGGCGGGC CAGATATAC (SEQ ID NO:10)
miR-132	GGCACAGATAATAACCTGCAAAAACGACCATGGCTGTAGA CTGTTACGACCATGGCTGTAGACTGTTACGACCATGGCTGT AGACTGTTACGACCATGGCTGTAGACTGTTAAATGCAGGC GGGCCAGATATAC (SEQ ID NO:11)

5 **Ejemplo 9: Genética combinatoria masivamente en paralelo usando recombinasas**

La genética combinatoria masivamente en paralelo también puede lograrse usando recombinasas. Tal como se muestra en la figura 9, la recombinación constituye la base tras técnicas tales como la clonación Gateway®. Brevemente, pueden intercambiarse fragmentos de ADN flaqueados por secuencias de recombinación compatibles usando mezclas de reacción enzimáticas Clonase™. Por ejemplo, en la clonación Gateway®, las secuencias de recombinación *attB* son compatibles con las secuencias *attP* y forman sitios *attL* y *attR* tras la reacción de recombinación. Los sitios *attL* y *attR* a su vez son compatibles y generan sitios *attB* y *attP*. En la tecnología Gateway®, existen cuatro conjuntos ortogonales de secuencias *attB*, *attP*, *attL* y *attR*.

15 En la figura 11 se muestra un ejemplo no limitativo de una tecnología genética combinatoria mediada por

recombinación. Esta realización emplea dos colecciones de vectores, en las que se ubican dos secuencias de recombinación entre el elemento de ADN y su elemento de código de barras. Los insertos pueden generarse mediante PCR y pueden contener un par flanqueante de secuencias de recombinación ortogonales. Las secuencias de recombinación *att* se representan por B, P, L ó R. BC = Código de barras, ADN = elemento de ADN.

Tal como se muestra en la figura 12, para generar combinaciones de manera iterativa, puede usarse un vector de cualquier colección. La reacción iterativa avanza alternando entre las dos colecciones de insertos, comenzando con los insertos generados a partir de la colección de vectores que no se usaron para la etapa de recombinación inicial.

Las secuencias de recombinación específica que van a usarse pueden modificarse siempre que se cumplan unos requisitos mínimos de compatibilidad para impedir la recombinación no deseada. Las longitudes aproximadas de las secuencias *attB*, *attP*, *attL*, y *attR* son 30, 240, 100, y 160 nucleótidos, respectivamente. En algunas realizaciones, la inserción de sitios *attB* y *attL* entre códigos de barras puede optimizar la longitud de secuencia que va a leerse mediante secuenciación u otros métodos.

Ejemplo 10: Genética combinatoria masivamente en paralelo usando un solo sitio de separación enzimática

La genética combinatoria masivamente en paralelo puede realizarse usando un solo sitio de restricción ubicado entre el elemento de ADN y el elemento de código de barras dentro de un constructo genético. La figura 15 demuestra un vector en el que un elemento de código de barras y el elemento de ADN están separados por un sitio de restricción para la enzima de restricción BgIII. También se representa un inserto que contiene un elemento de ADN y un elemento de código de barras. El inserto contiene un sitio de restricción para la enzima de restricción BgIII ubicada entre el elemento de ADN y el elemento de código de barras, y sitios de restricción para la enzima de restricción AlwNI ubicada fuera del elemento de ADN y el elemento de código de barras. Las enzimas de restricción BgIII y AlwNI generan extremos compatibles de manera que cuando se escinde el vector mediante BgIII y se escinde el inserto mediante AlwNI, el vector y el inserto se aparean entre sí, generando un vector que contiene dos elementos de ADN y dos elementos de código de barras separados por un solo sitio de restricción para la enzima BgIII.

La descripción proporciona además los siguientes párrafos numerados:

1. Un constructo genético que comprende:

un elemento de ADN;

un primer elemento de extremos compatibles y un segundo elemento de extremos compatibles que flanquean al elemento de ADN, en el que los elementos de extremos compatibles primero y segundo son capaces de aparearse entre sí;

un elemento de código de barras;

un tercer elemento de extremos compatibles y un cuarto elemento de extremos compatibles que flanquean al elemento de código de barras, en el que los elementos de extremos compatibles tercero y cuarto son capaces de aparearse entre sí pero no son capaces de aparearse a los elementos de extremos compatibles primero o segundo;

y un sitio de separación ubicado entre el cuarto elemento de extremos compatibles y el primer elemento de extremos compatibles, en el que el elemento de ADN, primer elemento de extremos compatibles y segundo elemento de extremos compatibles están en un lado del sitio de separación, y el elemento de código de barras, tercer elemento de extremos compatibles y cuarto elemento de extremos compatibles están en el otro lado del sitio de separación.

2. El constructo genético según el párrafo 1, en el que el elemento de ADN contiene al menos dos secuencias de nucleótidos en tándem, codificando cada secuencia de nucleótidos un gen de interés.

3. El constructo genético según el párrafo 2, en el que el constructo genético contiene al menos dos elementos de código de barras en tándem.

4. El constructo genético según uno cualquiera de los párrafos 1-3, en el que cada secuencia de nucleótidos está unida operativamente a un promotor inducible.

5. El constructo genético según uno cualquiera de los párrafos 2-4, en el que el gen de interés es un factor de transcripción y/o una enzima de modificación de histonas.

6. El constructo genético según uno cualquiera de los párrafos 2-5, en el que:

(a) al menos una de las secuencias de nucleótidos codifica *mtlR* y al menos una de las secuencias de nucleótidos

- codifica *uidR*;
- 5 (b) al menos una de las secuencias de nucleótidos codifica *qseB* y al menos una de las secuencias de nucleótidos codifica *bolA*;
- (c) al menos una de las secuencias de nucleótidos codifica *rstA* y al menos una de las secuencias de nucleótidos codifica *rob*;
- 10 (d) al menos una de las secuencias de nucleótidos codifica *rstA* y al menos una de las secuencias de nucleótidos codifica *mirA*;
- (e) al menos una de las secuencias de nucleótidos codifica *rcsB* y al menos una de las secuencias de nucleótidos codifica *mirA*;
- 15 (f) al menos una de las secuencias de nucleótidos codifica *feaR* y al menos una de las secuencias de nucleótidos codifica *hcaR*;
- (g) al menos una de las secuencias de nucleótidos codifica *feaR* y al menos una de las secuencias de nucleótidos codifica *hcaR*;
- 20 (h) al menos una de las secuencias de nucleótidos codifica *torR* y al menos una de las secuencias de nucleótidos codifica *metR*;
- (i) al menos una de las secuencias de nucleótidos codifica *nhaR* y al menos una de las secuencias de nucleótidos codifica *melR*;
- 25 (j) al menos una de las secuencias de nucleótidos codifica *allR* y al menos una de las secuencias de nucleótidos codifica *metJ*;
- (k) al menos una de las secuencias de nucleótidos codifica *malL* y al menos una de las secuencias de nucleótidos codifica *yfeT*;
- (l) al menos una de las secuencias de nucleótidos codifica *cadC* y al menos una de las secuencias de nucleótidos codifica *allR*;
- 35 (m) al menos una de las secuencias de nucleótidos codifica *torR*; o
- (n) al menos una de las secuencias de nucleótidos codifica *metR*.
- 40 7. El constructo genético según uno cualquiera de los párrafos 2-6, en el que las menos dos secuencias de nucleótidos en tándem potencian la letalidad de los antibióticos en una célula bacteriana.
8. El constructo genético según el párrafo 7, en el que la célula bacteriana es resistente a al menos un antibiótico.
- 45 9. El constructo genético según el párrafo 7 u 8, en el que la célula bacteriana es una célula de Nueva Delhi metalo-betalactamasa 1 (NDM-1) de *Escherichia coli*.
10. El constructo genético según el párrafo 8 ó 9, en el que el antibiótico es ceftriaxona, imipenem, piperacilina, tazobactam, gentamicina o una combinación de dos cualquiera o más de los anteriores.
- 50 11. El constructo genético según el párrafo 1, en el que el elemento de ADN contiene al menos una secuencia de microARN (miR), o al menos dos secuencias de miR en tándem.
12. El constructo genético según el párrafo 11, en el que el constructo genético contiene al menos dos elementos de código de barras en tándem.
- 55 13. El constructo genético según el párrafo 11 ó 12, en el que el miR está unido operativamente a un promotor inducible.
- 60 14. El constructo genético según uno cualquiera de los párrafos 11-13, en el que el miR se selecciona de miR-124, miR-128 y miR-138.
15. El constructo genético según uno cualquiera de los párrafos 11-13, en el que los al menos dos miR en tándem son una combinación seleccionada de miR-124, miR-128 y miR-138.
- 65 6. Un fagémido que comprende al menos un constructo genético según uno cualquiera de los párrafos 1-15.

17. Un bacteriófago recombinante que comprende el fagémido según el párrafo 16.

18. El bacteriófago recombinante según el párrafo 17, en el que el bacteriófago recombinante es de una familia seleccionada de *Myoviridae*, *Siphoviridae*, *Podoviridae*, *Tectiviridae*, *Corticoviridae*, *Lipothrixviridae*, *Plasmaviridae*, *Rudoviridae*, *Fuselloviridae*, *Inoviridae*, *Microviridae*, *Leviviridae* y *Cystoviridae*.

19. El bacteriófago recombinante según el párrafo 18, en el que el bacteriófago *Inoviridae* recombinante es un bacteriófago M13 o similar a M13.

20. Un método de cribado para genes que potencian la letalidad de los antibióticos, comprendiendo el método:

transformar células bacterianas resistentes a antibióticos con la colección que comprende dos o más constructos genéticos según uno cualquiera de los párrafos 1-15;

someter las células bacterianas transformadas a tratamiento con un inductor análogo, un antibiótico y una combinación de tratamiento con inductor análogo y antibiótico;

comparar fenotipos celulares entre cada condición de tratamiento; y

seleccionar fenotipos de destrucción sinérgica, crecimiento neutro y evaluar el fenotipo celular para cada tratamiento.

21. El método según el párrafo 20, que comprende además secuenciar constructos genéticos de célula bacterianas que tienen fenotipos seleccionados.

22. Un método para generar un constructo genético combinatorio, que comprende:

proporcionar un vector que contiene un primer constructo genético según uno cualquiera de los párrafos 1-15;

escindir el vector en el sitio de separación dentro del primer constructo genético, dando como resultado que el primer constructo genético se separe en segmentos primero y segundo;

proporcionar un segundo constructo genético según uno cualquiera de los párrafos A1-A15; y

aparear el segundo constructo genético al vector escindido, en el que el apareamiento se produce en elementos de extremos compatibles dentro de los constructos genéticos primero y segundo que son capaces de aparearse entre sí, y en el que tras el apareamiento, el segundo constructo genético se integra entre los segmentos primero y segundo del primer constructo genético, creando un constructo genético combinatorio.

23. El método según el párrafo 22, en el que el método es iterativo.

24. Un método para la identificación de un elemento de ADN o una pluralidad de elementos de ADN, que comprende:

proporcionar un constructo genético según uno cualquiera de los párrafos 1-15;

realizar un ensayo para determinar la secuencia de ADN del código de barras o la pluralidad de códigos de barras dentro del constructo genético y/o la secuencia de ADN del elemento de ADN o la pluralidad de elementos de ADN dentro del constructo genético; e

identificar el elemento de ADN o la pluralidad de elementos de ADN.

25. Una colección que comprende:

dos o más constructos genéticos según uno cualquiera de los párrafos 1-15.

26. Un método para generar un constructo genético combinatorio, que comprende:

proporcionar un vector que comprende:

un primer elemento de ADN,

un primer elemento de código de barras, y

dos elementos de recombinación específica de sitio ubicados entre el primer elemento de ADN y el primer

elemento de código de barras;

proporcionar un primer inserto que comprende:

5 un segundo elemento de ADN,

un segundo elemento de código de barras, y

10 elementos de recombinación específica de sitio que flanquean cada uno del segundo elemento de ADN y el segundo elemento de código de barras, de manera que dos elementos de recombinación específica de sitio están ubicados entre el segundo elemento de ADN y el segundo elemento de código de barras que no son compatibles con los elementos de recombinación específica de sitio dentro del vector, y dos elementos de recombinación específica de sitio están ubicados fuera del segundo elemento de ADN y el segundo elemento de código de barras que son compatibles con los elementos de recombinación específica de sitio dentro del vector;

15 realizar recombinación específica de sitio entre el vector y el primer inserto, en el que la recombinación específica de sitio se produce entre los elementos de recombinación específica de sitio dentro del vector ubicados entre el primer elemento de ADN y el primer elemento de código de barras y los elementos de recombinación específica de sitio compatibles dentro del primer inserto ubicados fuera del segundo elemento de ADN y el segundo elemento de código de barras, y en el que tras la recombinación específica de sitio, el primer inserto está ubicado dentro del vector, y el vector contiene múltiples elementos de ADN y múltiples elementos de código de barras, con dos elementos de recombinación específica de sitio ubicados entre los múltiples elementos de ADN y los múltiples elementos de código de barras;

25 proporcionar un segundo inserto que comprende:

un tercer elemento de ADN,

un tercer elemento de código de barras, y

30 elementos de recombinación específica de sitio que flanquean cada uno del tercer elemento de ADN y el tercer elemento de código de barras, de manera que dos elementos de recombinación específica de sitio están ubicados entre el tercer elemento de ADN y el tercer elemento de código de barras que no son compatibles con los dos elementos de recombinación específica de sitio ubicados entre los múltiples elementos de ADN y los múltiples elementos de código de barras del vector, y dos elementos de recombinación específica de sitio están ubicados fuera del tercer elemento de ADN y el tercer elemento de código de barras que son compatibles con los dos elementos de recombinación específica de sitio ubicados entre los múltiples elementos de ADN y los múltiples elementos de código de barras del vector;

40 realizar recombinación específica de sitio entre el vector y el segundo inserto, en el que la recombinación específica de sitio se produce entre los elementos de recombinación específica de sitio dentro del vector ubicados entre los múltiples elementos de ADN y los múltiples elementos de código de barras y los elementos de recombinación específica de sitio compatibles dentro del segundo inserto ubicados fuera del tercer elemento de ADN y el tercer elemento de código de barras, y en el que tras la recombinación específica de sitio, el segundo inserto está ubicado dentro del vector, y el vector contiene múltiples elementos de ADN y múltiples elementos de código de barras, con dos elementos de recombinación específica de sitio ubicados entre los múltiples elementos de ADN y los múltiples elementos de código de barras; y

50 repetir la recombinación específica de sitio un número n de veces, alternando entre recombinación específica de sitio entre el vector y el primer inserto y recombinación específica de sitio entre el vector y el segundo inserto, creando de ese modo un constructo genético combinatorio.

55 27. El método según el párrafo 26, en el que cada elemento de ADN contiene una secuencia de nucleótidos que codifica un factor de transcripción y/o enzima de modificación de histonas diferente, o único.

28. El método según el párrafo 26, en el que cada elemento de ADN contiene una secuencia de nucleótidos que codifica un microARN (miR) diferente, o único.

60 29. Un constructo genético combinatorio producido mediante el método según uno cualquiera de los párrafos 26-28.

30. Un método para la identificación de un elemento de ADN o una pluralidad de elementos de ADN, que comprende:

proporcionar un constructo genético combinatorio según el párrafo 29;

65 realizar un ensayo para determinar la secuencia de ADN de uno o más elementos de código de barras dentro del

constructo genético combinatorio y/o la secuencia de ADN de uno o más elementos de ADN dentro del constructo genético combinatorio; e

identificar el elemento de ADN o la pluralidad de elementos de ADN.

- 5
31. Un método para generar un constructo genético combinatorio, que comprende:
- proporcionar un vector que comprende:
- 10 un primer elemento de ADN,
- un primer elemento de código de barras, y
- 15 un sitio de reconocimiento para una primera enzima de restricción ubicado entre el primer elemento de ADN y el primer elemento de código de barras;
- proporcionar un inserto que comprende:
- 20 un segundo elemento de ADN,
- un segundo elemento de código de barras,
- 25 un sitio de reconocimiento para la primera enzima de restricción ubicado entre el segundo elemento de ADN y el segundo elemento de código de barras, y
- 30 dos sitios de reconocimiento para una o más enzimas de restricción que son distintas de la primera enzima de restricción ubicados fuera del segundo elemento de ADN y segundo elemento de código de barras, de manera que la digestión por restricción en el sitio de reconocimiento dentro del vector y en los dos sitios de reconocimiento ubicados fuera del segundo elemento de ADN y segundo elemento de código de barras dentro del inserto genera extremos compatibles;
- digerir el vector e inserto con enzimas de restricción;
- 35 aparear el inserto al vector, produciendo de ese modo un constructo genético combinatorio que contiene múltiples elementos de ADN y múltiples elementos de código de barras; y opcionalmente repetir el método un número n de veces.

32. El método según el párrafo 31, en el que cada elemento de ADN contiene una secuencia de nucleótidos que codifica un factor de transcripción y/o enzima de modificación de histonas diferente, o único.

33. El método según el párrafo 31, en el que cada elemento de ADN contiene una secuencia de nucleótidos que codifica un microARN (miR) diferente, o único.

34. Un constructo genético combinatorio producido mediante el método según uno cualquiera de los párrafos 31-33.

35. Un método para la identificación de un elemento de ADN o una pluralidad de elementos de ADN, que comprende:

proporcionar un constructo genético combinatorio según el párrafo 34;

realizar un ensayo para determinar la secuencia de ADN de uno o más elementos de código de barras dentro del constructo genético combinatorio y/o la secuencia de ADN de uno o más elementos de ADN dentro del constructo genético combinatorio; e

identificar el elemento de ADN o la pluralidad de elementos de ADN.

Secuencias

SEQ NO: 1 – Expresión de precursor de miR en tándem

CGATATCAAGCTTATCGATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTG
 ACTGGTATTCTTAACTATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAAT
 GCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTCATTTTCTCCTCCTTGTATAA
 ATCCTGGTTGCTGTCTCTTTATGAGGAGTTGTGGCCCGTTGTCAGGCAACGTGGC
 GTGGTGTGCACTGTGTTTGCTGACGCAACCCCCACTGGTTGGGGCATTGCCACCA
 CCTGTCAGCTCCTTTCCGGGACTTTCGCTTTCCTCCCTATTGCCACGGCGGAA
 CTCATCGCCGCTGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTG
 ACAATTCCGTGGTGTGTCGGGGAAATCATCGTCCTTTCCTTGGCTGCTCGCCTGT
 GTTGCCACCTGGATTCTGCGCGGGACGTCCTTCTGCTACGTCCCTTCGGCCCTCA
 ATCCAGCGGACCTTCCCTTCCCGCGGCCTGCTGCCGGCTCTGCGGCCTCTTCCGCG
 TCTTCGCCTTCGCCCTCAGACGAGTCGGATCTCCCTTTGGGCCGCTCCCCGCATC
 GATACCGTCGACCTCGAGACCTAGAAAAACATGGAGCAATCACAAGTAGCAATA
 CAGCAGCTACCAATGCTGATTGTGCCTGGCTAGAAGCACAAAGAGGAGGAGGAGG
 TGGGTTTTCCAGTCACACCTCAGGTACCTTTAAGACCAATGACTTACAAGGCAGC
 TGTAGATCTTAGCCACTTTTTAAAGAAAAGGGGGGACTGGAAGGGCTAATTCA
 CTCCCAACGAAGACAAGATATCCTTGATCTGTGGATCTACCACACACAAGGCTAC
 TTCCCTGATTGGCAGAACTACACACCAGGGCCAGGGATCAGATATCCACTGACCT
 TTGGATGGTGCTACAAGCTAGTACCAGTTGAGCAAGAGAAGGTAGAAGAAGCCA
 ATGAAGGAGAGAACACCCGCTTGTTACACCCTGTGAGCCTGCATGGGATGGATG
 ACCCGGAGAGAGAAGTATTAGAGTGGAGGTTTGACAGCCGCCTAGCATTTTCATC
 ACATGGCCCGAGAGCTGCATCCGGACTGTAAGGGTCTCTCTGGTTAGACCAGAT
 CTGAGCCTGGGAGCTCTCTGGCTAACTAGGGAACCCACTGCTTAAGCCTCAATAA
 AGCTTGCCTTGAGTGCTTCAAGTAGTGTGTGCCCGTCTGTTGTGTGACTCTGGTAA
 CTAGAGATCCCTCAGACCCTTTTAGTCAGTGTGGAAAATCTCTAGCAGGGCCCGT
 TTAACCCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTT

GCCCTCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCCTGTCCTTTCC
 TAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGG
 GGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGG
 CATGCTGGGGATGCGGTGGGCTCTATGGCTTCTGAGGCGGAAAGAACCAGCTGG
 GGCTCTAGGGGGTATCCCCACGCGCCCTGTAGCGGCGCATTAAAGCGCGGCGGGT
 GTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCCGCTC
 CTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCCGCCGCTTTCCCCGTCAAGCTC
 TAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCC
 CAAAAAATTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGAC
 GGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTTTGTTCC
 AAAGTGAACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGAT
 TTTGCCGATTTCCGCCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAAC
 GCGAATTAATTCTGTGGAATGTGTGTCAGTTAGGGTGTGGAAAGTCCCCAGGCTC
 CCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCAGGTG
 TGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAA
 TTAGTCAGCAACCATAGTCCCCGCCCTAACTCCGCCCATCCCGCCCCCTAACTCCG
 CCCAGTTCGCCCATTTCTCCGCCCATGGCTGACTAATTTTTTTTTATTTATGCAGA
 GGCCGAGGCCGCCTCTGCCTCTGAGCTATTCCAGAAGTAGTGAGGAGGCTTTTTT
 GGAGGCCTAGGCTTTTTGCAAAAAGCTCCCGGGAGCTTGTATATCCATTTTCGGAT
 CTGATCAGCACGTGTTGACAATTAATCATCGGCATAGTATATCGGCATAGTATAA
 TACGACAAGGTGAGGAACTAAACCATGGCCAAGTTGACCAGTGCCGTTCCGGTG
 CTCACCGCGCGCGACGTCGCCGGAGCGGTGAGTTCTGGACCGACCGGCTCGGG
 TTCTCCCGGACTTCGTGGAGGACGACTTCGCCGGTGTGGTCCGGGACGACGTGA
 CCCTGTTTCATCAGCGCGGTCCAGGACCAGGTGGTGCCGGACAACACCCTGGCCT
 GGGTGTGGGTGCGCGGCCCTGGACGAGCTGTACGCCGAGTGGTCCGGAGGTCGTGT
 CCACGAACCTCCGGGACGCCCTCCGGGCGGCCATGACCGAGATCGGGCAGCAGC
 CGTGGGGGCGGGAGTTCGCCCTGCGCGACCCGGCCGGCAACTGCGTGCCTTCG
 TGGCCGAGGAGCAGGACTGACACGTGCTACGAGATTTTCGATTCCACCGCCGCCTT
 CTATGAAAGGTTGGGCTTCGGAATCGTTTTCCGGGACGCCGGCTGGATGATCCTC
 CAGCGCGGGGATCTCATGCTGGAGTTCTTCGCCACCCCAACTTGTTTATTGCAG
 CTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTACAAATAAAGCATT
 TTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTATCATG
 TCTGTATACCGTCGACCTCTAGCTAGAGCTTGGCGTAATCATGGTCATAGCTGTTT
 CCTGTGTGAAATTGTTATCCGCTCACAATTCACACAACATACGAGCCGGAAGCA
 TAAAGTGTAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTT
 GCGCTCACTGCCCGCTTTCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGA
 ATCGGCCAACGCGCGGGGAGAGGCGGTTTTCGCTATTGGGCGCTCTTCCGCTTCT
 CGCTCACTGACTCGCTGCGCTCGGTGCTTCGGCTGCGGCGAGCGGTATCAGCTCA
 CTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAA
 CATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCT
 GCGTTTTTCCATAGGCTCCGCCCCCTGACGAGCATCACAAAAATCGACGCTCA
 AGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATAACCAGGCGTTTCCCCCT
 GGAAGCTCCCTCGTGCCTCTCCTGTTCCGACCTGCCGCTTACCGGATACCTGT
 CCGCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTAT
 CTCAGTTCGGTGTAGGTGCTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCG
 TTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGT
 AAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGC
 GAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTAC
 ACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAA
 AAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTT

TTTTGTTTGC AAGCAGCAGATTACGCGCAGAAAAAAGGATCTCAAGAAGATCC
TTTGATCTTTTCTACGGGGTCTGACGCTCAGTGAACGAAAACCTCACGTTAAGGG
ATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTA
AATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTA
CCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTTCGTTTCATCCA
TAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATC
TGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTA
TCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACT
TTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGAAGCTAGAGTAAGTAGTT
CGCCAGTTAATAGITTTGCGCAACGTGTGTTGCCATTGCTACAGGCATCGTGGTGTG
ACGCTCGTTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCCGA
GTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCCTCCGA
TCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACT
GCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGT
ACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCC
GGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCAT
CATTGGAAAACGTTCTTCGGGGCGAAAACCTCTCAAGGATCTTACCGCTGTTGAGA
TCCAGTTCGATGTAACCCACTCGTGCACCCAAGTATCTTCAGCATCTTTTACTTT
CACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAAATGCCGCAAAAAGG
GAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTCAATATTA
TTGAAGCATTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTT
AGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCGAAAAGTGCCACCTG
ACGTCCGACGGATCGGGAGATCTCCCGATCCCCTATGGTGCACCTCTCAGTACAATC
TGCTCTGATGCCGCATAGTTAAGCCAGTATCTGCTCCCTGCTTGTGTGTTGGAGGT
CGCTGAGTAGTGCGCGAGCAAAATTTAAGCTACAACAAGGCAAGGCTTGACCGA
CAATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGCTTCGCGATGTA
CGGGCCAGATATACGCGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAA
TTACGGGGTTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTAC
GGTAAATGGCCCCGCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAAT
AATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGG
GTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGC
CAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCCGCTGGCATTATGC
CCAGTACATGACCTTATGGGACTTTTCTACTTGGCAGTACATCTACGTATTAGTC
ATCGCTATTACCATGGTGTATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGC
GGTTTACTCACGGGGATTTCCAAGTCTCCACCCATTGACGTCAATGGGAGTTT
GTTTTGGCACCAAAAATCAACGGGACTTTCCAAAATGTCGTAACAACCTCCGCCCCA
TTGACGCAAATGGGCGGTAGGCGGTACGGTGGGAGGTCTATATAAGCAGCGCG
TTTTGCCTGTACTGGGTCTCTCTGGTTAGACCAGATCTGAGCCTGGGAGCTCTCTG
GCTAACTAGGGAACCCACTGCTTAAGCCTCAATAAAGCTTGCCTTGAGTGCTTCA
AGTAGTGTGTGCCCCGCTGTTGTGTGACTCTGGTAACTAGAGATCCCTCAGACCC
TTTTAGTCAGTGTGGAAAATCTCTAGCAGTGGCGCCCGAACAGGGACTTGAAAG
CGAAAGGGAAACCAGAGGAGCTCTCTCGACGCAGGACTCGGCTTGCTGAAGCGC
GCACGGCAAGAGGCGAGGGGCGGGGACTGGTGTGAGTACGCCAAAAATTTTACTA
GCGGAGGCTAGAAGGAGAGAGATGGGTGCGAGAGCGTCAGTATTAAGCGGGGG
AGAATTAGATCGCGATGGGAAAAAATTCGGTTAAGGCCAGGGGGAAAGAAAA
ATATAAATTAACATATAGTATGGGCAAGCAGGGAGCTAGAACGATTCCGAGT
TAATCCTGGCCTGTTAGAAACATCAGAAGGCTGTAGACAAATACTGGGACAGCT
ACAACCATCCCTTCAGACAGGATCAGAAGAAGTATAGATCATTATATAATACAGTA
GCAACCCTCTATTGTGTGCATCAAAGGATAGAGATAAAGACACCAAGGAAGCT
TTAGACAAGATAGAGGAAGAGCAAAAACAAAAGTAAGACCACCGCACAGCAAGC

GGCCGCTGATCTTCAGACCTGGAGGAGGAGATATGAGGGACAATTGGAGAAGTG
 AATTATATAAATATAAAGTAGTAAAAATTGAACCATTAGGAGTAGCACCCACCA
 AGGCAAAGAGAAGAGTGGTGCAGAGAGAAAAAAGAGCAGTGGGAATAGGAGCT
 TTGTTCTTGGGTTCTTGGGAGCAGCAGGAAGCACTATGGGCGCAGCGTCAATGA
 CGCTGACGGTACAGGCCAGACAATTATTGTCTGGTATAGTGCAGCAGCAGAACA
 ATTTGCTGAGGGCTATTGAGGCGCAACAGCATCTGTTGCAACTCACAGTCTGGGG
 CATCAAGCAGCTCCAGGCAAGAATCCTGGCTGTGGAAAGATACTAAAGGATCA
 ACAGCTCCTGGGGATTTGGGGTTGCTCTGGAAAACCTCATTTGCACCACTGCTGTG
 CCTTGGAAATGCTAGTTGGAGTAATAAATCTCTGGAACAGATTTGGAATCACACGA
 CCTGGATGGAGTGGGACAGAGAAATTAACAATTACACAAGCTTAATACACTCCT
 TAATTGAAGAATCGCAAACCAGCAAGAAAAGAATGAACAAGAATTATTGGAAT
 TAGATAAATGGGCAAGTTTGTGGAATTGGTTTAAACATAACAAATTGGCTGTGGTA
 TATAAAATTATTCATAATGATAGTAGGAGGCTTGGTAGGTTTAAAGAATAGTTTTT
 GCTGTACTTTCTATAGTGAATAGAGTTAGGCAGGGATATTCACCATTATCGTTTC
 AGACCCACCTCCCAACCCCGAGGGGACCCGACAGGCCCGAAGGAATAGAAGAA
 GAAGGTGGAGAGAGAGACAGAGACAGATCCATTTCGATTAGTGAACGGATCGGC
 ACTGCGTGCGCAATTCTGCAGACAAATGGCAGTATTCATCCACAATTTTAAAAG
 AAAAGGGGGGATTGGGGGGTACAGTGCAGGGGAAAGAATAGTAGACATAATAG
 CAACAGACATACAAATAAAGAATTACAAAAACAAATTACAAAAATTCAAAATT
 TTCGGGTTTATTACAGGGACAGCAGAGATCCAGTTTGGTTAATTAAGGGTGCAGC
 GGCTCCGCGCCGGGTTTTGGCGCCTCCCGCGGGCGCCCCCTCCTCACGGCGAG
 CGCTGCCACGTCAGACGAAGGGCGCAGGAGCGTTCCTGATCCTTCCGCCCCGAC
 GCTCAGGACAGCGCCCCGCTGCTCATAAGACTCGGCCTTAGAACCCCAAGTATCA
 GCAGAAGGACATTTTAGGACGGGACTTGGGTGACTCTAGGGCACTGGTTTTCTTT
 CCAGAGAGCGGAACAGGCGAGGAAAAGTAGTCCCTTCTCGGCGATTCTGCGGAG
 GGATCTCCGTGGGGCGGTGAACGCCGATGATTATATAAGGACGCGCCGGGTGTG
 GCACAGCTAGTTCCGTCGCAGCCGGGATTTGGGTGCGGTTCTTGTTTGTGGATC
 GCTGTGATCGTCACTTGGTGAGTTGCGGGCTGCTGGGCTGGCCGGGGCTTTCGTG
 GCCGCCGGGCCGCTCGGTGGGACGGAAGCGTGTGGAGAGACCGCCAAGGGCTGT
 AGTCTGGGTCCGCGAGCAAGGTTGCCCTGAACTGGGGGTTGGGGGGAGCGCACA
 AAATGGCGGCTGTTCCCGAGTCTTGAATGGAAGACGCTTGTAAAGGCGGGCTGTG
 AGGTCGTTGAAACAAGGTGGGGGGCATGGTGGGCGGCAAGAACCCAAGGTCTTG
 AGGCCTTCGCTAATGCGGGAAAGCTCTTATTCGGGTGAGATGGGCTGGGGCACC
 ATCTGGGGACCCTGACGTGAAGTTTGTCACTGACTGGAGAACTCGGGTTTGTCTG
 CTGGTTGCGGGGGCGGCAGTTATGCGGTGCCGTTGGGCAGTGCACCCGTACCTTT
 GGGAGCGCGCCCTCGTCTGTCGTGACGTCACCCGTTCTGTTGGCTTATAATGC
 AGGGTGGGGCCACCTGCCGGTAGGTGTGCGGTAGGCTTTTCTCCGTCGCAGGACG
 CAGGGTTCGGGCCTAGGGTAGGCTCTCCTGAATCGACAGGCGCCGGACCTCTGGT
 GAGGGGAGGGATAAGTGAGGCGTCAGTTTCTTTGGTTCGGTTTTATGTACCTATCT
 TCTTAAGTAGCTGAAGCTCCGGTTTTGAACTATGCGCTCGGGGTTGGCGAGTGTG
 TTTTGTGAAGTTTTTTAGGCACCTTTTGAATGTAATCATTTGGGTCAATATGTAA
 TTTTCAGTGTTAGACTAGTAAAGCTTCTGCAGGTGCACTCTAGAAAATTGTCCGC
 TAAATTCTGGCCGTTTTTTGGCTTTTTTGTTAGACAGTTAATTAAGCCACCATGAGC
 GAGCTGATCAAGGAGAACATGCACATGAAGCTGTACATGGAGGGCACCGTGAAC
 AACCACCACTTCAAGTGCACATCCGAGGGCGAAGGCAAGCCCTACGAGGGCACC
 CAGACCATGAAGATCAAGGTGGTTCGAGGGCGGCCCTCTCCCTTCGCCTTCGACA
 TCCTGGCTACCAGCTTCATGTACGGCAGCAAAGCCTTCATCAACCACACCCAGGG
 CATCCCCGACTTCTTTAAGCAGTCCCTCCCTGAGGGCTTACATGGGAGAGAATC
 ACCACATAACGAAGACGGGGGCGTGCTGACCGCTACCCAGGACACCAGCTTCCAG
 AACGGCTGCATCATCTACAACGTCAAGATCAACGGGGTGAACCTCCCATCCAAC

GGCCCTGTGATGCAGAAGAAAACACGCGGCTGGGAGGCCAACACCGAGATGCTG
TACCCCGCTGACGGCGGCTGAGAGGCCACAGCCAGATGGCCCTGAAGCTCGTG
GGCGGGGGCTACCTGCACTGCTCCTTCAAGACCACATACAGATCCAAGAAACCC
GCTAAGAACCTCAAGATGCCCGGCTTCCACTTCGTGGACCACAGACTGGAAAGA
ATCAAGGAGGCCGACAAAGAGACCTACGTTCGAGCAGCACGAGATGGCTGTGGCC
AAGTACTGCGACCTCCCTAGCAAACCTGGGGCACAGATAATAACCTGCAGGCGGG
CCAGATATACGCGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTAC
GGGGTCATTAGTTCATAGCCCATATATGGAGTTCGCGGTTACATAACTTACGGTA
AATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATG
ACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGG
AGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAG
TACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAG
TACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGC
TATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTT
GACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGT
GGCACCAAATCAACGGGACTTTCAAAATGTTCGTAACAACCTCCGCCCCATTGA
CGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGCGGTAACC
CCCGGGTACCGGTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGG
TGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCG
TGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCA
TCTGCACCACGGCAAGCTGCCCGTGGCCACCCTCGTGACCACCCTGAC
CTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTC
TTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGG
ACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGGCGACACCCTGG
TGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGG
GGCACAAGCTGGAGTACAACACTACAACAGCCACAACGTCTATATCATGGCCGACA
AGCAGAAGAACGGCATCAAGGTGAACCTTCAAGATCCGCCACAACATCGAGGACG
GCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCC
CCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAG
ACCCCAACGAGAAGCGCGATCACATGGTCCCTGCTGGAGTTCGTGACCGCCGCG
GGATCACTCTCGGCATGGACGAGCTGTACAAGTAAAGCGGCCGCGACTCTAGAG
ATCTCTACAGGAGGATCTCATGCATCGGATCCGCGAAAAGTCCGGAATTCATACT
GTGAAGTACACTGCATATAAGGAGTGTGGTATAGTATAAAGAACTTTCTGCAG
GTAGTAATTATAGTGAAGATTTTAGGTTTACAAAGCCCTAGCTGTTTTCTGTGTA
GTTTTTATTATTCTTATGACTCTTGACAAGTTTGTAGCTTCACCATATACATTTAA
TATTTTGAATAATTGGCCTTGTTCTGAGCTGTTGGATTCCGGGGCCGTAGCACTG
TCTGAGAGGTTTACATTTCTCACAGTGAACCGGTCTCTTTTTTACGCTGCTTCCTGG
CTTCTTTTTACTCAGGTTTCCACTGCTTTTTTTGCTTTTTTTAATGCTGTATGAAGGT
GTTAACATTTGTTTATATTTTTTCATTAATTGTAATACCTTTAAATCATGCATCATA
CTCAGAAATAGGGATTAGAATTTAAGTGACATCTTTGGCCCAATTCTTAGCCCCG
CAGACACTAGCGCCACCCCCGCGCCCGCGGTGCTGACGTCAGCCTGCAAGCCC
CGCCCCCGCTCTCCAGGGCAACCGTGGCTTTCGATTGTTACTGTGGGAACCGGA
GGTAACAGTCTACAGCCATGGTTCGCCCCGACGACGCCACGCTCCCCACCACTC
CCGAGTTCTGCCAGCCTGGGTTTGGGCAGATACAGAGCAAGAGGAGGCGGGGCA
ATTG

ES 2 663 424 T3

Nombre	de	Tipo	Mínimo	Máximo	Longitud
Precursor miR132		ncARN	11.755	11.979	225
Precursor miR128		ncARN	11.314	11.748	435
EGFP CDS		gen	10.525	11.244	720
CMVp CDS		CDS	9.918	10.494	577
turbo RFP		gen	9.191	9.886	696
hUbCp CDS		CDS	7.946	9.174	1.229
cPPT CDS		CDS	7.772	7.787	16
RRE CDS		CDS	7.008	7.241	234
Psi CDS		CDS	6.454	6.498	45
5'LTR	CDS;	CDS	6.163	6.343	181
del3'LTR CDS					
CMVp CDS		CDS	5.567	6.143	577
AMP CDS		CDS	4.238	5.238	1.001
SV40-pA CDS		CDS	3.007	3.126	120
Bleo CDS		CDS	2.500	2.871	372
EM7p CDS		CDS	2.432	2.499	68
SV40p CDS		CDS	2.070	2.338	269
Origen de F1		misc_feature	1.632	1.938	307
bGH-Ter CDS		CDS	1.342	1.569	228
5'LTR	CDS;	CDS	1.133	1.313	181
del3'LTR CDS					
U3PPT CDS		CDS	796	817	22
WPRE CDS		CDS	22	609	588
misc_signal	de	misc_signal	1	7.926	7.926
estructura principal de FuGW					

SEQ ID NO: 2 - Cebador directo:

5 AATGATACGGCGACCACCGAGATCTACACCGCTGGCAAGTGTAGC (SEQ ID NO: 2)

SEQ ID NO: 3 – Cebador inverso con código de barras:

10 CAAGCAGAAGACGGCATAACGAGATNNNNNNGGGAGGGCCCGTTG (SEQ ID NO: 3)

SEQ ID NO: 4 – Cebador de secuenciación de Illumina:

15 CCACGAGGATTCGAAAAGGTGAACCGACCCGGTCGATGCACTAGT (SEQ ID NO: 4)

SEQ ID NO: 5 – Cebador de indexación de Illumina:

20 CCTAGGAGCAAGTACGAACAACGGGCCCTCCC (SEQ ID NO: 5)

SEQ ID NO: 6 - Precursor de miR-124

AGGTGGGAGTACTGCTCAGAGCTACAACCTCTAGGAGTAGGGACTCCAAGCCTAG
 AGCTCCAAGAGAGGGTGAAGGGCAGGGAGAAAATTATAGTAATAGTTGCAATGA
 GTCACCTTGCTTCTAGATCAAGATCAGAGACTCTGCTCTCCGTGTTCCACAGCGGAC
 CTTGATTTAATGTCATAACAATTAAGGCACGCGGTGAATGCCAAGAGCGGAGCCT
 ACAGCTGCACTTGAAGGACATCCGAGAGAAGTTAGGAAGGGTGGGGAGAAACA
 ATTCTAGAATGAACCCATCCTGTGCGACAC

25 SEQ ID NO: 7 - Precursor de miR-128

ATACTGTGAAGTACACTGCATATAAGGAGTGTGGTATAGTATAAAGAACTTTCT
GCAGGTAGTAATTATAGTGAAGATTTTAGGTTTACAAAGCCCTAGCTGTTTTCTG
TGTAGCTTTTATTATTCTTATGACTCTTGACAAGTTTGTAGCTTCACCATATACAT
TTAATATTTTGCAATAATTGGCCTTGTTCTGAGCTGTTGGATTCGGGGCCGTAGC
ACTGTCTGAGAGGTTTACATTTCTCACAGTGAACCGGTCTCTTTTTTCAGCTGCTTC
CTGGCTTCTTTTTACTCAGGTTTCCACTGCTTTTTTTGCTTTTTTTAATGCTGTATGA
AGGTGTTAACATTTGTTTATATTTTTTCATTAATTGTAATACCTTTAAATCATGCAT
CATACTCAGAAATAGGGATTAGAATTTAAGTGACATCTTTGGCC

SEQ ID NO: 8 - Precursor de miR-132

CTAGCCCCGCAGACACTAGCGCCACCCCCGCCGCCCGCGGTGCTGACGTCAGCCT
GCAAGCCCCGCCCGCGTCTCCAGGGCAACCGTGGCTTTCGATTGTTACTGTGG
GAACCGGAGGTAACAGTCTACAGCCATGGTCGCCCCGCAGCACGCCACGCTCC
CCACCACTCCCGAGTTCTGCCAGCCTGGGTTTGGGCAGATACAGAGCAAGAGGA
5 GGCGGGG

SEQ ID NO: 9 – Sensor de miR-124

GGCACAGATAATAACCTGCAAAAAGGCATTCACCGCGTGCCTTAGGCATTCACCG
CGTGCCTTAGGCATTCACCGCGTGCCTTAGGCATTCACCGCGTGCCTTAAATGCAG
10 GCGGGCCAGATATAC

SEQ ID NO: 10 – Sensor de miR-128

GGCACAGATAATAACCTGCAAAAAGAGACCGGTTCACTGTGAAAAGAGACCGGT
TCACTGTGAAAAGAGACCGGTTCACTGTGAAAAGAGACCGGTTCACTGTGAAAT
GCAGGCGGGCCAGATATAC

SEQ NO: 11- Sensor de miR-132

GGCACAGATAATAACCTGCAAAAACGACCATGGCTGTAGACTGTTACGACCATGG
CTGTAGACTGTTACGACCATGGCTGTAGACTGTTACGACCATGGCTGTAGACTGTT
15 AAATGCAGGCGGGCCAGATATAC

Bibliografía

- 5 1. Kitagawa M, Ara T, Arifuzzaman M, Ioka-Nakamichi T, Inamoto E, *et al.* 2005. Complete set of ORF clones of *Escherichia coli* ASKA library (a complete set of *E. coli* K-12 ORF archive): unique resources for biological research. DNA research: an international journal for rapid publication of reports on genes and genomes 12:291-9
 - 10 2. Soo VW, Hanson-Manful P, Patrick WM. 2011. Artificial gene amplification reveals an abundance of promiscuous resistance determinants in *Escherichia coli*. Proceedings of the National Academy of Sciences of the United States of America 108:1484-9
 - 15 3. Butland G, Babu M, Diaz-Mejia JJ, Bohdana F, Phanse S, *et al.* 2008. eSGA: *E. coli* synthetic genetic array analysis. Nature Methods 5:789-95
 - 20 4. Pan X, Yuan DS, Xiang D, Wang X, Sookhai-Mahadeo S, *et al.* 2004. A robust toolkit for functional profiling of the yeast genome. Molecular Cell 16:487-96
 - 25 5. Tong AH, Lesage G, Bader GD, Ding H, Xu H, *et al.* 2004. Global mapping of the yeast genetic interaction network. Science 303:808-13
 - 30 6. Gibson DG, Young L, Chuang RY, Venter JC, Hutchison CA, 3rd, Smith HO. 2009. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nature Methods 6:343-5
 - 35 7. Anderson JC, Dueber JE, Leguia M, Wu GC, Goler JA, *et al.* 2010. BglBricks: A flexible standard for biological part assembly. Journal of Biological Engineering 4:1
 - 40 8. Son EY, Ichida JK, Wainger BJ, Toma JS, Rafuse VF, *et al.* 2011. Conversion of mouse and human fibroblasts into functional spinal motor neurons. Cell Stem Cell 9:205-18
 - 45 9. Merryman CG, D.G. 2012. Methods and applications for assembling large DNA constructs. Metabolic Engineering 14:196-204
 - 50 10. Roth FP, Suzuki Y, Mellor J. Publicación de patente estadounidense n.º US 2009/0098555.
 - 55 11. Wang HH, Isaacs FJ, Carr PA, Sun ZZ, Xu G, *et al.* 2009. Programming cells by multiplex genome engineering and accelerated evolution. Nature 460:894-8
 - 60 12. Mochon, A. B. *et al.* 2011. New Delhi metallo-beta-lactamase (NDM-1)-producing *Klebsiella pneumoniae*: case report and laboratory detection strategies. Journal of clinical microbiology 49:1667-1670.
 13. Yu, H. *et al.* Next-generation sequencing to generate interactome datasets. Nature methods 8, 478-480, doi:10.1038/nmeth.1597 (2011).
 14. Berezikov, E. *et al.* Evolution of microRNA diversity and regulation in animals. Nat Rev Genet 12, 846-860 (2011).
 15. Ambros, V. The functions of animal microRNAs. Nature 431, 350-355 (2004).
 16. Gangarajiu, V.K. and Lin, H. MicroRNAs: key regulators of stem cells. Nat Rev Mol Cell Bio 10, 116-125 (2009).
 17. Inui, M. *et al.* MicroRNA control of signal transduction. Nat Rev Mol Cell Bio 11, 222-263 (2010).
 18. Esteller M., *et al.*, Non-coding RNAs in human disease. Nat Rev Genet 12, 861-874 (2011).
 19. Eacker, S.M. *et al.* Understanding microRNAs in neurodegeneration. Nat Rev Neurosci 10, 837-841 (2009).
 20. Kong, Y.W. *et al.* MicroRNAs in cancer management. Lancet Oncol 13, e249-258 (2012).
- Habiendo descrito, por tanto, varios aspectos de al menos una realización de esta invención, ha de apreciarse que a los expertos en la técnica se les ocurrirán fácilmente diversas alteraciones, modificaciones y mejoras. Por consiguiente, la descripción anterior y los dibujos son a modo de ejemplo únicamente.

Equivalentes

Aunque en el presente documento se han descrito e ilustrado varias realizaciones inventivas, los expertos habituales en la técnica contemplarán fácilmente una variedad de otros medios y/o estructuras para realizar la función y/u obtener los resultados y/o una o más de las ventajas descritas en el presente documento, y se considera que cada una de tales variaciones y/o modificaciones está dentro del alcance de las realizaciones inventivas descritas en el presente documento. Más generalmente, los expertos en la técnica apreciarán fácilmente que se pretende que todos los parámetros, dimensiones, materiales y configuraciones descritos en el presente documento sean a modo de ejemplo y que los parámetros, dimensiones, materiales y/o configuraciones reales dependerán de la aplicación o aplicaciones específicas para las que se usa(n) las enseñanzas inventivas. Los expertos en la técnica reconocerán, o podrán establecer usando únicamente experimentación de rutina, muchos equivalentes a las realizaciones inventivas específicas descritas en el presente documento. Por tanto, ha de entenderse que las realizaciones anteriores se presentan a modo de ejemplo únicamente y que, dentro del alcance de las reivindicaciones adjuntas y equivalentes a las mismas, pueden ponerse en práctica realizaciones inventivas distintas a las descritas y reivindicadas específicamente. Las realizaciones inventivas de la presente divulgación se refieren a cada característica, sistema, artículo, material, kit y/o método individual descrito en el presente documento. Además, cualquier combinación de dos o más de tales características, sistemas, artículos, materiales, kits y/o métodos, si tales características, sistemas, artículos, materiales, kits y/o métodos no son incompatibles entre sí, está incluido dentro del alcance inventivo de la presente divulgación.

Debe entenderse que todas las definiciones, tal como se definen y se usan en el presente documento, dominan con respecto a definiciones de diccionarios, y/o significados habituales de los términos definidos.

Los artículos indefinidos “un” y “una”, tal como se usa en el presente documento en la memoria descriptiva y en las reivindicaciones, a menos que se indique claramente lo contrario, debe entenderse que significan “al menos uno”.

La expresión “y/o”, tal como se usa en el presente documento en la memoria descriptiva y en las reivindicaciones, debe entenderse que significa “cualquiera o ambos” de los elementos unidos, es decir, elementos que están presentes de manera conjuntiva en algunos casos y que están presentes de manera disyuntiva en otros casos. Múltiples elementos enumerados con “y/o” deben interpretarse del mismo modo, es decir, “uno o más” de los elementos unidos. Opcionalmente pueden estar presentes otros elementos distintos de los elementos identificados específicamente por la conjunción “y/o”, ya estén relacionados o no con los elementos identificados específicamente. Por tanto, como ejemplo no limitativo, una referencia a “A y/o B”, cuando se usa conjuntamente con vocabulario abierto tal como “que comprende” puede referirse, en una realización, sólo a A (incluyendo opcionalmente elementos distintos de B); en otra realización, sólo a B (incluyendo opcionalmente elementos distintos de A); aún en otra realización, tanto a A como a B (incluyendo opcionalmente otros elementos); etc.

Tal como se usa en el presente documento en la memoria descriptiva y en las reivindicaciones, debe entenderse que “o” tiene el mismo significado que “y/o”, tal como se definió anteriormente. Por ejemplo, cuando se separan elementos en una lista, se interpretará que “o” o “y/o” son inclusivos, es decir, la inclusión de al menos uno, pero también incluyendo más de uno, de varios o de una lista de elementos, y, opcionalmente, elementos no enumerados adicionales. Sólo los términos que indiquen claramente lo contrario, tal como “sólo uno de” o “exactamente uno de”, o, cuando se usen en las reivindicaciones, “que consiste en”, harán referencia a la inclusión de exactamente un elemento de varios o de una lista de elementos. En general, se interpretará que el término “o” tal como se usa en el presente documento indica alternativas exclusivas (es decir, “uno o los otros pero no ambos”) cuando va precedido por términos de exclusividad, tales como “cualquiera”, “uno de”, “sólo uno de”, o “exactamente uno de”. “Que consiste esencialmente en”, cuando se usa en las reivindicaciones, tendrá su significado habitual tal como se usa en el campo de la ley de patentes.

Tal como se usa en el presente documento en la memoria descriptiva y en las reivindicaciones, debe entenderse que la expresión “al menos uno”, en referencia a una lista de uno o más elementos, significa al menos un elemento seleccionado de uno cualquiera o más de los elementos en la lista de elementos, pero sin incluir necesariamente al menos uno de todos y cada uno de los elementos enumerados específicamente dentro de la lista de elementos y sin excluir ninguna combinación de elementos en la lista de elementos. Esta definición también permite que puedan estar presentes opcionalmente elementos distintos de los elementos identificados específicamente dentro de la lista de elementos a la que se refiere la expresión “al menos uno”, ya estén relacionados o no con los elementos identificados específicamente. Por tanto, como ejemplo no limitativo, “al menos uno de A y B” (o, de manera equivalente, “al menos uno de A o B”, o, de manera equivalente “al menos uno de A y/o B”) puede referirse, en una realización, a al menos uno, incluyendo opcionalmente más de uno, de A, sin estar B presente (e incluyendo opcionalmente elementos distintos de B); en otra realización, a al menos uno, incluyendo opcionalmente más de uno, de B, sin estar A presente (e incluyendo opcionalmente elementos distintos de A); aún en otra realización, a al menos uno, incluyendo opcionalmente más de uno, de A, y al menos uno, incluyendo opcionalmente más de uno, de B (e incluyendo opcionalmente otros elementos); etc.

También debe entenderse que, a menos que se indique claramente lo contrario, en cualquiera de los métodos reivindicados en el presente documento que incluyen más de una etapa o acto, el orden de las etapas o actos del

método no se limita necesariamente al orden en el que se enumeran las etapas o actos del método.

5 En las reivindicaciones, así como en la memoria descriptiva anterior, ha de entenderse que todas las expresiones de transición tales como "que comprende", "que incluye", "que porta", "que tiene", "que contiene", "que implica", "que mantiene", "compuesto por", y similares son abiertas, es decir, significan incluyendo pero sin limitarse a. Sólo las expresiones de transición "que consiste en" y "que consiste esencialmente en" serán expresiones de transición cerradas o semicerradas, respectivamente, tal como se expone en el Manual de la Oficina de Patentes de los Estados Unidos de los Procedimientos de Examen de Patentes, sección 2111.03.

Lista de secuencias

	<110> Instituto Tecnológico de Massachusetts	
5	<120> Genética combinatoria masivamente en paralelo	
	<130> M0656.70273WO00	
	<140> No asignado aún	
10	<141> 28-06-2013	
	<150> 61/666.081	
	<151> 29-06-2012	
15	<160> 11	
	<170> PatentIn versión 3.5	
	<210> 1	
20	<211> 11985	
	<212> ADN	
	<213> Secuencia artificial	
	<220>	
25	<223> Precursor de miR en tandem	
	<400> 1	
	cgatatcaag cttatcgata atcaacctct ggattacaaa atttgtgaaa gattgactgg	60
	tattcttaac tatgttgctc cttttacgct atgtggatac gctgctttaa tgcctttgta	120
	tcatgctatt gcttcccgta tggctttcat tttctcctcc ttgtataaat cctggttgct	180
	gtctctttat gaggagtgtg ggcccgttgt caggcaacgt ggcgtggtgt gcactgtgtt	240
	tgctgacgca acccccactg gttggggcat tgccaccacc tgtcagctcc tttccgggac	300
	tttcgctttc ccctcccta ttgccacggc ggaactcatc gccgcctgcc ttgccgctg	360
	ctggacaggg gctcggctgt tgggactga caattccgtg gtgttgcgg ggaatcatc	420
	gtcctttcct tggctgctcg cctgtgttgc cacctggatt ctgcgcggga cgtccttctg	480
	ctacgtccct tcggccctca atccagcggc ccttccttcc cgcggcctgc tgccggctct	540
	gcggcctctt ccgctcttc gccttcgcc tcagacgagt cggatctccc tttgggccgc	600
	ctccccgcat cgataccgtc gacctcgaga cctagaaaaa catggagcaa tcacaagtag	660
	caatacagca gctaccaatg ctgattgtgc ctggctagaa gcacaagagg aggaggaggt	720
	gggttttcca gtcacacctc aggtacctt aagaccaatg acttacaagg cagctgtaga	780
	tcttagccac tttttaaag aaaagggggg actggaagg ctaattcact cccaacgaag	840
	acaagatct cttgatctgt ggatctacca cacacaaggc tacttccctg attggcagaa	900
	ctacacacca gggccaggga tcagatatcc actgaccttt ggatggtgct acaagctagt	960
	accagttgag caagagaagg tagaagaagc caatgaagga gagaacaccc gcttgttaca	1020
	ccctgtgagc ctgcatggga tggatgacct ggagagagaa gtattagagt ggaggtttga	1080

ES 2 663 424 T3

cagccgccta gcatttcata acatggcccg agagctgcat ccggactgta ctgggtctct 1140
 ctggttagac cagatctgag cctgggagct ctctggctaa ctaggaacc cactgcttaa 1200
 gcctcaataa agcttgccctt gagtgcttca agtagtgtgt gcccgctctgt tgtgtgactc 1260
 tggtaactag agatccctca gaccctttta gtcagtgtgg aaaatctcta gcagggcccg 1320
 tttaaaccog ctgatcagcc tcgactgtgc cttctagttg ccagccatct gttgtttgcc 1380
 cctccccogt gccttccttg accctggaag gtgccactcc cactgtcctt tcctaataaa 1440
 atgaggaaat tgcacgcat tgtctgagta ggtgtcattc tattctgggg ggtggggtgg 1500
 ggcaggacag caagggggag gattgggaag acaatagcag gcatgctggg gatgcggtgg 1560
 gctctatggc ttctgaggcg gaaagaacca gctggggctc tagggggtat cccacgcgc 1620
 cctgtagcgg cgcattaagc gcggcgggtg tggtggttac gcgcagcgtg accgctacac 1680
 ttgccagcgc cctagcgcgc gctcctttcg ctttcttccc tcctttctc gccacgttcg 1740
 ccggttttcc ccgtcaagct ctaaactcggg ggctcccttt agggttccga tttagtgttt 1800
 tacggcacct cgaccccaaa aaacttgatt aggggtgatgg ttcacgtagt gggccatcgc 1860
 cctgatagac ggtttttctgc cctttgacgt tggagtccac gttctttaat agtggactct 1920
 tgttccaaac tggacaaca ctcaacccta tctcggctta ttcttttgat ttataagggg 1980
 ttttgccgat ttccgcctat tggttaaaaa atgagctgat ttaacaaaaa tttaacgcga 2040
 attaattctg tggaatgtgt gtcagttagg gtgtggaaag tccccaggct cccagcagg 2100
 cagaagtatg caaagcatgc atctcaatta gtcagcaacc aggtgtggaa agtccccagg 2160
 ctccccagca ggcagaagta tgcaaagcat gcatctcaat tagtcagcaa ccatagtccc 2220
 gccctaact ccgcccattc cgcacctaac tccgcccagt tccgcccatt ctccgcccc 2280
 tggctgacta atttttttta tttatgcaga ggccgaggcc gcctctgcct ctgagctatt 2340
 ccagaagtag tgaggaggct tttttggagg cctaggcttt tgcaaaaagc tcccgggagc 2400
 ttgtatatcc attttccgat ctgatcagca cgtgttgaca attaatacgc ggcatagtat 2460
 atcggcatag tataatacga caaggtgagg aactaaacca tggccaagtt gaccagtgcc 2520
 gttccgggtgc tcaccgcgcg cgacgtcgcc ggagcggctc agttctggac cgaccggctc 2580
 gggttctccc gggacttcgt ggaggacgac ttccgccgtg tggccggga cgacgtgacc 2640
 ctgttcatca gcgcgggtcca ggaccagggtg gtgccggaca acacctggc ctgggtgtgg 2700
 gtgcgcggcc tggacgagct gtacgcagag tggtcggagg tcgtgtccac gaacttccgg 2760
 gacgcctccg ggccggccat gaccgagatc ggcgagcagc cgtggggggcg ggagttccgc 2820
 ctgcgcgacc cggccggcaa ctgcgtgcac ttcgtggccg aggagcagga ctgacacgtg 2880
 ctacgagatt tcgattccac cgcgccttc tatgaaaggt tgggcttcgg aatcgttttc 2940
 cgggacgccg gctggatgat cctccagcgc ggggatctca tgctggagtt cttcgcacc 3000

ES 2 663 424 T3

cccaacttgt ttattgcagc ttataatggt tacaaataaa gcaatagcat cacaaatttc 3060
 acaaataaag catttttttc actgcattct agttgtgggt tgtccaaact catcaatgta 3120
 tcttatcatg tctgtatacc gtcgacctct agctagagct tggcgtaatc atgggtcatag 3180
 ctgtttcctg tgtgaaattg ttatccgctc acaattccac acaacatacg agccggaagc 3240
 ataaagtgta aagcctgggg tgcctaataga gtgagctaac tcacattaat tgcgttgctc 3300
 tcaactgcccg ctttccagtc gggaaacctg tcgtgccagc tgcattaatg aatcggccaa 3360
 cgcgcgggga gaggcggtt gcgtattggg cgctcttccg cttcctcgct cactgactcg 3420
 ctgcgctcgg tcgttcggct gcggcgagcg gtatcagctc actcaaaggc ggtaatacgg 3480
 ttatccacag aatcagggga taacgcagga aagaacatgt gagcaaaagg ccagcaaaag 3540
 gccaggaacc gtaaaaaggc cgcggttgctg gcgtttttcc ataggtccg cccccctgac 3600
 gagcatcaca aaaatcgagc ctcaagtcag aggtggcgaa acccgacagg actataaaga 3660
 taccaggcgt ttccccctgg aagctccctc gtgcgctctc ctgttccgac cctgccgctt 3720
 accggatacc tgtccgcctt tctcccttcc ggaagcgtgg cgctttctca tagctcacgc 3780
 tgtaggtatc tcagttcggg ttaggtcgtt cgctccaagc tgggctgtgt gcacgaacct 3840
 cccgttcagc ccgaccgctg cgccttatcc ggtaactatc gtcttgagtc caaccggta 3900
 agacacgact tatcgccact ggcagcagcc actggtaaca ggattagcag agcgaggtat 3960
 gtaggcggtg ctacagagtt cttgaagtgg tggcctaact acggctacac tagaagaaca 4020
 gtatttggtg tctgcgctct gctgaagcca gttaccttcc gaaaaagagt tggtagctct 4080
 tgatccggca acaaaccac cgctggtagc ggtggttttt ttgtttgcaa gcagcagatt 4140
 acgcgacagaa aaaaaggatc tcaagaagat cctttgatct tttctacggg gtctgacgct 4200
 cagtggaacg aaaactcacg ttaagggatt ttgggtcatga gattatcaa aaggatcttc 4260
 acctagatcc ttttaaatta aaaatgaagt tttaaatcaa tctaaagtat atatgagtaa 4320
 acttggctctg acagttacca atgcttaatc agtgaggcac ctatctcagc gatctgtcta 4380
 tttcgttcat ccatagttgc ctgactcccc gtcgtgtaga taactacgat acgggagggc 4440
 ttaccatctg gccccagtgc tgcaatgata ccgcgagacc cacgctcacc ggctccagat 4500
 ttatcagcaa taaaccagcc agccggaagg gccgagcgca gaagtgggcc tgcaacttta 4560
 tccgcctcca tccagtctat taattggtgc cgggaagcta gagtaagtag ttcgccagtt 4620
 aatagtttgc gcaacgttgt tgccattgct acaggcatcg tgggtgcacg ctcgctgctt 4680
 ggtatggctt cattcagctc cggttcccaa cgatcaaggc gagttacatg atccccatg 4740
 ttgtgcaaaa aagcggtag ctcttccggt cctccgatcg ttgtcagaag taagttggcc 4800
 gcagtggtat cactcatggt tatggcagca ctgcataatt ctcttactgt catgccatcc 4860

ES 2 663 424 T3

gtaagatgct tttctgtgac tggtgagtac tcaaccaagt cattctgaga atagtgtatg 4920
cggcgaccga gttgctcttg cccggcgtca atacgggata ataccgcgcc acatagcaga 4980
actttaaaag tgctcatcat tggaaaacgt tcttcggggc gaaaactctc aaggatctta 5040
ccgctggtga gatccagttc gatgtaacct actcgtgcac ccaactgac ttcagcatct 5100
tttactttca ccagcgtttc tgggtgagca aaaacaggaa ggcaaatgc cgcaaaaaag 5160
ggaataaggg cgacacggaa atggtgaata ctcatactct tcctttttca atattattga 5220
agcatttatac agggttattg tctcatgagc ggatacatat ttgaatgtat ttagaaaaat 5280
aaacaaatag gggttccgcg cacatctccc cgaaaagtgc cacctgacgt cgacggatcg 5340
ggagatctcc cgatccccta tggtgactc tcagtacaat ctgctctgat gccgcatagt 5400
taagccagta tctgctccct gcttgtgtgt tggaggctgc tgagtagtgc gcgagcaaaa 5460
ttaaagctac aacaaggcaa ggcttgaccg acaattgcat gaagaatctg cttagggtta 5520
ggcgttttgc gctgcttcgc gatgtacggg ccagatatac gcgttgacat tgattattga 5580
ctagttatta atagtaatca attacggggc cattagttca tagcccatat atggagttcc 5640
gcgttacata acttacggta aatggcccgc ctggctgacc gcccaacgac ccccgcccat 5700
tgacgtcaat aatgacgtat gttcccatag taacgccaat agggactttc cattgacgtc 5760
aatgggtgga gtatttacgg taaactgcc acttggcagt acatcaagtg tatcatatgc 5820
caagtaogcc ccctattgac gtcaatgacg gtaaatggcc cgcctggcat tatgccagt 5880
acatgacctt atgggacttt cctacttggc agtacatcta cgtattagtc atcgctatta 5940
ccatggtgat gcggttttgg cagtacatca atgggcgtgg atageggttt gactcacggg 6000
gatttccaag tctccacccc attgacgtca atgggagttt gttttggcac caaaatcaac 6060
gggactttcc aaaatgtcgt aacaactccg cccattgac gcaaatgggc ggtaggcgtg 6120
tacggtggga ggtctatata agcagcgcgt tttgcctgta ctgggtctct ctggttagac 6180
cagatctgag cctgggagct ctctggctaa ctagggaacc cactgcttaa gcctcaataa 6240
agcttgccct gagtgcttca agtagtgtgt gcccgctctgt tgtgtgactc tggtaactag 6300
agatccctca gaccctttta gtcagtgtgg aaaatctcta gcagtggcgc ccgaacaggg 6360
acttgaaagc gaaagggaaa ccagaggagc tctctcgacg caggactcgg cttgctgaag 6420
cgcgcacggc aagaggcgag gggcggcgac tggtgagtac gccaaaaatt ttgactagcg 6480
gaggctagaa ggagagagat gggtgcgaga gcgtcagtat taagcggggg agaattagat 6540
cgcgatggga aaaaattcgg ttaaggccag ggggaaagaa aaaatataaa ttaaacata 6600
tagtatgggc aagcagggag ctagaacgat tcgcagttaa tcctggcctg ttagaaacat 6660
cagaaggctg tagacaaata ctgggacagc tacaaccatc ccttcagaca ggatcagaag 6720
aacttagatc attatataat acagtagcaa ccctctattg tgtgcatcaa aggatagaga 6780

ES 2 663 424 T3

taaaagacac	caaggaagct	ttagacaaga	tagaggaaga	gcaaaacaaa	agtaagacca	6840
ccgcacagca	agcggccgct	gatcttcaga	cctggaggag	gagatatgag	ggacaattgg	6900
agaagtgaat	tatataaata	taaagtagta	aaaattgaac	cattaggagt	agcaccacc	6960
aaggcaaaga	gaagagtgg	gcagagagaa	aaaagagcag	tgggaatagg	agctttgttc	7020
cttgggttct	tgggagcagc	aggaagcact	atgggcccag	cgtcaatgac	gctgacggta	7080
caggccagac	aattattgtc	tggatatagt	cagcagcaga	acaatttgct	gagggctatt	7140
gaggcgcaac	agcatctgtt	gcaactcaca	gtctggggca	tcaagcagct	ccaggcaaga	7200
atcctggctg	tggaaagata	cctaaaggat	caacagctcc	tggggatttg	gggttgctct	7260
ggaaaactca	tttgcaccac	tgctgtgcct	tggaatgcta	gttgagtaa	taaactctctg	7320
gaacagattt	ggaatcacac	gacctggatg	gagtgggaca	gagaaattaa	caattacaca	7380
agcttaatac	actccttaat	tgaagaatcg	caaaaccagc	aagaaaagaa	tgaacaagaa	7440
ttattggaat	tagataaatg	ggcaagtttg	tggaattggt	ttaacataac	aaattggctg	7500
tggatatataa	aattattcat	aatgatagta	ggaggcttgg	taggtttaag	aatagttttt	7560
gctgtacttt	ctatagttaa	tagagttagg	cagggatatt	caccattatc	gtttcagacc	7620
cacctcccaa	ccccgagggg	acccgacagg	cccgaaggaa	tagaagaaga	aggtggagag	7680
agagacagag	acagatccat	tcgattagtg	aacggatcgg	cactgcgtgc	gccaattctg	7740
cagacaaatg	gcagtattca	tccacaattt	taaaagaaaa	ggggggattg	gggggtacag	7800
tgcaggggaa	agaatagtag	acataatagc	aacagacata	caactaaag	aattacaaaa	7860
acaaattaca	aaaattcaaa	atthtcgggt	ttattacagg	gacagcagag	atccagtttg	7920
gttaattaag	ggtgcagcgg	cctccgcgcc	gggttttggc	gcctcccgcg	ggcgcccccc	7980
tcctcacggc	gagcgtgcc	acgtcagacg	aagggcgcag	gagcgttcct	gatccttccg	8040
cccggacgct	caggacagcg	gcccgtgct	cataagactc	ggccttagaa	ccccagtatc	8100
agcagaagga	cattttagga	cgggacttgg	gtgactctag	ggcactggtt	ttctttccag	8160
agagcggaac	aggcgaggaa	aagtagtccc	ttctcggcga	ttctgcggag	ggatctccgt	8220
ggggcggtga	acgccgatga	ttatataagg	acgcgccggg	tgtggcacag	ctagttccgt	8280
cgcagccggg	atttgggtcg	cggttcttgt	ttgtggatcg	ctgtgatcgt	cacttgggtga	8340
gttgccggct	gctgggctgg	ccggggcttt	cgtggccgcc	gggccgctcg	gtgggacgga	8400
agcgtgtgga	gagaccgcca	agggctgtag	tctgggtccg	cgagcaaggt	tgccctgaac	8460
tgggggttgg	ggggagcgca	caaaatggcg	gctgttcccg	agtcttgaat	ggaagacgct	8520
tgtaaggcgg	gctgtgaggt	cgttgaaaca	aggtgggggg	catggtgggc	ggcaagaacc	8580
caaggtcttg	aggccttcgc	taatgcggga	aagctcttat	tcgggtgaga	tgggctgggg	8640

ES 2 663 424 T3

caccatctgg ggaccctgac gtgaagtttg tcaactgactg gagaactcgg gtttgtcgtc 8700
 tggttgcggg ggcggcagtt atgcggtgcc gttgggcagt gcacccgtac ctttgggagc 8760
 gcgcgcctcg tcgtgtcgtg acgtcaccocg ttctgttggc ttataatgca ggggtggggcc 8820
 acctgccggt aggtgtgcgg taggcctttc tccgtcgcag gacgcagggt tcgggcctag 8880
 ggtaggctct cctgaatcga caggcgcocgg acctctggtg aggggagggga taagtgagggc 8940
 gtcagtttct ttggtcgggtt ttatgtacct atcttcttaa gtagctgaag ctccggtttt 9000
 gaactatgcg ctccggggttg gcgagtgtgt tttgtgaagt tttttaggca cttttgaaa 9060
 tgtaatcatt tgggtcaata tgtaattttc agtgtttagac tagtaaagct tctgcaggtc 9120
 gactctagaa aattgtccgc taaattctgg ccgtttttgg cttttttggt agacagttaa 9180
 ttaagccacc atgagcgcgc tgatcaagga gaacatgcac atgaagctgt acatggaggg 9240
 caccgtgaac aaccaccact tcaagtgcac atccgagggc gaaggcaagc cctacgaggg 9300
 caccagacc atgaagatca aggtggtcga gggcggccct ccccccttcg ccttcgacat 9360
 cctggctacc agcttcatgt acggcagcaa agccttcac c aaccacacc agggcatccc 9420
 cgacttcttt aagcagtcct tccctgaggg cttcacatgg gagagaatca ccacatacga 9480
 agacgggggc gtgctgaccg ctaccagga caccagcttc cagaacggct gcatcatcta 9540
 caacgtcaag atcaacgggg tgaacttccc atccaacggc cctgtgatgc agaagaaaac 9600
 acgcggtcgg gaggccaaca ccgagatgct gtaccccgct gacggcggcc tgagaggcca 9660
 cagccagatg gccctgaagc tcgtgggocgg gggctacctg cactgctcct tcaagaccac 9720
 atacagatcc aagaaaccocg ctaagaacct caagatgccc ggcttccact tcgtggacca 9780
 cagactggaa agaatcaagg aggcgcgcaa agagacctac gtcgagcagc acgagatggc 9840
 tgtggccaag tactgcgacc tccctagcaa actggggcac agataataac ctgcaggcgg 9900
 gccagatata cgcggttgaca ttgattattg actagttatt aatagtaatc aattacgggg 9960
 tcattagttc atagccata tatggagttc cgcggtacat aacttacggt aaatggccocg 10020
 cctggctgac cgcccaacga cccccgcca ttgacgtcaa taatgacgta tgttcccata 10080
 gtaacgcaa tagggacttt ccattgacgt caatgggtgg agtatttacg gtaaactgcc 10140
 cacttggcag tacatcaagt gtatcatatg ccaagtacgc cccctattga cgtcaatgac 10200
 ggtaaatggc ccgcctggca ttatgccag tacatgacct tatgggactt tcctacttgg 10260
 cagtacatct acgtattagt catcgtatt accatggtga tgcggttttg gcagtacatc 10320
 aatgggcgtg gatagcgggt tgactcacgg ggatttccaa gtctccacc cattgacgtc 10380
 aatgggagtt tgttttggca ccaaaatcaa cgggactttc caaaatgtcg taacaactcc 10440
 gccccattga cgc aaatggg cggtaggcgt gtacgggtgg aggtctatat aagcagcgg 10500
 aacccccggg taccggtcgc caccatggtg agcaagggcg aggagctggt caccgggggtg 10560

ES 2 663 424 T3

gtgcccaccc tggtcgagct ggacggcgac gtaaaccggcc acaagttcag cgtgtccggc 10620
gagggcgagg gcgatgccac ctacggcaag ctgacctga agttcatctg caccaccggc 10680
aagctgcccc tgccctggcc caccctcgtg accacctga cctacggcgt gcagtgttc 10740
agccgctacc cggaccacat gaagcagcac gacttcttca agtccgcat gcccgaggc 10800
tacgtccagg agcgcaccat cttcttcaag gacgacggca actacaagac ccgcgccgag 10860
gtgaagttcg agggcgacac cctggtgaac cgcacgagc tgaagggcat cgacttcaag 10920
gaggacggca acatcctggg gcacaagctg gagtacaact acaacagcca caacgtctat 10980
atcatggccg acaagcagaa gaacggcatc aaggtgaact tcaagatccg ccacaacatc 11040
gaggacggca gcgatgcagct cgccgaccac taccagcaga acaccccat cggcgacggc 11100
cccgtgtgc tgcccgacaa cactacctg agcaccagc cgcacctgag caaagacccc 11160
aacgagaagc gcgatcacat ggtcctgctg gagttcgtga ccgccgccgg gatcactctc 11220
ggcatggacg agctgtacaa gtaaagcggc cgcgactcta gagatctcta caggaggatc 11280
tcatgcatcg gatccgcaa aagtccgaa ttcatactgt gaagtacact gcatataagg 11340
agtgtggtat agtataaaga aactttctgc aggtagtaat tatagtgaag attttaggtt 11400
tacaagccc tagctgtttt ctgtgtagct tttattattc ttatgactct tgacaagttt 11460
gtagcttcac catatacatt taatattttg caataattgg cttgttctct gagctgttgg 11520
attcggggcc gtagcactgt ctgagagggt tacatttctc acagtgaacc ggtctctttt 11580
tcagctgctt cctggcttct ttttactcag gtttccactg cttttttgct ttttttaatg 11640
ctgtatgaag gtgttaacat ttgtttatat ttttcattaa ttgtaatacc tttaaatcat 11700
gcatcact cagaaatagg gattagaatt taagtacat ctttgccca attcctagcc 11760
ccgcagacac tagcgcacc cccgccccc gcggtgctga cgtcagcctg caagccccgc 11820
ccccgcgtct ccagggcaac cgtggctttc gattgttact gtgggaaccg gaggtaacag 11880
tctacagcca tggtcgcccc gcagcacgcc cacgctcccc accactcccg agttctgcca 11940
gcctgggttt gggcagatac agagcaagag gaggcggggc aattg 11985

5 <210> 2
<211> 45
<212> ADN
<213> Secuencia artificial

10 <220>
<223> Cebador directo

<400> 2
aatgatacgg cgaccaccga gatctacacc gctggcaagt gtagc

45

15 <210> 3
<211> 44

ES 2 663 424 T3

<212> ADN
 <213> Secuencia artificial

<220>
 5 <223> Cebador inverso de código de barras

<220>
 <221> misc_feature
 <222> (25)..(30)
 10 <223> n es a, c, g, o t

<400> 3
 caagcagaag acggcatacg agatnnnnn gggagggccc gttg 44

15 <210> 4
 <211> 45
 <212> ADN
 <213> Secuencia artificial

20 <220>
 <223> Cebador de secuenciación de Illumina

<400> 4
 25 ccacgaggat tcgaaaaggt gaaccgaccc ggtcgatgca ctagt 45

<210> 5
 <211> 32
 <212> ADN
 <213> Secuencia artificial

30 <220>
 <223> Cebador de indexación de Illumina

<400> 5
 35 cctaggagca agtacgaaca acgggcctc cc 32

<210> 6
 <211> 300
 <212> ADN
 40 <213> Secuencia artificial

<220>
 <223> Precursor de miR-124

45 <400> 6
 aggtgggagt actgctcaga gctacaactc taggagtagg gactccaagc ctagagctcc 60
 aagagagggt gaagggcagg gagaaaatta tagtaatagt tgcaatgagt cacttgcttc 120
 tagatcaaga tcagagactc tgctctccgt gttcacagcg gaccttgatt taatgtcata 180
 caattaaggc acgcggtgaa tgccaagagc ggagcctaca gctgcacttg aaggacatcc 240
 gagagaagtt aggaaggggtg gggagaaaca attctagaat gaacccatcc tgtgcgacac 300

<210> 7
 <211> 435
 50 <212> ADN
 <213> Secuencia artificial

<220>
 <223> Precursor de miR-128

55 <400> 7

ES 2 663 424 T3

	atactgtgaa gtacactgca tataaggagt gtggtatagt ataaagaaac tttctgcagg	60
	tagtaattat agtgaagatt ttaggtttac aaagccctag ctgttttctg tgtagctttt	120
	attattctta tgactcttga caagtttgta gottcaccat atacatttaa tttttgcaa	180
	taattggcct tgttcctgag ctggttgatt cggggccgta gcactgtctg agaggtttac	240
	atttctcaca gtgaaccggt ctctttttca gctgcttcct ggcttctttt tactcaggtt	300
	tccactgctt ttttgctttt tttaatgctg tatgaagggtg ttaacatttg tttatatattt	360
	tcattaattg taataccttt aatcatgca tcatactcag aatagggat tagaatttaa	420
	gtgacatctt tggcc	435
5	<210> 8 <211> 225 <212> ADN <213> Secuencia artificial	
10	<220> <223> Precursor de miR-132	
	<400> 8 ctagccccgc agacactagc gccacccccg ccgcccgcgg tgctgacgtc agcctgcaag	60
	ccccgcccc gcgtctccag ggcaaccgtg gctttcgatt gttactgtgg gaaccggagg	120
	taacagtcta cagccatggt cgccccgcag cacgcccacg ctccccacca ctcccagatt	180
	ctgccagcct gggtttgggc agatacagag caagaggagg cgggg	225
15	<210> 9 <211> 126 <212> ADN <213> Secuencia artificial	
20	<220> <223> Sensor de miR-124	
	<400> 9 ggcacagata ataacctgca aaaaggcatt caccgcgtgc cttaggcatt caccgcgtgc	60
	cttaggcatt caccgcgtgc cttaggcatt caccgcgtgc cttaaagca ggcgggcccag	120
	atatac	126
25	<210> 10 <211> 127 <212> ADN <213> Secuencia artificial	
30	<220> <223> Sensor de miR-128	
	<400> 10 ggcacagata ataacctgca aaaagagacc ggttcactgt gaaaagagac cggttcactg	60
	tgaaaagaga ccggttcact gtgaaaagag accggttcac tgtgaaatgc aggcgggcca	120
35	gatatac	127

ES 2 663 424 T3

<210> 11
<211> 134
<212> ADN
<213> Secuencia artificial

5

<220>
<223> Sensor de miR-132

<400> 11
ggcacagata ataacctgca aaaacgacca tggctgtaga ctggttacgac catggctgta 60

gactgttacg accatggctg tagactgta cgaccatggc ttagactgt taaatgcagg 120

10 cgggccagat atac 134

REIVINDICACIONES

1. Constructo genético que comprende:
- 5 un elemento de ADN;
- un primer elemento de extremos compatibles y un segundo elemento de extremos compatibles que flanquean al elemento de ADN, en el que los elementos de extremos compatibles primero y segundo son capaces de aparearse entre sí;
- 10 un elemento de código de barras;
- un tercer elemento de extremos compatibles y un cuarto elemento de extremos compatibles que flanquean al elemento de código de barras, en el que los elementos de extremos compatibles tercero y cuarto son capaces de aparearse entre sí pero no son capaces de aparearse a los elementos de extremos compatibles primero o segundo; y
- 15 un sitio de separación ubicado entre el cuarto elemento de extremos compatibles y el primer elemento de extremos compatibles, en el que el elemento de ADN, primer elemento de extremos compatibles y segundo elemento de extremos compatibles están en un lado del sitio de separación, y el elemento de código de barras, tercer elemento de extremos compatibles y cuarto elemento de extremos compatibles están en el otro lado del sitio de separación.
2. Constructo genético que comprende:
- 25 una pluralidad de elementos de ADN;
- un primer elemento de extremos compatibles y un segundo elemento de extremos compatibles que flanquean a la pluralidad de elementos de ADN, en el que los elementos de extremos compatibles primero y segundo son capaces de aparearse entre sí;
- 30 una pluralidad de elementos de código de barras;
- un tercer elemento de extremos compatibles y un cuarto elemento de extremos compatibles que flanquean a la pluralidad de elementos de código de barras, en el que los elementos de extremos compatibles tercero y cuarto son capaces de aparearse entre sí pero no son capaces de aparearse a los elementos de extremos compatibles primero o segundo; y
- 35 un sitio de separación ubicado entre la pluralidad de elementos de ADN y la pluralidad de elementos de código de barras.
3. Un método para generar un constructo genético combinatorio, que comprende:
- 45 proporcionar un vector que contiene un primer constructo genético según la reivindicación 1 o la reivindicación 2;
- escindir el vector en el sitio de separación dentro del primer constructo genético, dando como resultado que el primer constructo genético se separe en segmentos primero y segundo;
- 50 proporcionar un segundo constructo genético según la reivindicación 1 o la reivindicación 2; y aparear el segundo constructo genético al vector escindido, en el que el apareamiento se produce en elementos de extremos compatibles dentro de los constructos genéticos primero y segundo que son capaces de aparearse entre sí, y en el que tras el apareamiento, el segundo constructo genético se integra entre los segmentos primero y segundo del primer constructo genético, creando un constructo genético combinatorio.
- 55 4. Método según la reivindicación 3, en el que el método es iterativo.
5. Constructo genético combinatorio producido mediante el método según la reivindicación 3 o la reivindicación 4.
- 60 6. Método para la identificación de un elemento de ADN o una pluralidad de elementos de ADN, que comprende:
- 65 proporcionar un constructo genético según la reivindicación 1 o la reivindicación 2;
- realizar un ensayo para determinar la secuencia de ADN del código de barras o la pluralidad de códigos de

barras dentro del constructo genético y/o la secuencia de ADN del elemento de ADN o la pluralidad de elementos de ADN dentro del constructo genético; e

identificar el elemento de ADN o la pluralidad de elementos de ADN.

5

7. Colección que comprende:

dos o más constructos genéticos según la reivindicación 1 o la reivindicación 2.

Figura 1

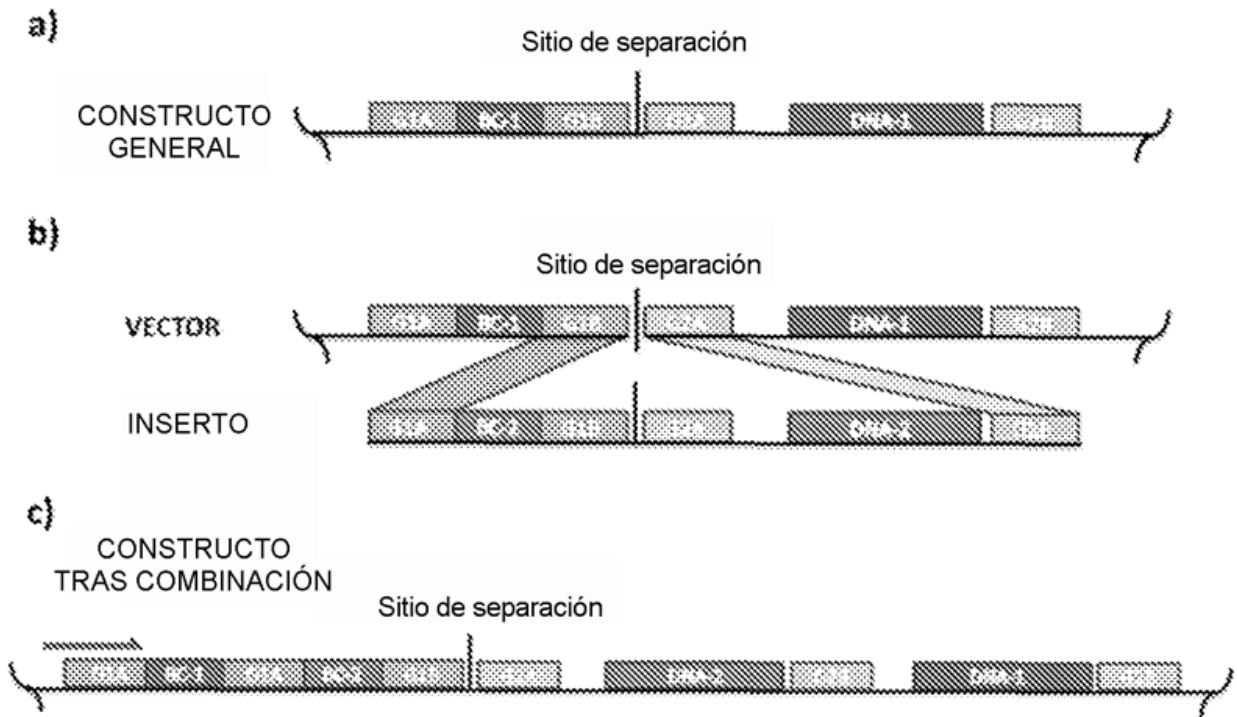


Figura 2

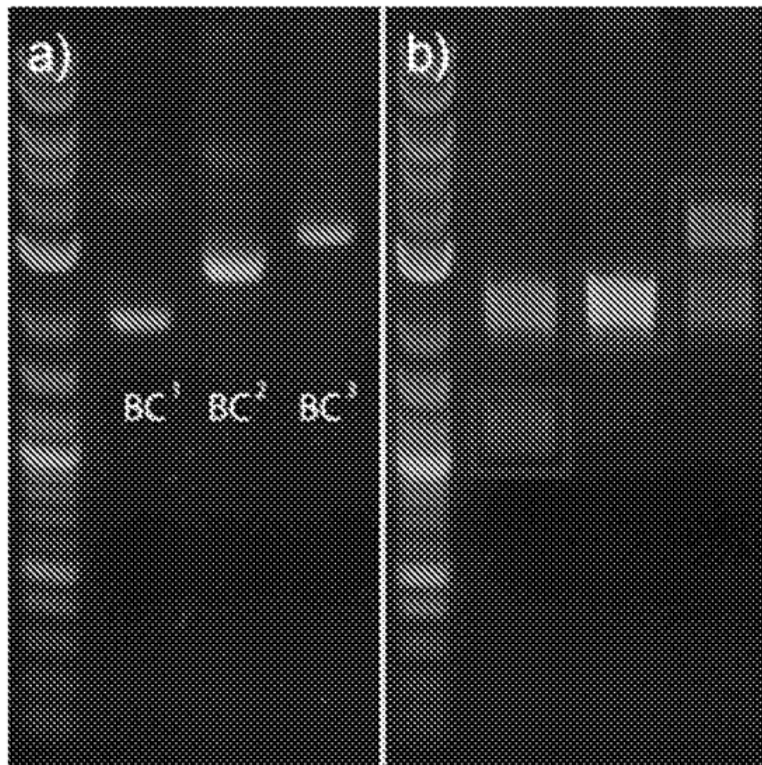


Figura 3

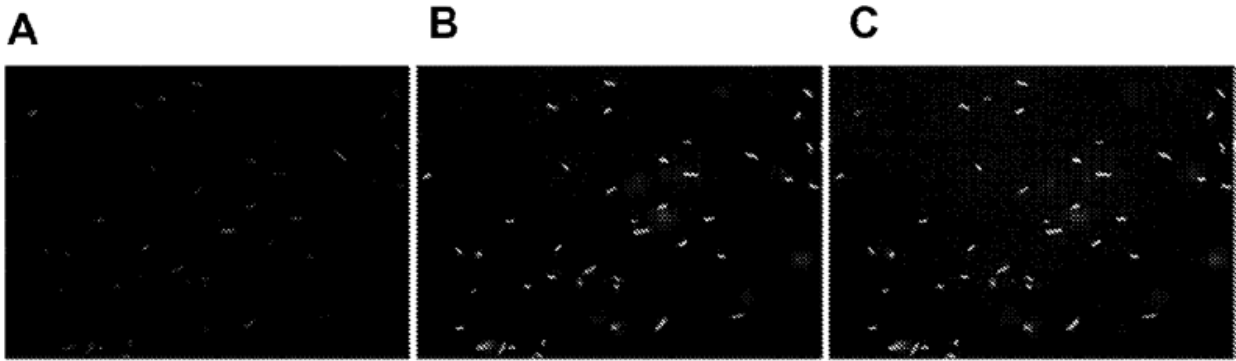


Figura 5

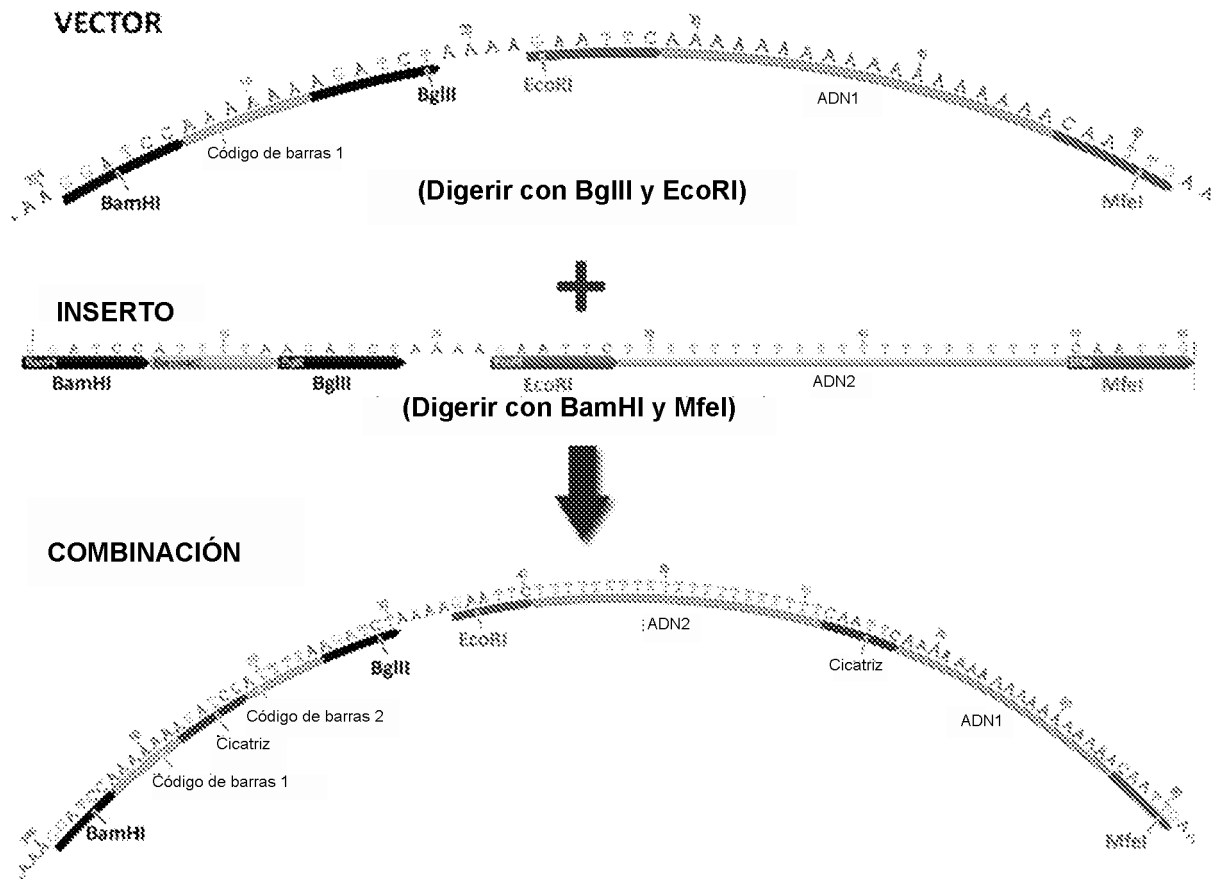


Figura 7

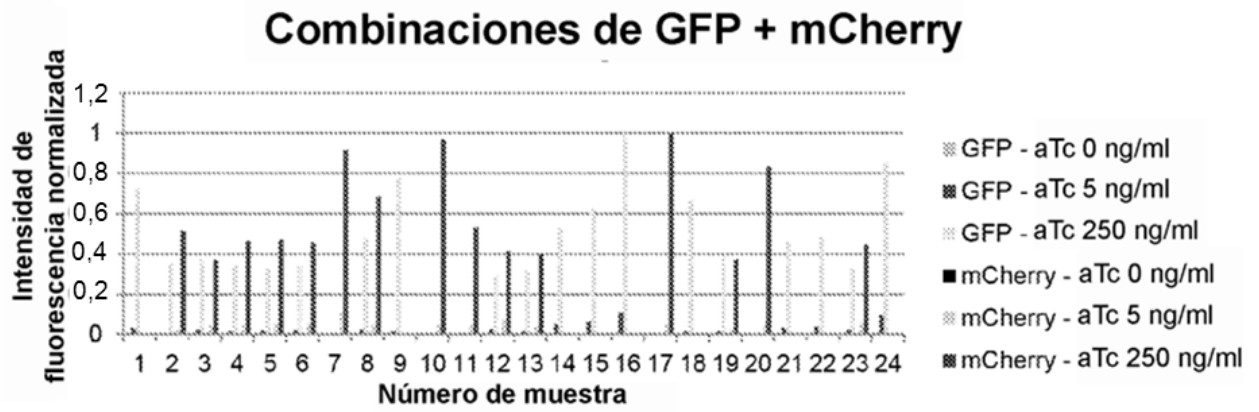


Figura 8

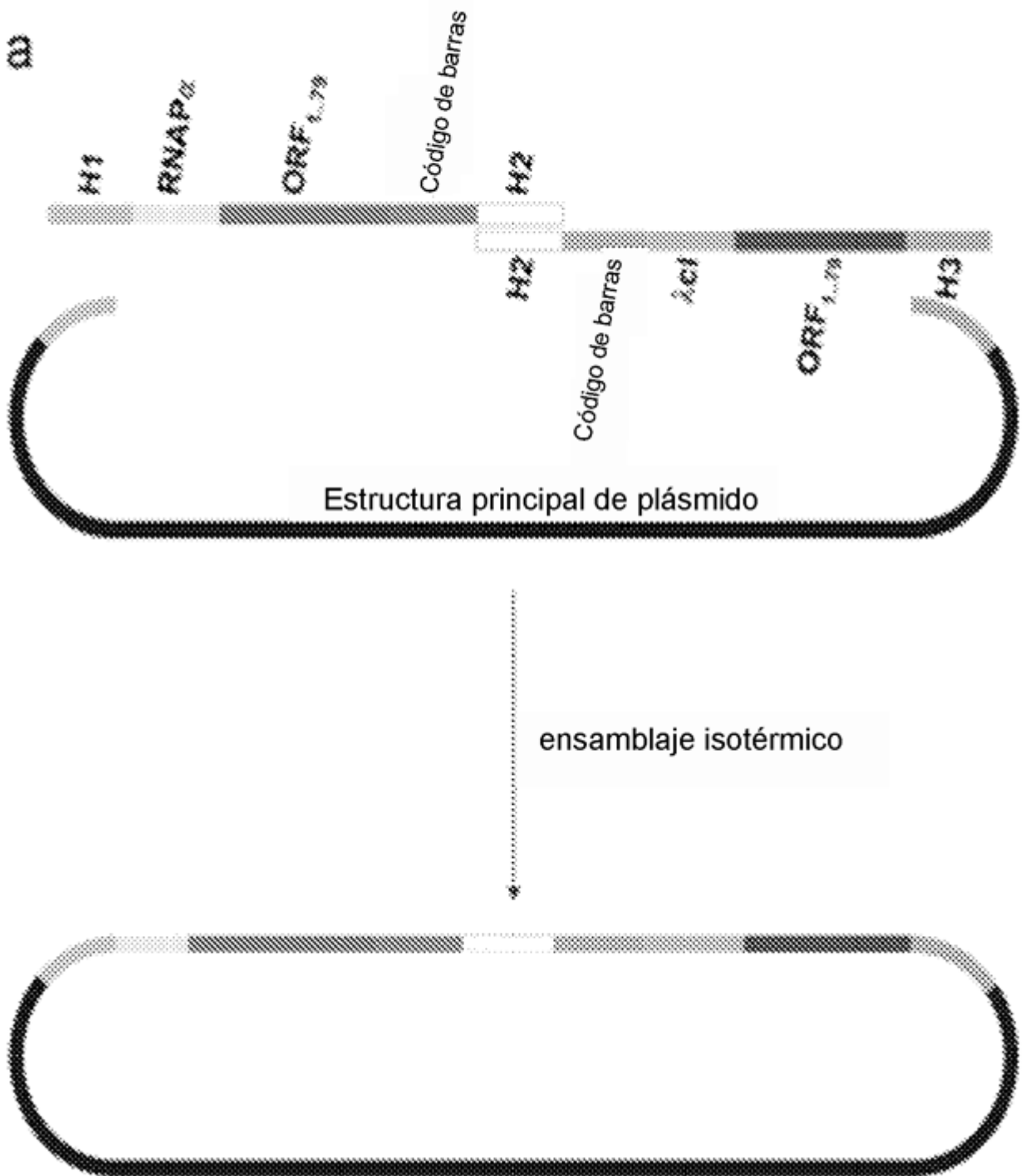
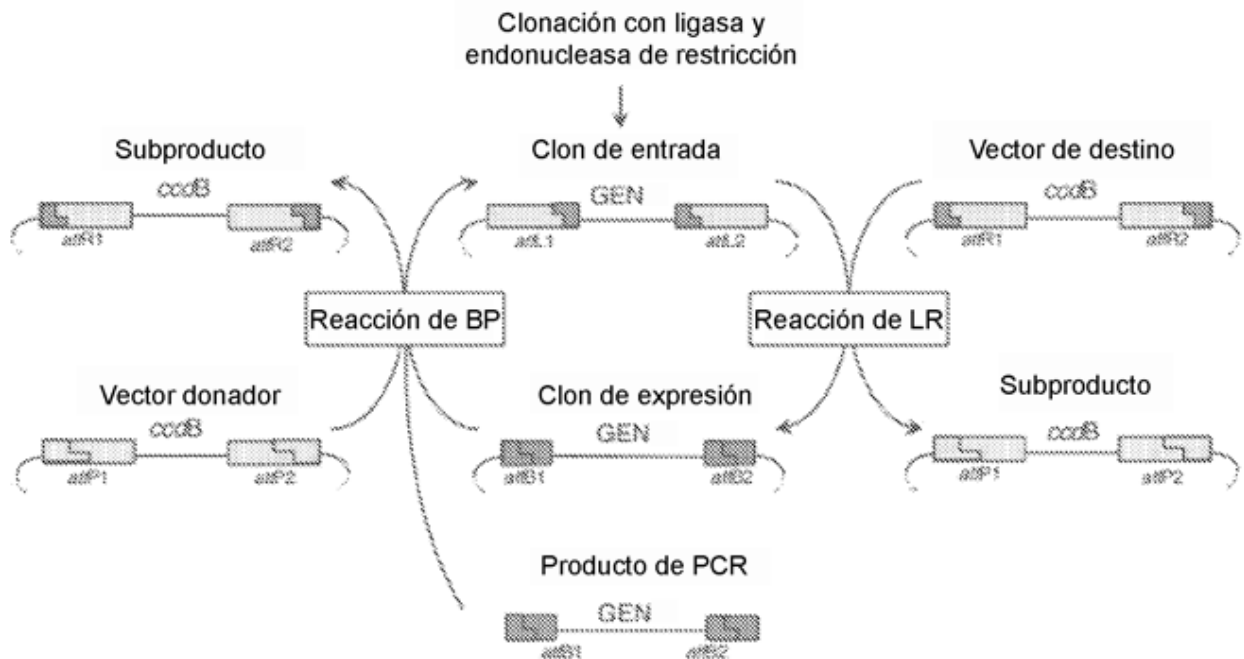


Figura 9



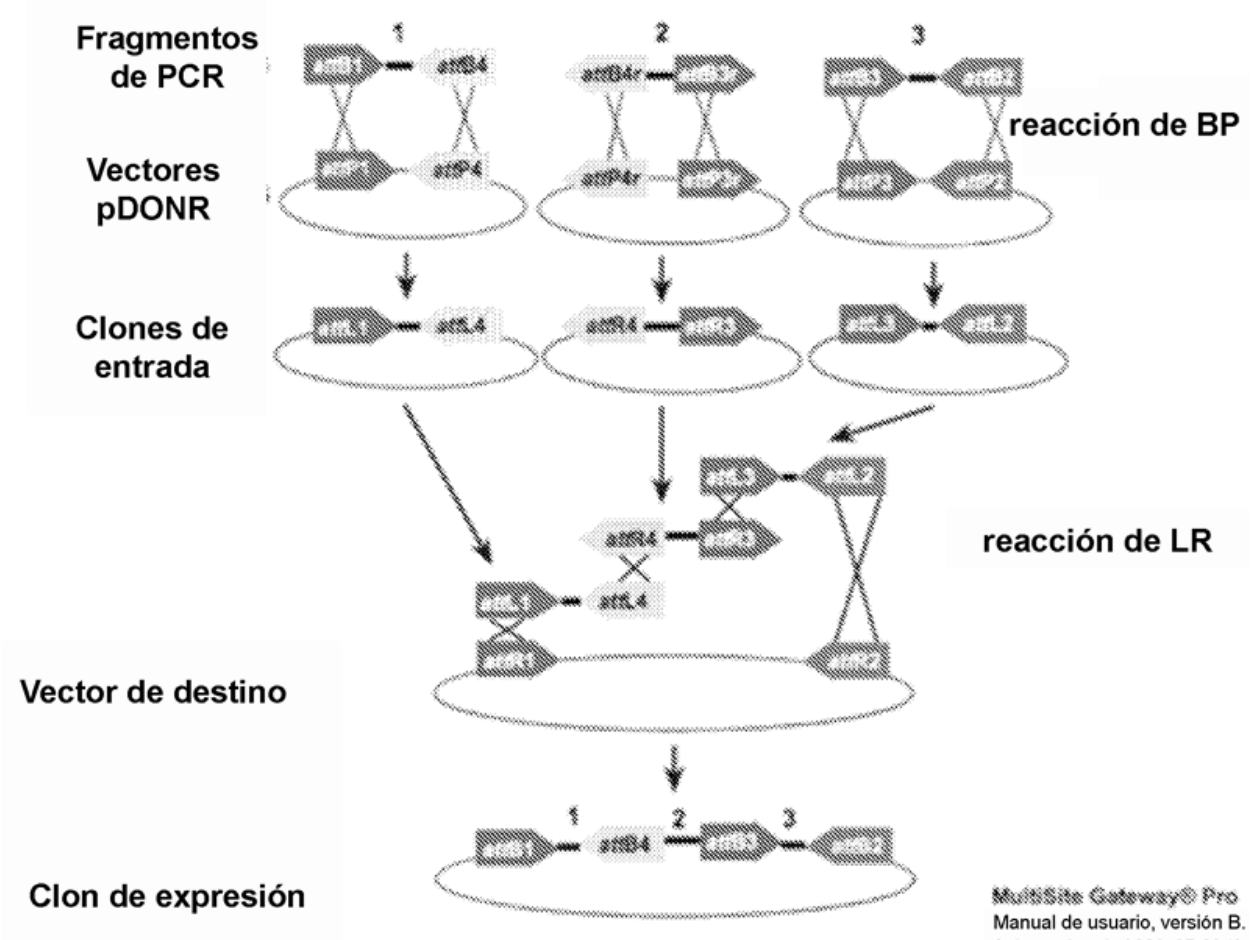
Clave:

B1 __ * __ B2 + P1 _____ P2

L1 __ * __ L2 + R1 _____ R2

<http://wolfson.huji.ac.il/expression/gatewayman.pdf>

Figura 10



Multisite Gateway® Pro
Manual de usuario, versión B.
3 de octubre de 2006, 25-0942

Figura 11

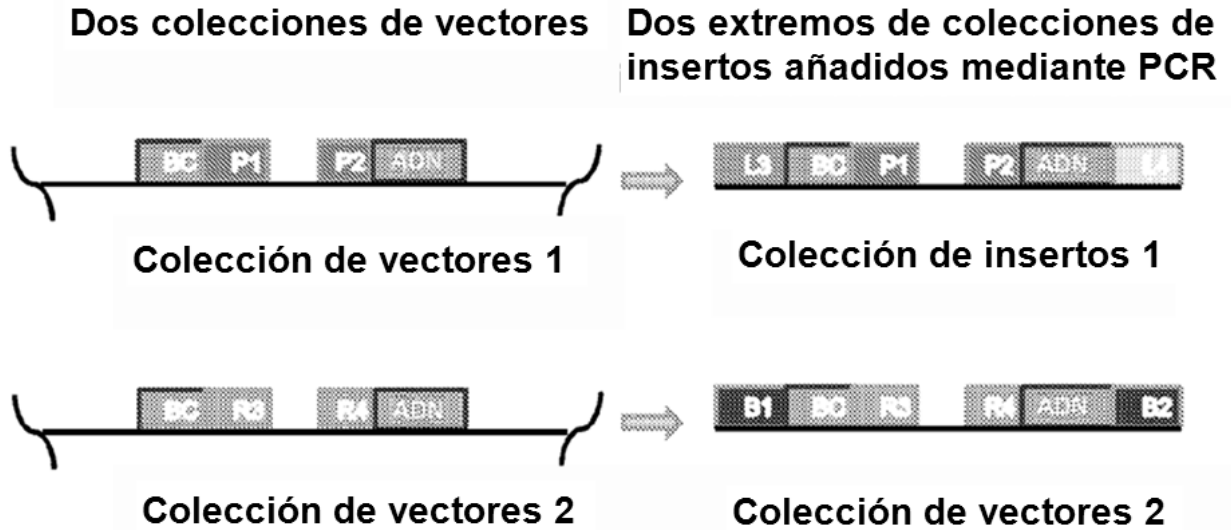


Figura 12

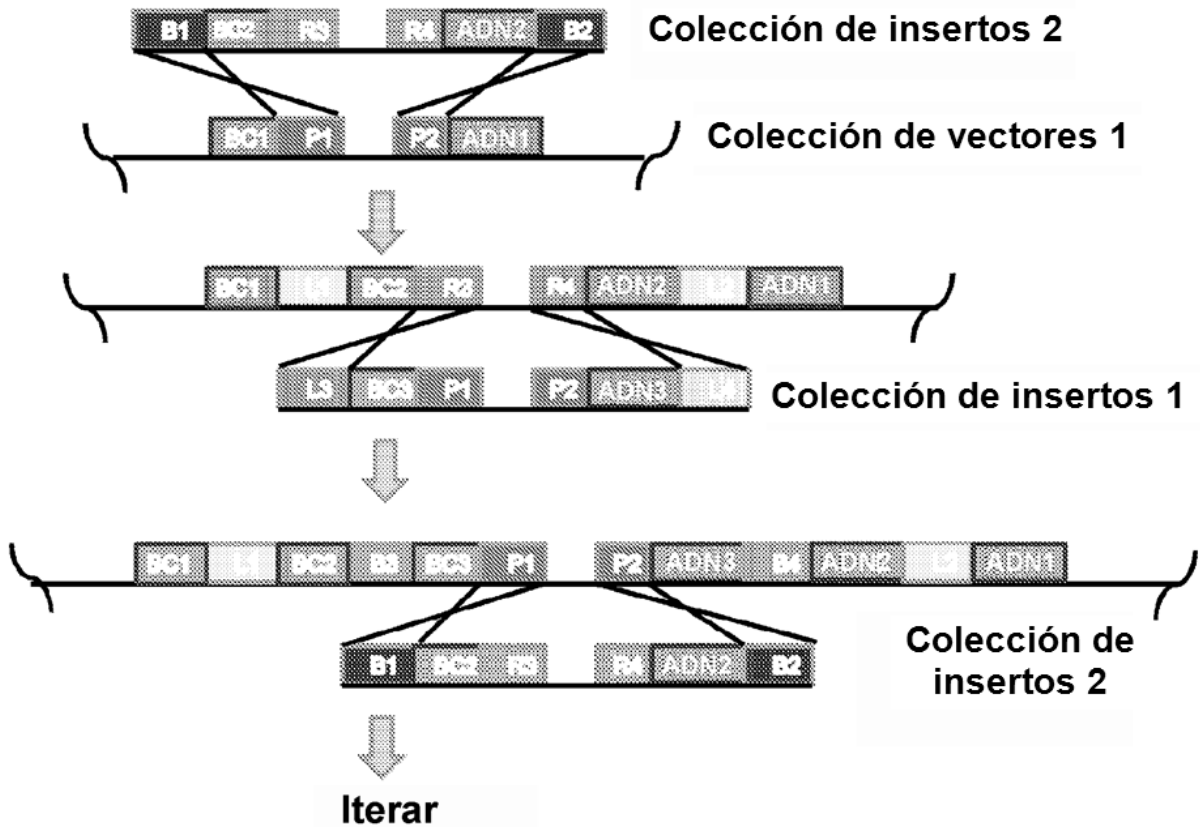


Figura 13

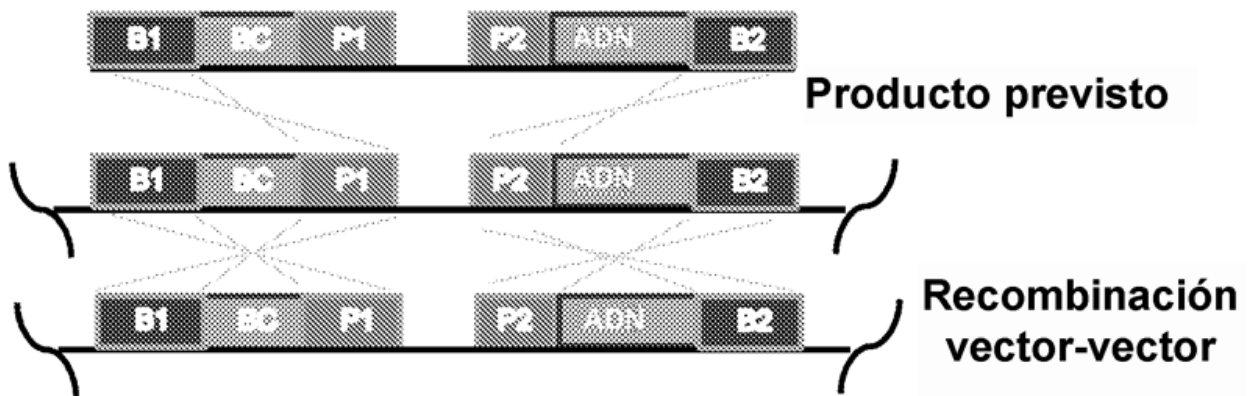


Figura 14

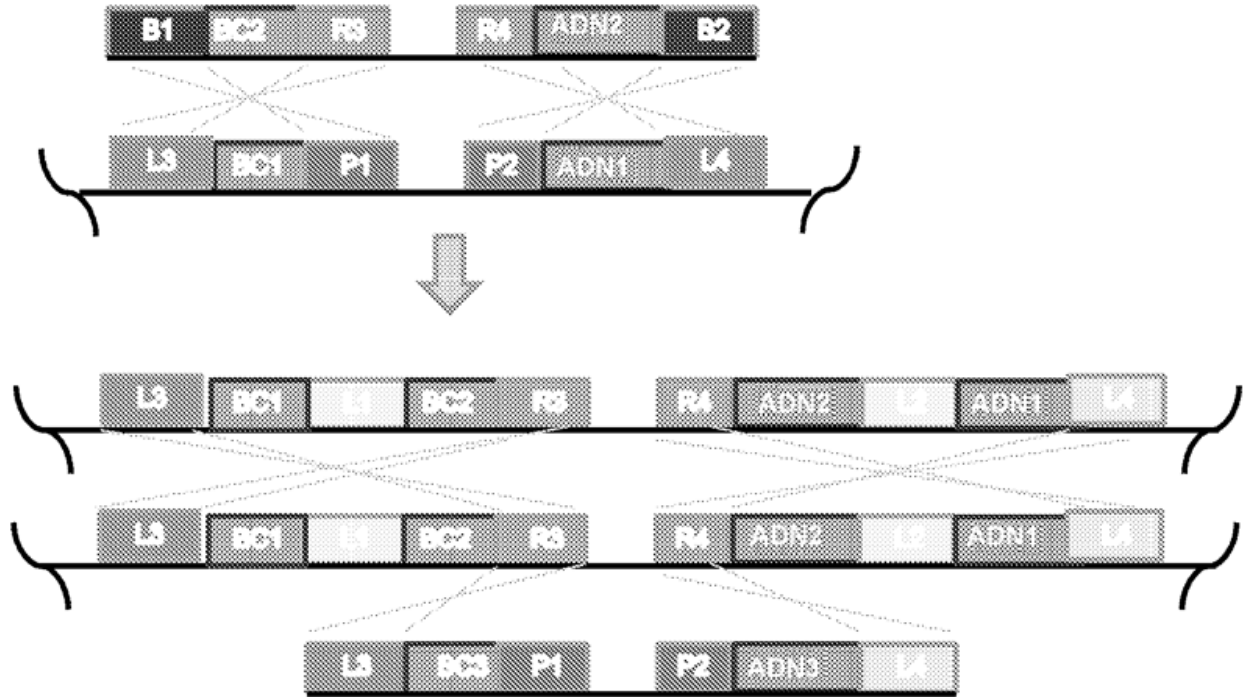


Figura 15

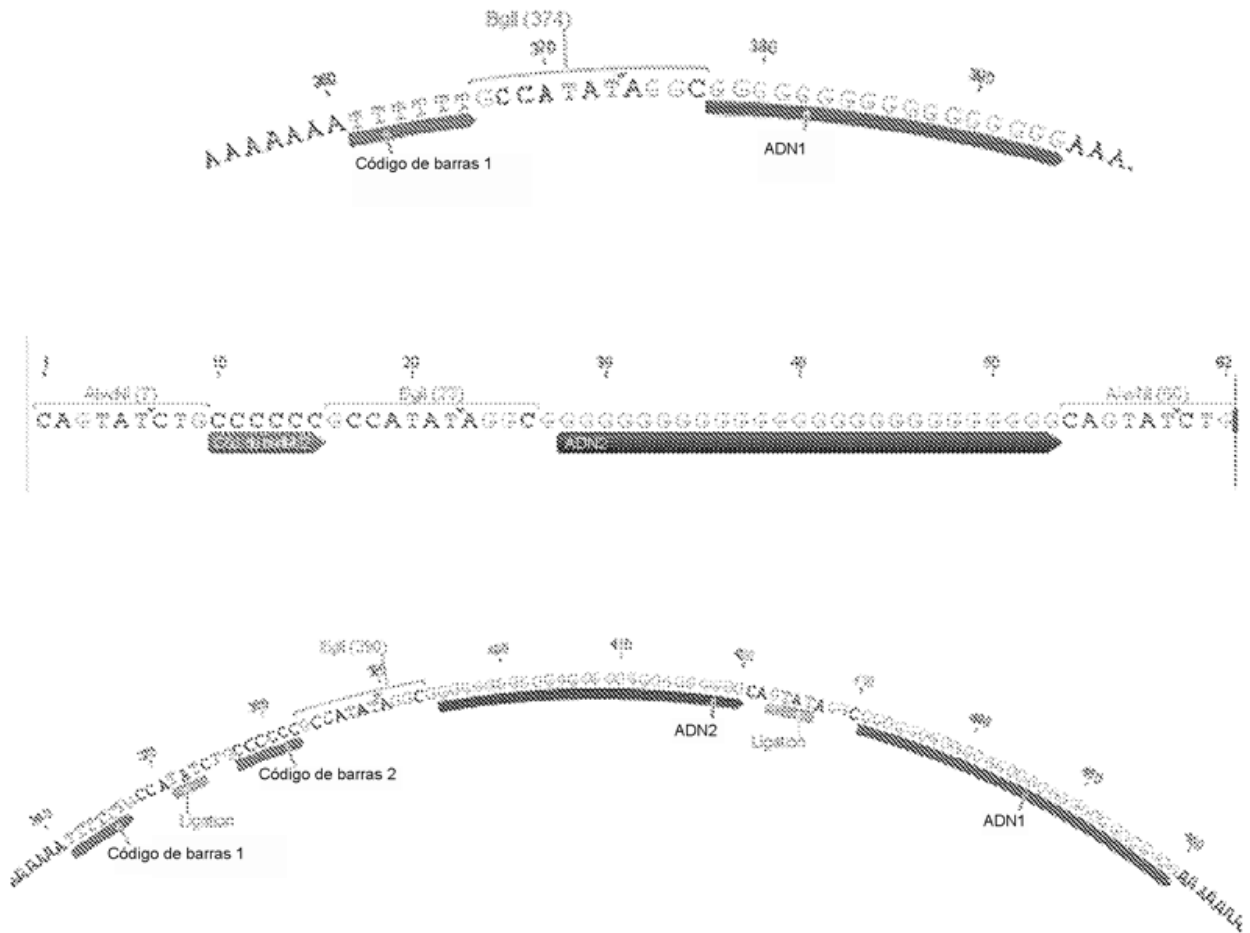


Figura 16

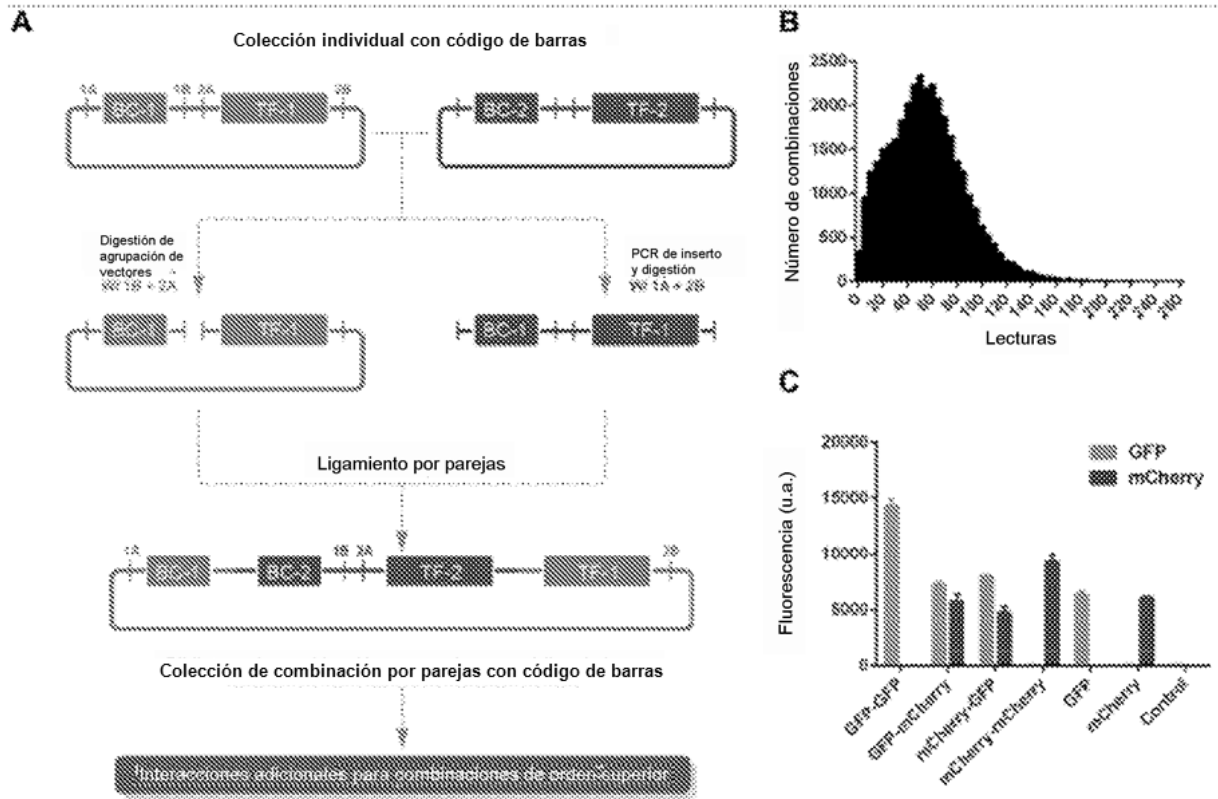


Figura 17

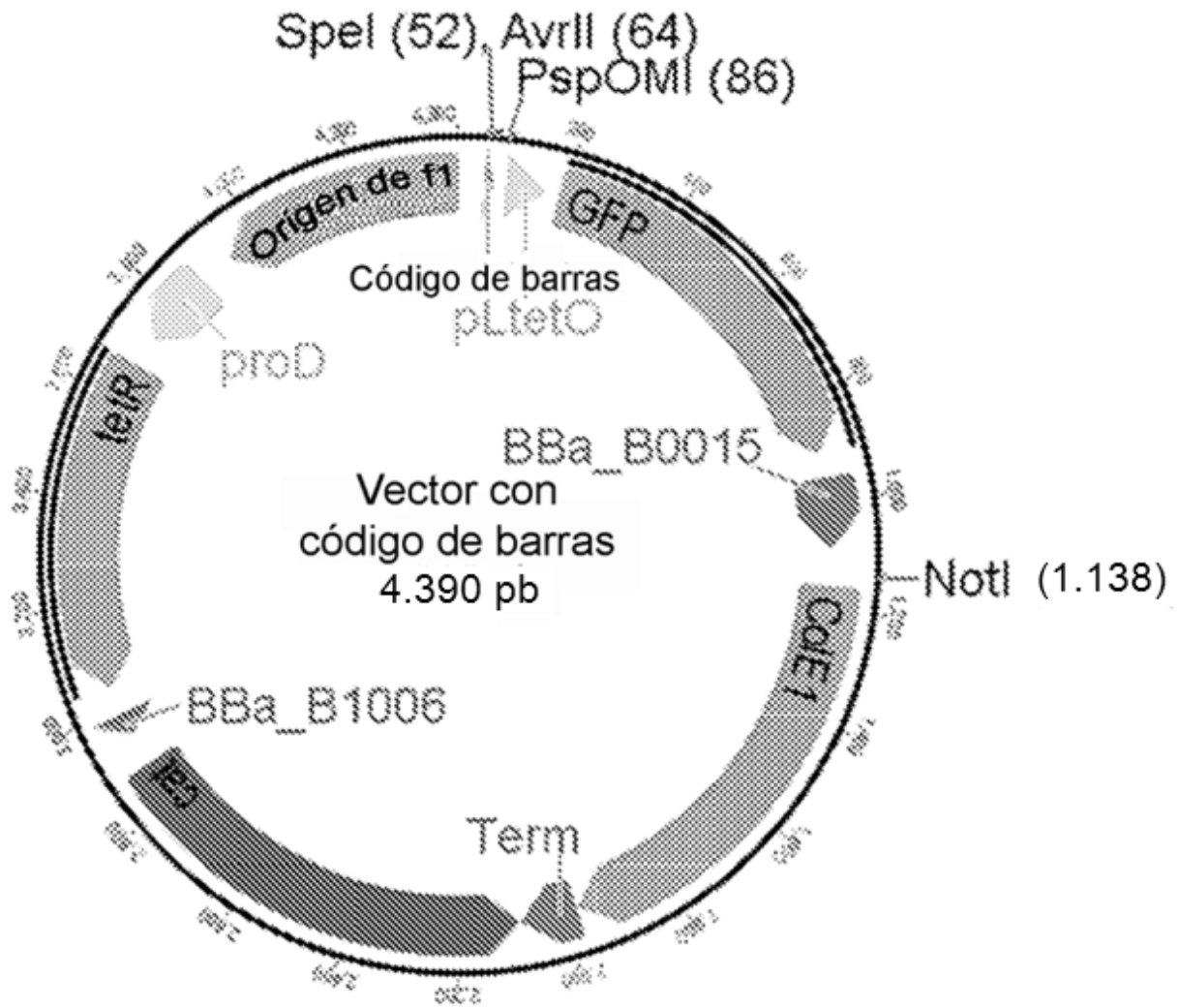


Figura 18

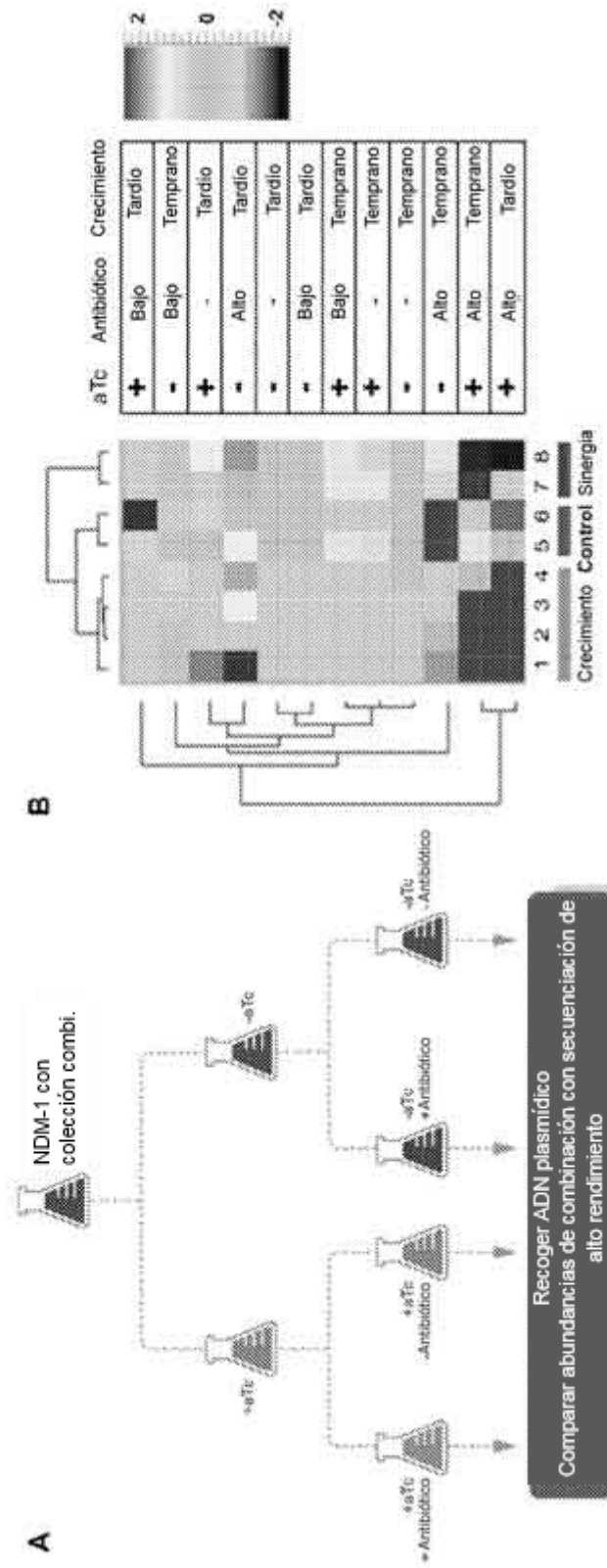


Figura 19

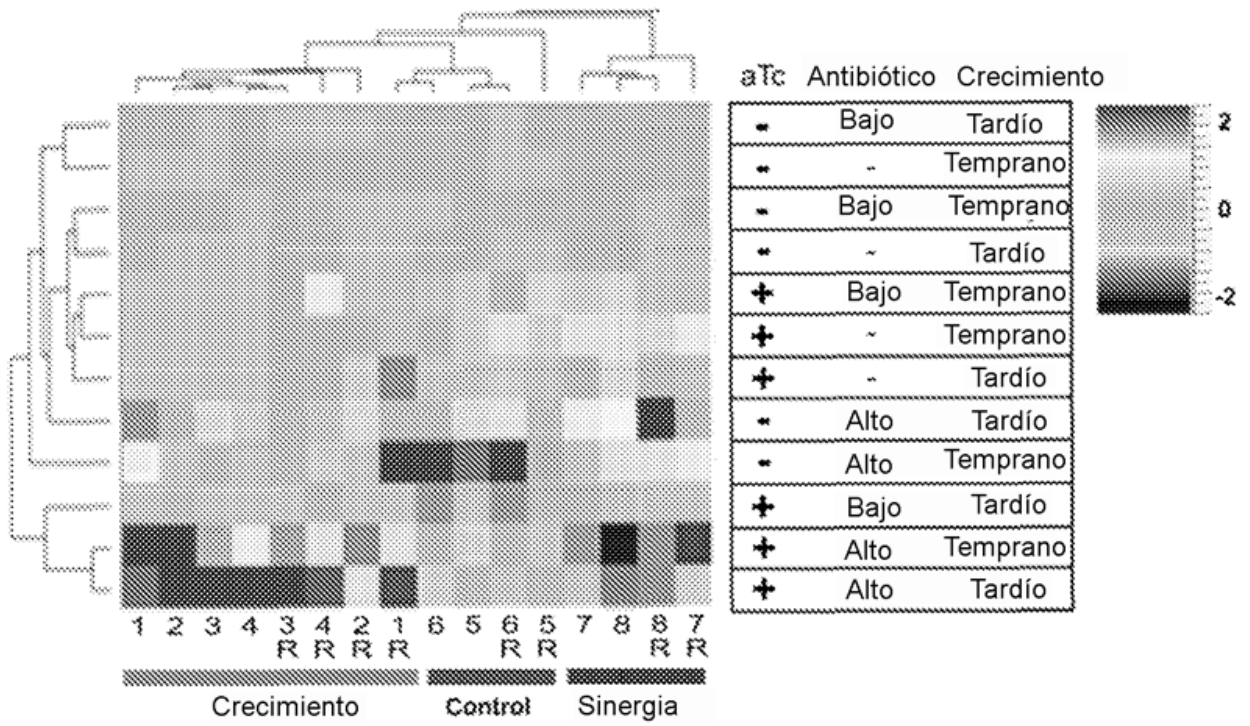


Figura 20

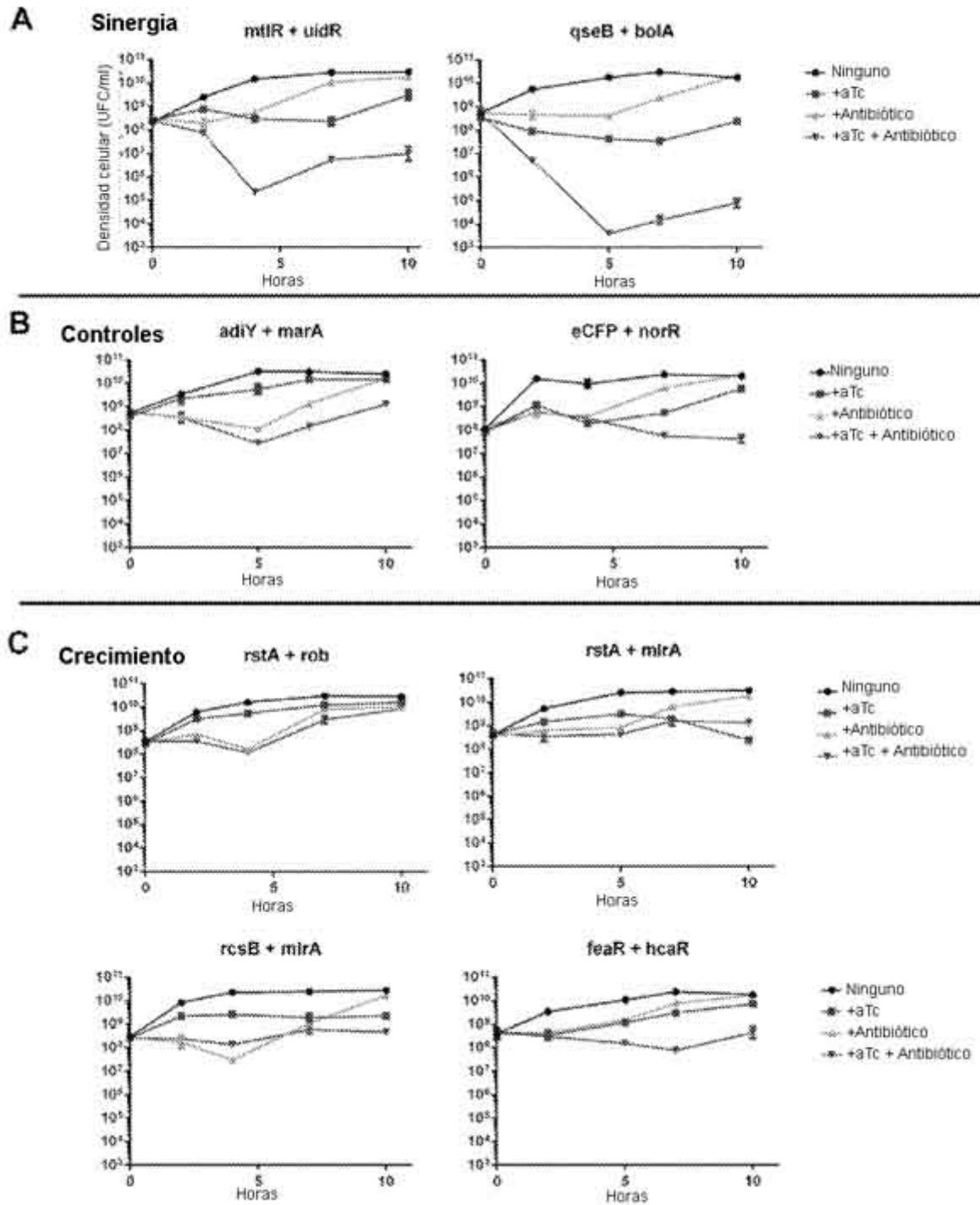


Figura 21

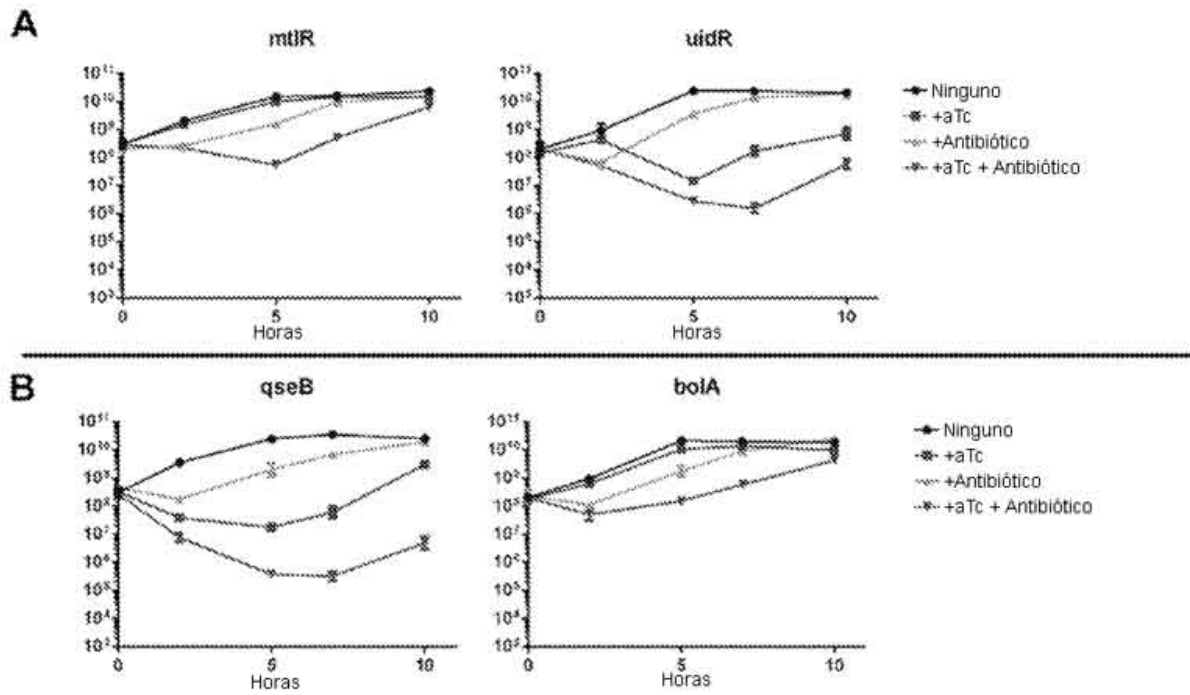


Figura 22

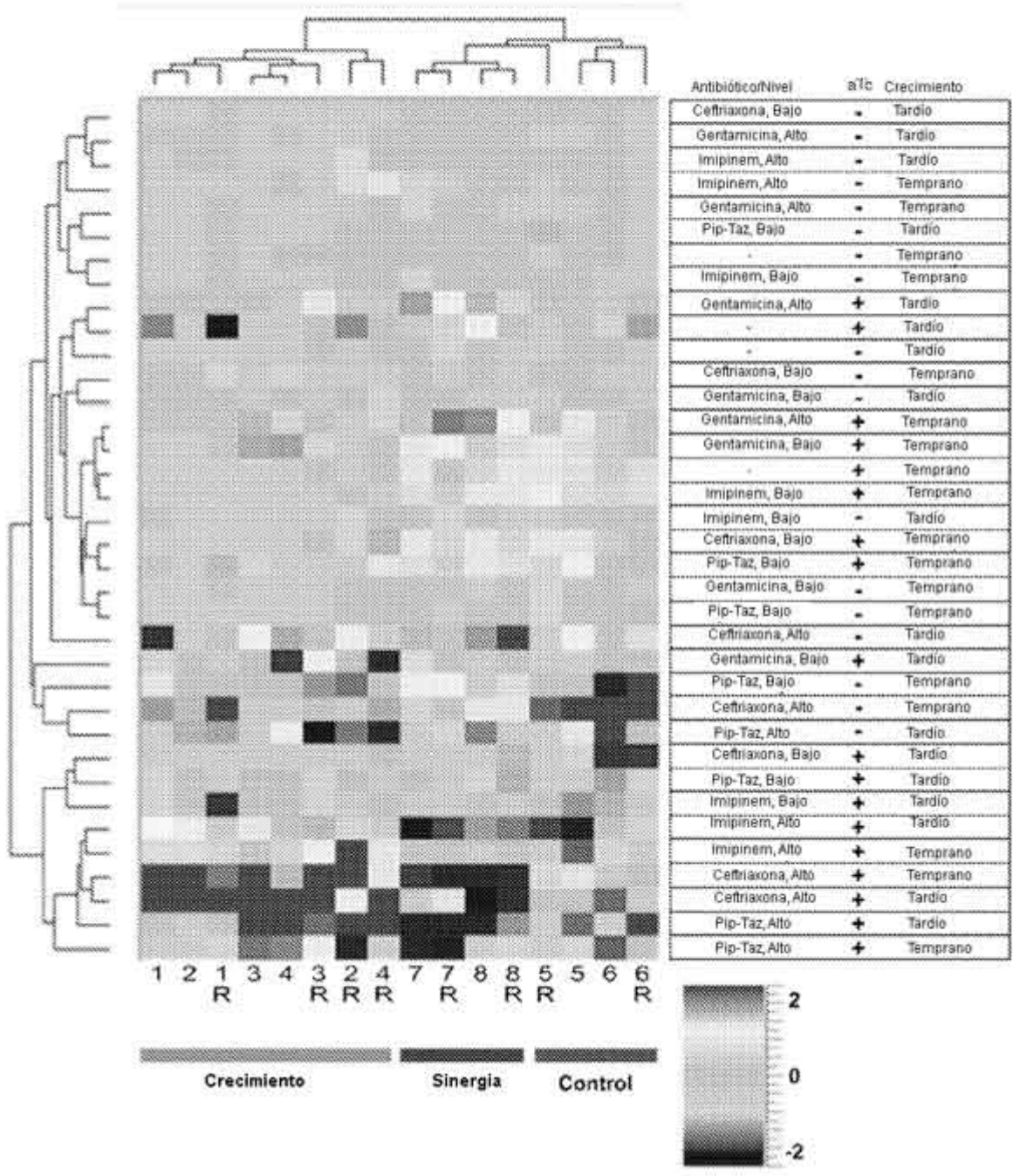


Figura 23

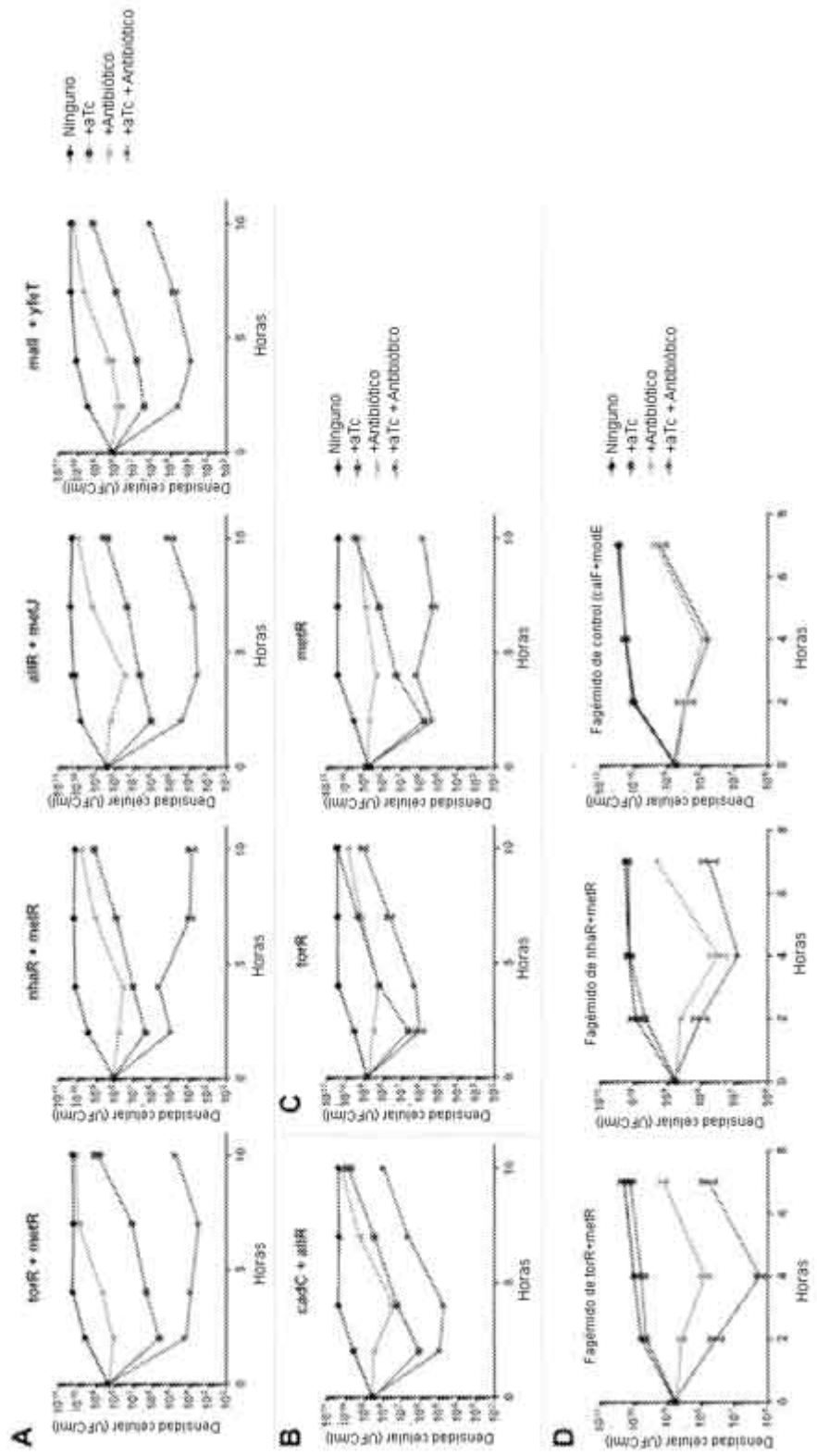


Figura 24

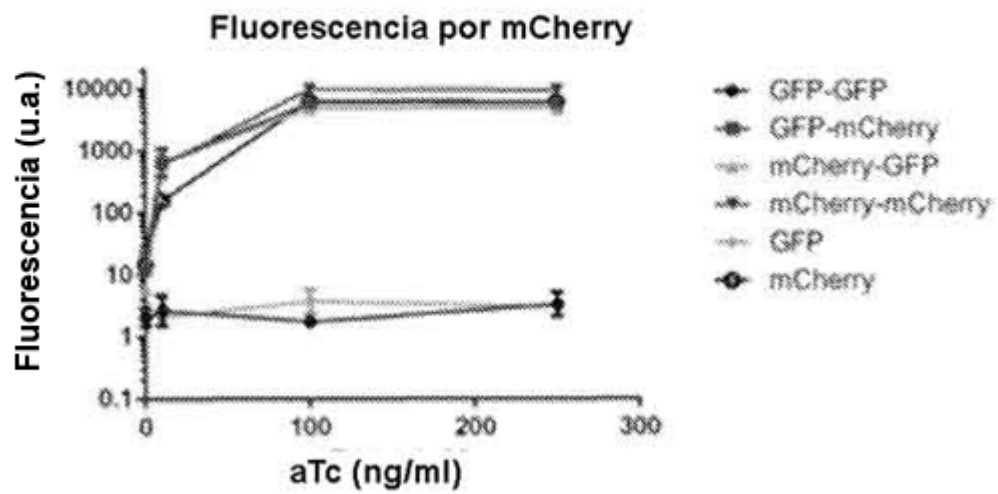
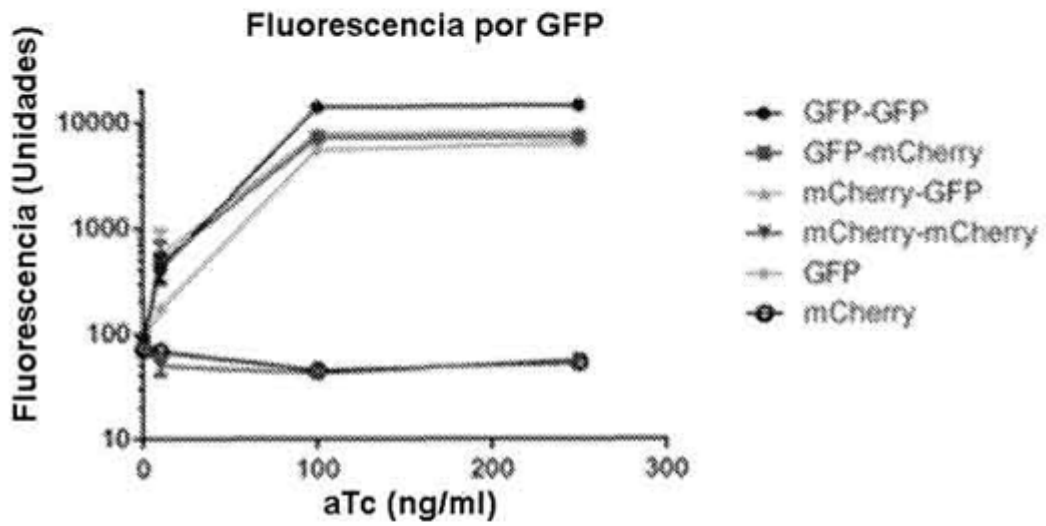


Figura 25

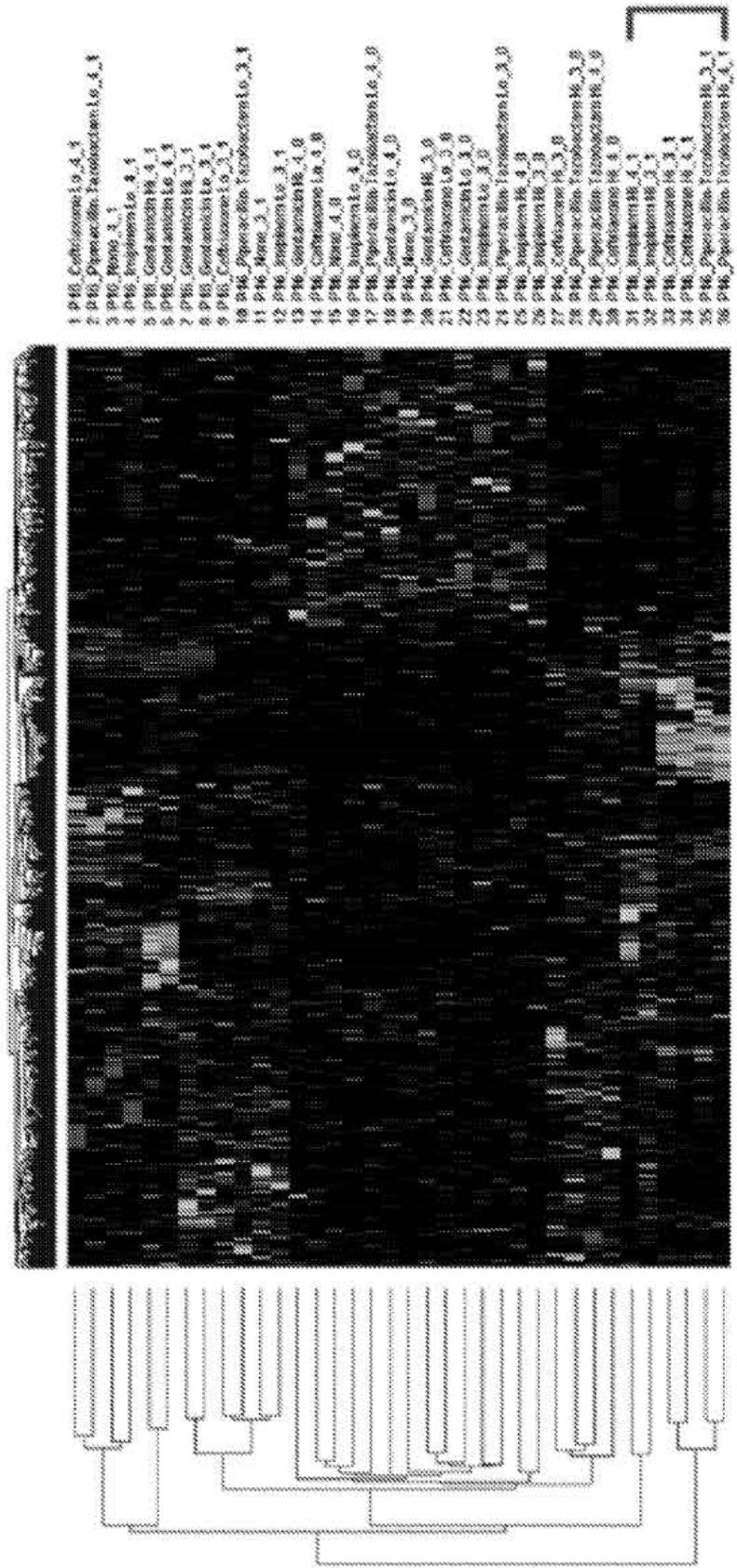


Figura 26

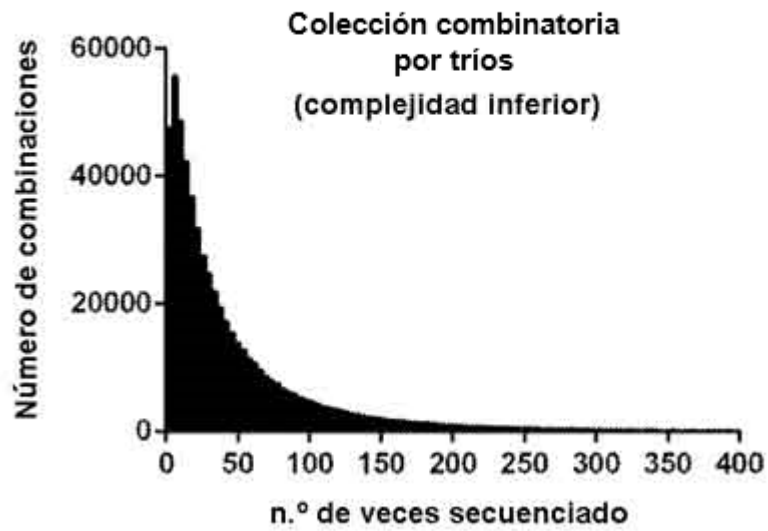


Figura 27

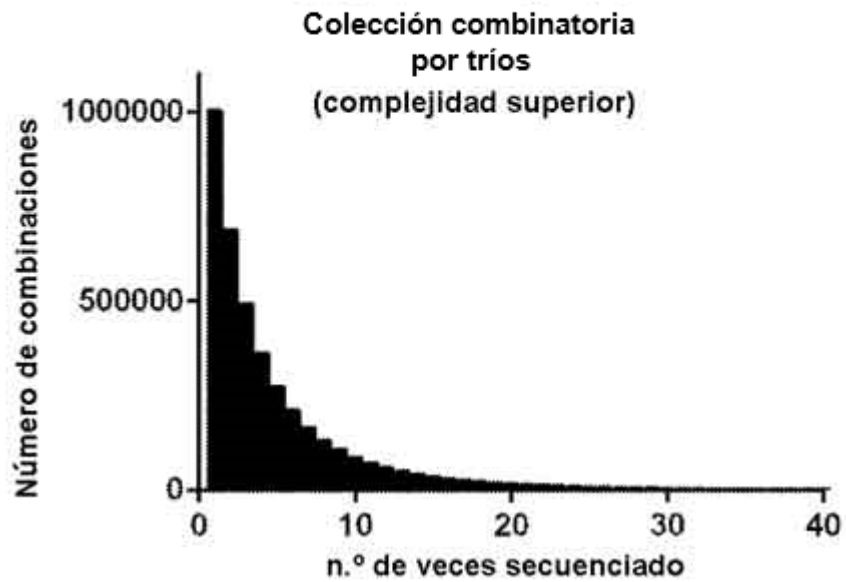


Figura 28

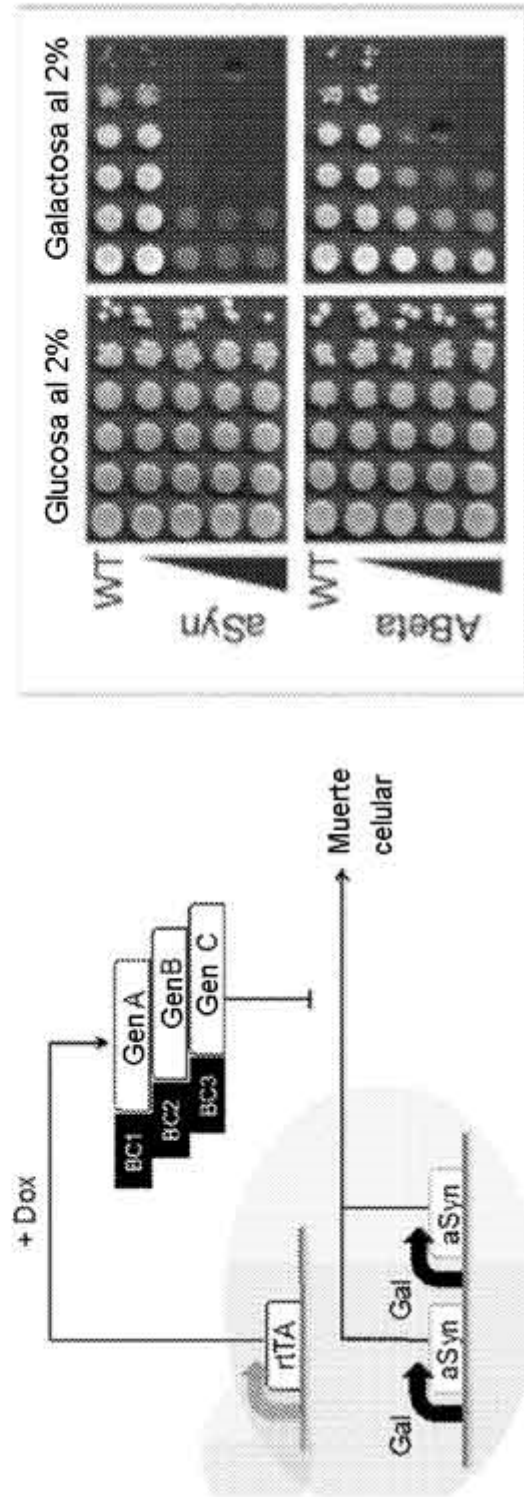


Figura 29

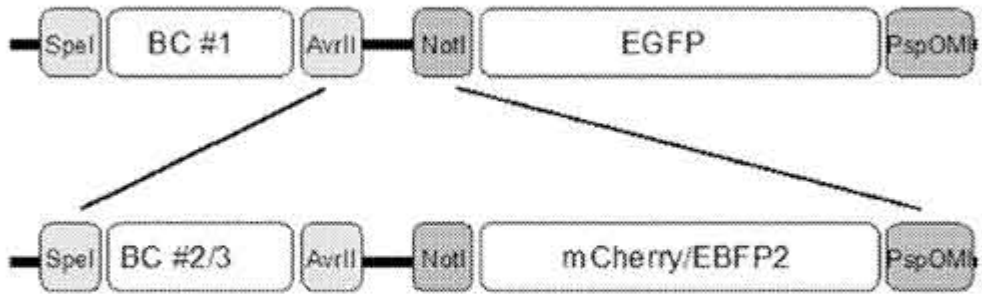


Figura 30

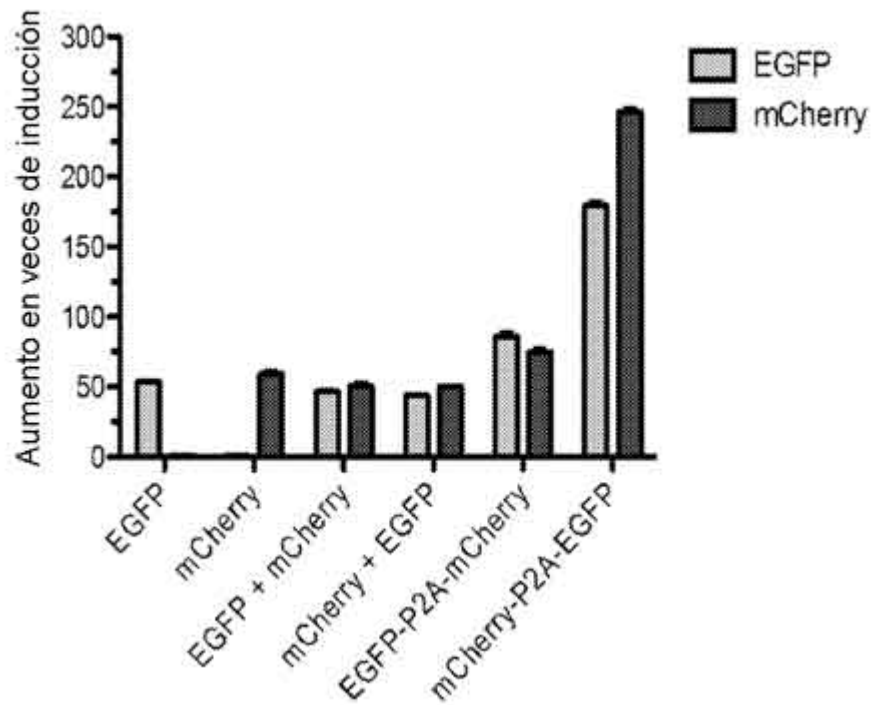


Figura 31

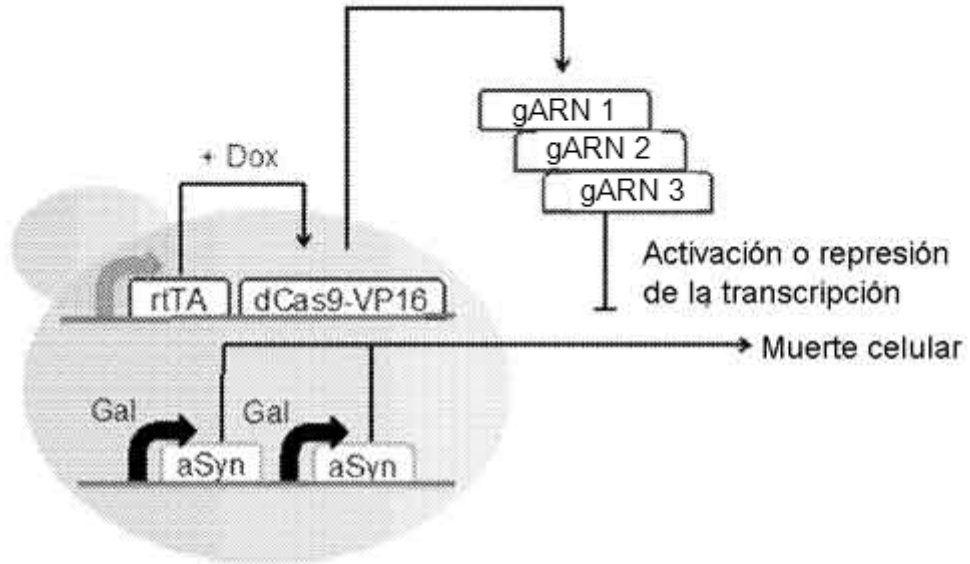


Figura 32

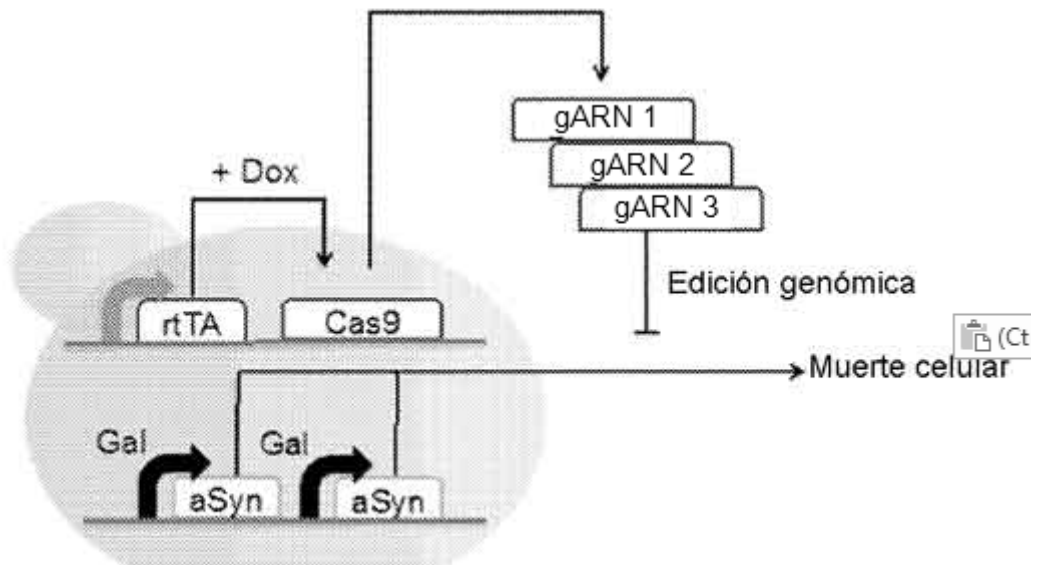


Figura 33

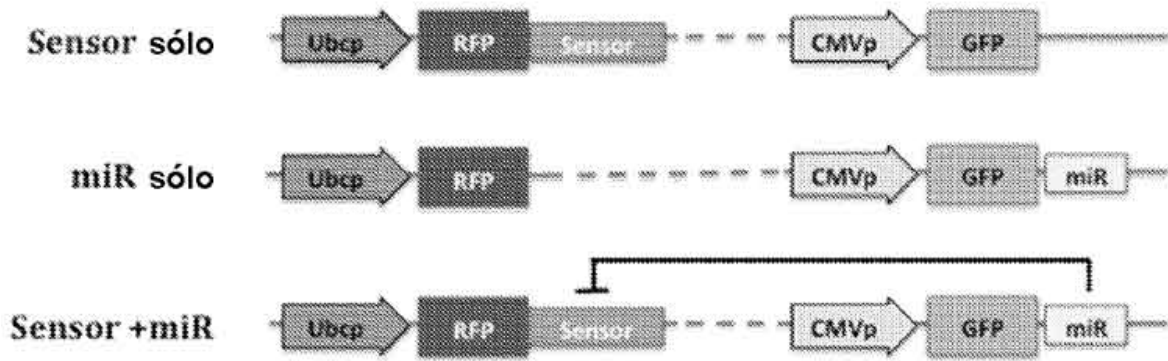


Figura 34

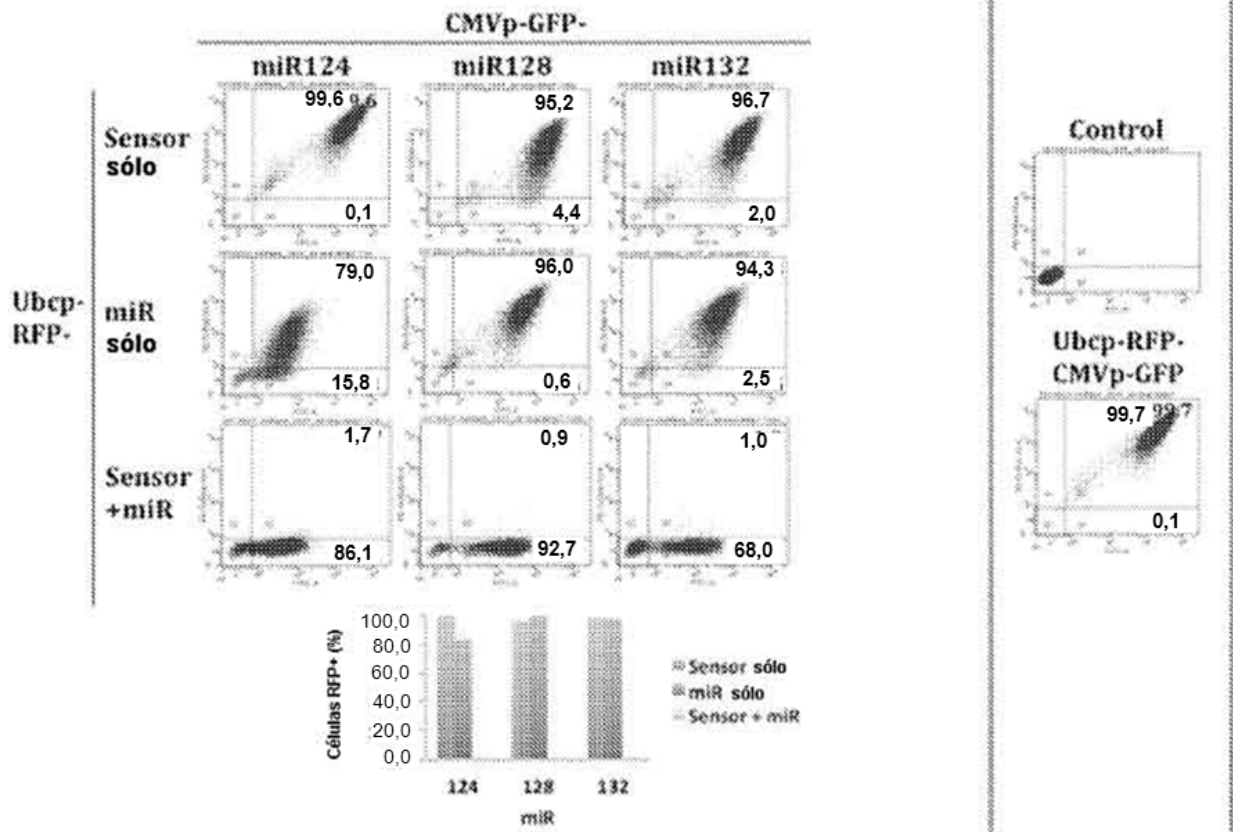


Figura 35

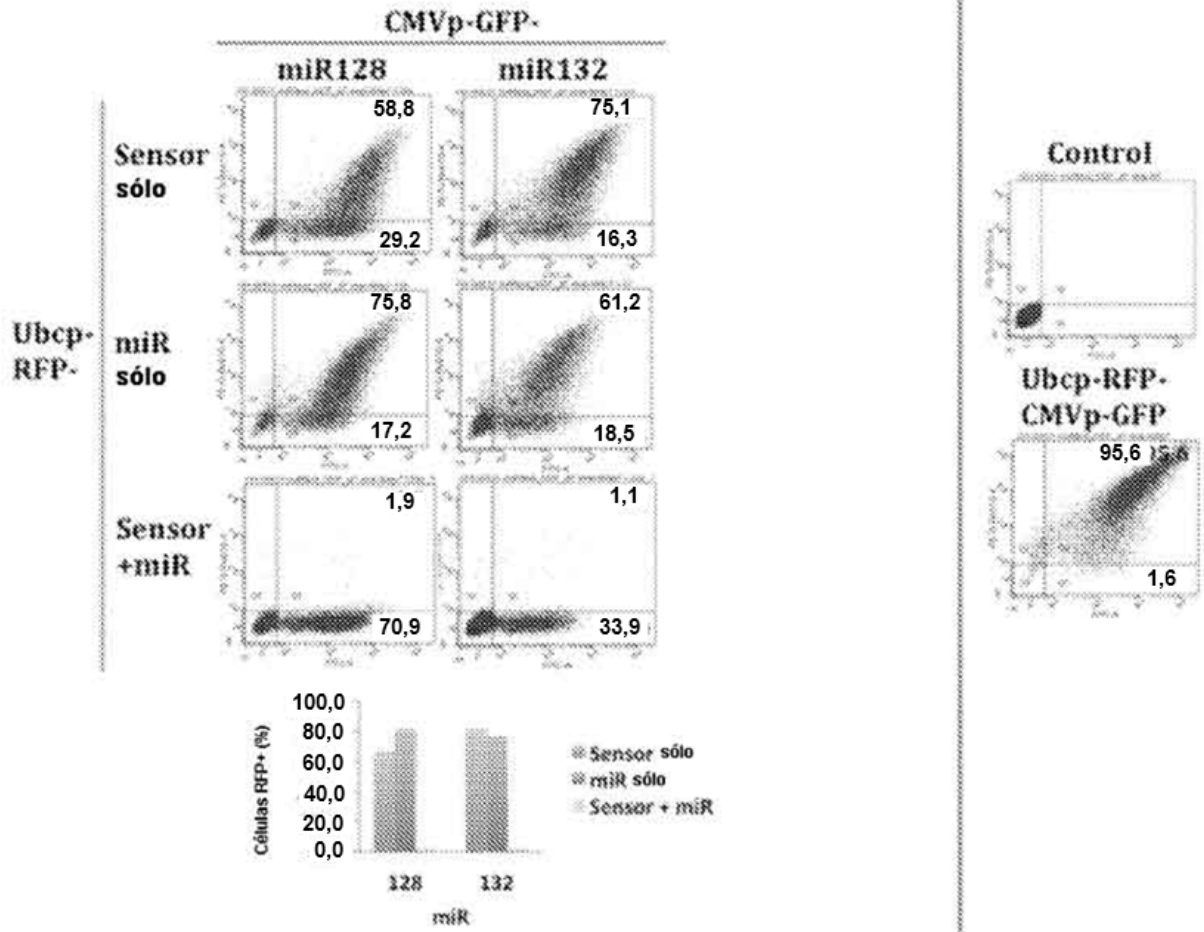
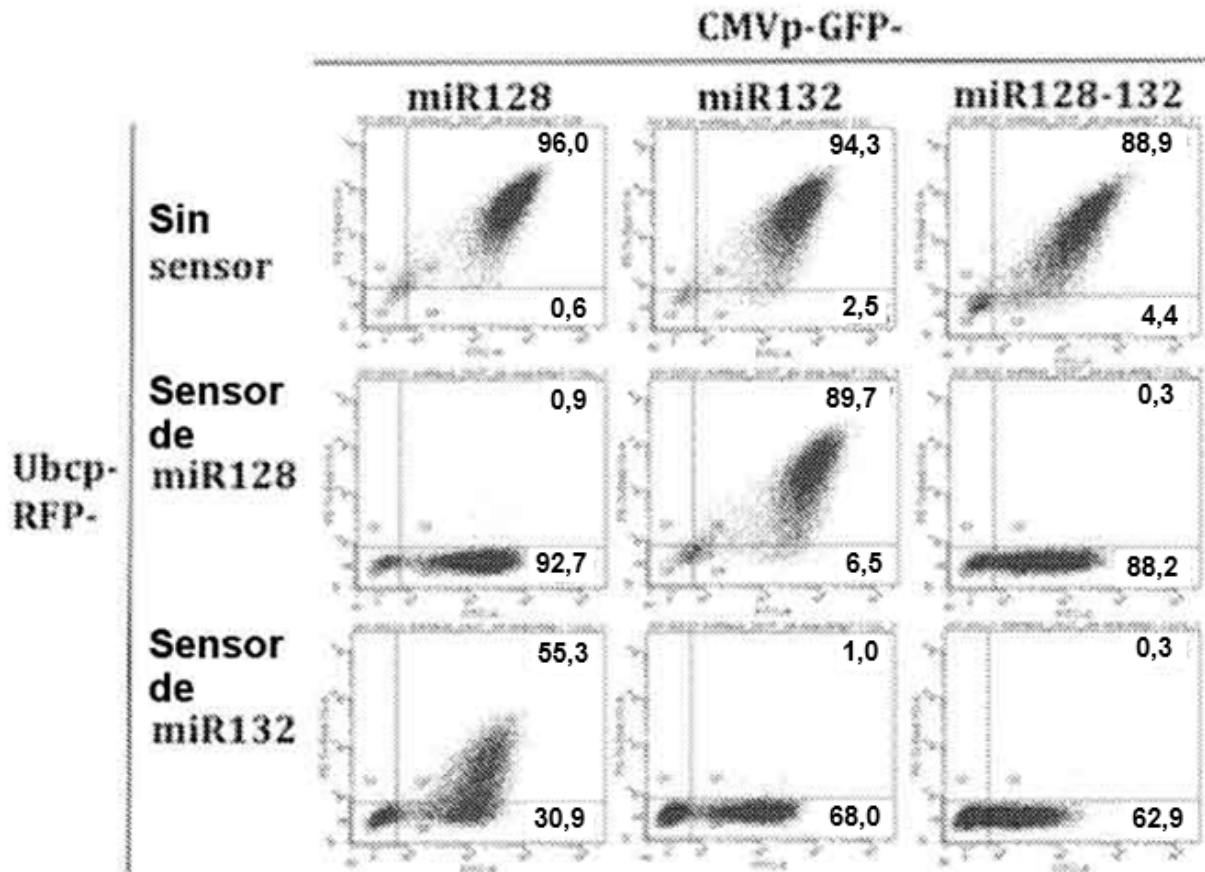


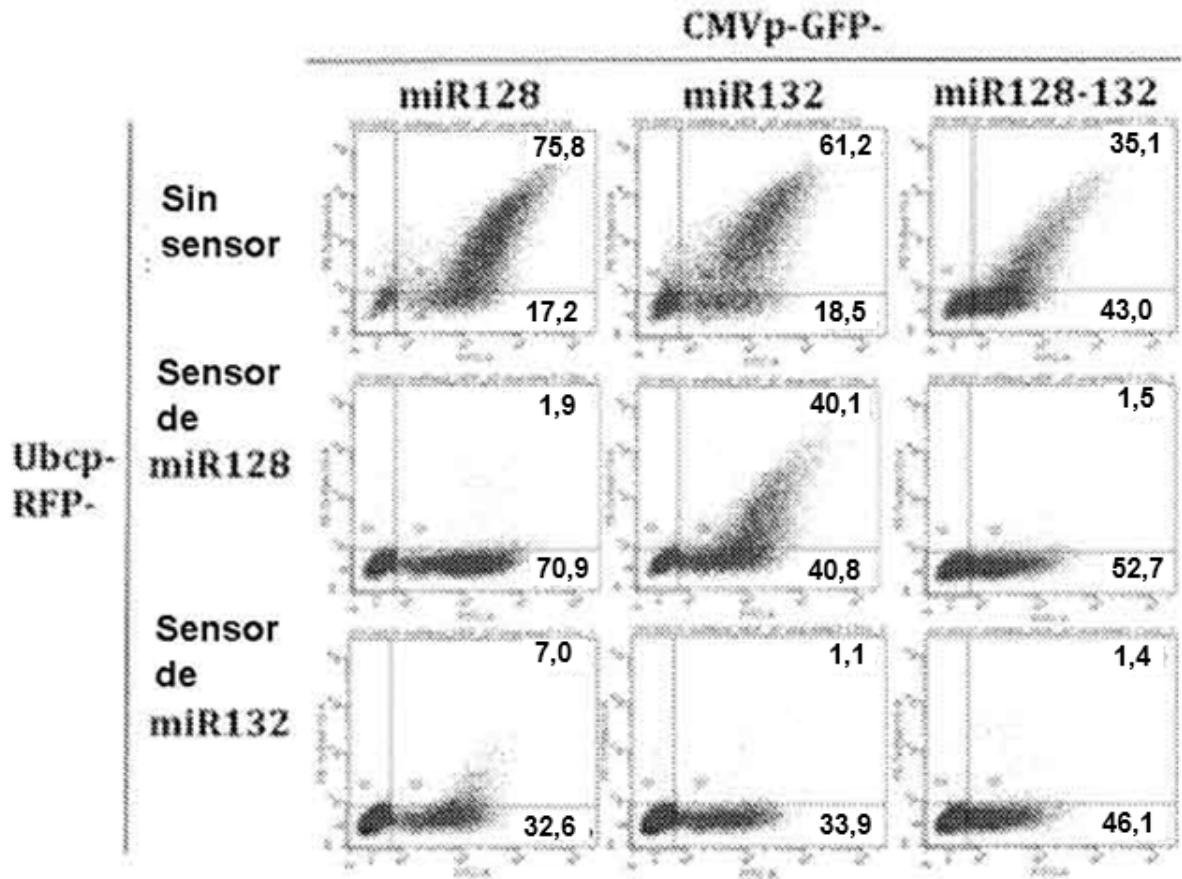
Figura 36



Células RFP+ (%)

		miR		
		128	132	128-132
Sensor	no	99,4	97,4	95,3
	128	1,0	93,2	0,3
	132	64,2	1,4	0,5

Figura 37



Células RFP+ (%)		miR		
		128	132	128-132
Sensor	no	81,5	76,8	44,5
	128	2,6	49,6	2,8
	132	17,7	3,1	2,9

Figura 38

